Making cough count:
The application of cough as a biomarker for respiratory disease screening and monitoring

Alexandra Jaye Zimmer

Department of Epidemiology, Biostatistics, and Occupational Health
Faculty of Medicine and Health Sciences
McGill University
March 2025

A thesis to be submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy (PhD)

Montreal, Quebec, Canada

Table of Contents

Front matter

Abstract

Cough is a symptom common to many acute and chronic respiratory diseases, including COVID-19, tuberculosis (TB), chronic obstructive pulmonary disease, and lung cancer. Its non-specific nature makes it difficult for clinicians to diagnose patients based on cough alone, and patients often struggle to accurately remember cough duration and characteristics. To address these limitations, recent developments in digital health and artificial intelligence (AI) aim to transform cough into an objective biomarker for clinical decision-making. These tools analyze cough sounds to screen for diseases or explore longitudinal cough counts among patients undergoing treatment. This emerging field, termed "acoustic epidemiology", uses technology to detect and analyze bodily sounds to better understand disease dynamics. As technologies and AI advance, there is a need to understand the clinical utility of cough-based tools, the complexities of cough dynamics, and implementation challenges in diverse settings. The objective of this thesis is to advance acoustic epidemiology by evaluating cough as a potential biomarker for respiratory diseases.

The first manuscript is a perspective on recent advances in digital cough analysis tools for TB care. We explored cough's potential as a biomarker across various stages of the TB care cascade.

The second manuscript is a scoping review of digital cough counting tools for longitudinal cough monitoring in respiratory diseases. We identified four key clinical use cases: diagnosis, treatment monitoring, outcome prediction, and syndromic surveillance. Moderate correlations between objective cough counts and patient-reported outcomes highlight cough's complex nature. Our review uncovered implementation challenges, emphasizing the need for real-world validation studies.

The third manuscript examined the relationship between daily cough counts and TB bacterial burden during the first two weeks of treatment. Among 209 TB-positive individuals, cough counts declined over time in all participants. Multivariable analyses showed that participants with lower molecular Xpert semi-quantitative groups ('Medium', 'Low', 'Very low') had lower cough counts

compared to the 'High' group, with rate ratios (RRs) of 0.79 (95% confidence interval [CI]: 0.59, 1.05), 0.64 (95% CI: 0.47, 0.87) and 0.61 (95% CI: 0.41, 0.91) respectively. Lower digital chest X-ray severity scores were also associated with lower cough frequencies (RR: 0.80; 95% CI: 0.67, 0.95). These findings showed that cough frequency correlates with markers of bacterial burden and declines with treatment.

The fourth manuscript explored the external validity of cough-based COVID-19 triage models across populations. We analyzed cough recordings from 605 coughing adults in Lima, Peru and Montreal, Canada. Cough feature analyses revealed significant heterogeneity between the cohorts. Cough-based machine learning algorithms performed well in Lima, achieving an area under the curve (AUC) of 0.71 (standard error [SE]: ± 0.08). Performance was lower within the Montreal dataset with an AUC of 0.53 (± 0.04). Both models showed poor external validity (AUC ± SE: 0.5 ± 0.03 for Lima, 0.51 ± 0.01 for Montreal).

The fifth and final manuscript examined the external validity of cough-based TB triage AI models. We evaluated models from the CODA TB DREAM Challenge using a cough dataset collected from 303 coughing adults in Peru. Model performance decreased from an AUC range of 0.689-0.743 in the CODA Challenge to a range of 0.480-0.615 in the external validation, emphasizing the need for developing and validating cough-based tools using data from the intended use populations.

These studies collectively advance our understanding of cough as a biomarker, highlighting its potential in disease monitoring and screening while underscoring some of the challenges for eventual clinical application.

Resumé

La toux est un symptôme fréquent des maladies respiratoires, notamment la COVID-19, la tuberculose (TB), la bronchopneumopathie chronique et le cancer du poumon. Sa nature non spécifique complique le diagnostic, et les patients peinent à en décrire la durée précise. Les développements en santé numérique et en intelligence artificielle (IA) cherchent à transformer la toux en biomarqueur objectif pour la décision clinique. Ces outils analysent les sons de toux pour dépister les maladies ou suivre leur évolution pendant le traitement. Ce domaine émergent, l'"épidémiologie acoustique", utilise la technologie pour analyser les sons corporels et comprendre la dynamique des maladies. Avec l'avancement des technologies et de l'IA, il faut évaluer l'utilité clinique de ces outils, comprendre la dynamique de la toux et les défis de mise en œuvre. Cette thèse vise à faire progresser l'épidémiologie acoustique en évaluant la toux comme biomarqueur des maladies respiratoires.

Le premier manuscrit est une perspective sur les récentes avancées des outils d'analyse numérique de la toux pour les soins de la tuberculose. Nous avons exploré le potentiel de la toux comme biomarqueur à différentes étapes de la cascade de soins de la tuberculose.

Le deuxième manuscrit revoit les outils numériques de comptage de la toux pour le suivi longitudinal. Nous avons identifié quatre cas d'utilisation clinique: le diagnostic, le suivi du traitement, la prédiction des résultats et la surveillance syndromique. Des corrélations modérées entre les décomptes objectifs de la toux et les résultats rapportés par les patients soulignent la nature complexe de la toux. Notre revue a révélé des défis de mise en œuvre, soulignant le besoin d'études de validation en conditions réelles.

Le troisième manuscrit examine la relation entre les décomptes de toux et la charge bactérienne TB durant les deux premières semaines de traitement. Parmi 209 individus TB-positifs, les décomptes de toux ont diminué au fil du temps. Les analyses multivariées ont montré que les participants avec des groupes Xpert plus faibles ('Moyen', 'Faible', 'Très faible') avaient des décomptes de toux plus faibles comparés au groupe 'Élevé', avec des ratios de taux (RT) de 0,79 (intervalle de confiance [IC] à 95%: 0,59, 1,05), 0,64 (95%: 0,47, 0,87) et 0,61 (95%: 0,41, 0,91).

Des scores de gravité plus faibles à la radiographie thoracique étaient aussi associés à des fréquences de toux plus faibles (RT: 0,80; 95%: 0,67, 0,95). Ces résultats montrent que la fréquence de la toux corrèle avec les marqueurs de charge bactérienne et diminue avec le traitement.

Le quatrième manuscrit explore la validité externe des modèles de triage COVID-19 basés sur la toux. Nous avons analysé les toux de 605 adultes à Lima, Pérou et Montréal, Canada. Les analyses des caractéristiques de la toux ont révélé une hétérogénéité significative entre les cohortes. Les algorithmes d'apprentissage automatique ont bien performé à Lima, atteignant une aire sous la courbe (ASC) de 0,71 (erreur standard [ES]: ± 0,08). La performance était plus faible à Montréal avec une ASC de 0,53 (± 0,04). Les deux modèles ont montré une faible validité externe (ASC ± ES: 0,5 ± 0,03 pour Lima, 0,51 ± 0,01 pour Montréal).

Le dernier manuscrit examine la validité externe des modèles d'IA de triage TB basés sur la toux. Nous avons évalué les modèles du CODA TB DREAM Challenge avec des données de toux de 303 adultes au Pérou. La performance des modèles a diminué d'une plage d'ASC de 0,689-0,743 dans le CODA Challenge à 0,480-0,615 dans la validation externe, soulignant le besoin de développer et valider des outils basés sur la toux avec des données des populations cibles.

Ces études font progresser notre compréhension de la toux comme biomarqueur, soulignant son potentiel dans la surveillance et le dépistage tout en révélant les défis d'application clinique.

Acknowledgements

My PhD journey began during the midst of a global pandemic. While the traditional doctoral experience typically unfolds on campus, mine took place in virtual spaces, navigating research and relationships through computer screens. It would be more than a year before I would meet my supervisor and colleagues in person, and most of my research would take place several time zones away from Montreal. Though challenging, these unique circumstances taught me adaptability and resilience—qualities that were sustained by the unwavering support of numerous individuals who stood beside me throughout these past four years.

First and foremost among these individuals is my supervisor, Dr. Madhu Pai. While Madhu's academic guidance and passionate advocacy for global health issues were invaluable in shaping my development as a researcher, it was his unwavering trust that truly defined my doctoral experience. His trust in my ability to forge my own path, make independent research decisions, and deliver results gave me the confidence to grow as a researcher. This model of mentorship is one I hope to pass forward to future mentees in my own career.

I am also deeply grateful to my thesis committee members: Dr. Samira Rahimi, Dr. Andrea Benedetti, and Dr. Simon Grandjean Lapierre. Samira helped me navigate the initially daunting field of artificial intelligence, providing me with the tools and resources to transform an unfamiliar territory into a field I can confidently engage with. Andrea offered invaluable statistical support for my complex datasets, generously making time to walk me through analyses and code. And, of course, Simon, who went above and beyond his role as a committee member and without whom this thesis simply wouldn't exist.

A big thank you to all former and current members of Team Pai. I am deeply grateful to be part of this supportive network. Among this group, I would particularly like to thank Ms. Caroline Vadnais for her invaluable administrative support, help with grant preparation, and guidance through the maze of the IRB. Special thanks Dr. Sophie Huddart, whose expertise in cough analysis, acoustic epidemiology, and TB has been invaluable to my work. Finally, I want to thank Dr. Lena Faust, who has been both a valued colleague and became dear friend throughout this journey. From

sharing research insights to our adventures together during data collection in Peru, she was there every step of the way.

This work in Peru relied on an exceptional team. I am particularly grateful to Dr. Cesar Ugarte-Gil and Dr. Patricia Espinoza Lopez from Universidad Peruana Cayetano Heredia (UPCH). Cesar's trust in our project and his help in establishing our field site were instrumental to its success. Visiting the UPCH facilities in Lima was a highlight, and I look forward to future collaborations. Patricia's dedication and work ethic as the study coordinator were truly inspiring—she consistently turned the impossible into possible with unwavering motivation. Her passion for research was contagious, and I am excited to watch her career in global health research continue to flourish.

While I am endlessly grateful to all my friends and family who have given me encouragement and support throughout this journey, I would be remiss not to single out two individuals in my close circle who truly shaped me as a person during this transformative time. I first met Doris Durán over Zoom—two PhD students struggling with Stata code for an assignment. What began as peer support evolved into an invaluable friendship. As I approach the end of this journey, I cannot imagine having done it without her. She was there during my loneliest moments—even from thousands of kilometers away—turning isolation into connection, and challenges into shared victories. Finally, my partner Alex (#2) Pabouctsidis. Words fall short in expressing my gratitude for your unwavering support throughout this PhD journey. Your encouragement, patience, and endless optimism (even in the face of my pessimism) gave me strength when I needed it most. Thank you for keeping me grounded, well-fed, and reminding me of life beyond my research.

Special mention to my cat Pixel, who arrived in my life during the final year of my PhD. Her companionship brought both comfort and occasional chaos to this journey's conclusion.

Financial support

I am grateful for the financial support I received throughout my PhD. I held a Harrison-Watson Faculty of Medicine Studentship from 2021-2022. I was awarded a Canadian Institutes of Health Research (CIHR) Doctoral Research Award in 2022, that supported my research until the end of my PhD. Throughout my degree, I received supplementary funding from the McGill Global Health Scholars program as well as a travel award from the McGill International TB Center. Throughout my degree, I received supplemental funding from my supervisor Dr. Madhu Pai.

Contribution of authors

As the first author on all manuscripts, I co-wrote grants, developed study protocols, prepared ethical documents, oversaw data collection, analyzed the data, and wrote the manuscripts. I performed all of this with the guidance and support of my supervisor, committee members, and co-authors of each manuscript. Contributions of co-authors to each manuscript are detailed below.

Manuscript I: Making cough count in tuberculosis care

This perspective manuscript on the field of cough and tuberculosis (TB) care was conceptualized by my supervisor Dr. Madhukar Pai, my committee member Dr. Simon Grandjean Lapierre, and me. I wrote the first manuscript draft, with Dr. Pai and Dr. Grandjean Lapierre providing feedback. Additional input was obtained from other experts on cough and TB, including Dr. César Ugarte-Gil, Dr. Rahul Pathri, Dr. Puneet Dewan, Dr. Devan Jaganath, and Dr. Adithya Cattamanchi. All co-authors provided critical feedback and approved the final manuscript.

Manuscript II: Objective cough counting in clinical practice and public health: a scoping review

I conceptualized this scoping review, with input from committee member Dr. Simon Grandjean Lapierre. I wrote the scoping review's protocol, with input from Dr. Madhukar Pai, Dr. Simon Grandjean Lapierre, Dr. César Ugarte-Gil, Mrs. Genevieve Gore, Mr. Rishav Das, Ms. Patricia Espinoza, and Mrs. Vaidehi Nafade. I developed the search term and executed the search strategy with input from expert research librarian Mrs. Gore. Publication screening and data extraction was performed by Mr. Das, Ms. Espinoza, Mrs. Nafade, and me. I performed all data cleaning and synthesis tasks under the supervision of Dr. Grandjean Lapierre. I drafted the manuscript. All co-authors provided critical feedback and approved the final manuscript.

Manuscript III: Baseline tuberculosis severity as a predictor of patient cough trajectory during the first two weeks of anti-tuberculosis therapy

This study was conceptualized by Dr. Simon Grandjean Lapierre and me. The secondary data used in this study was provided by Dr. Grandjean Lapierre and co-authors from the University of California, San Francisco who are part of the Rapid Research in Diagnostics Development for TB Network (R2D2). I performed the data cleaning and all analyses, with statistical support from my

committee member Dr. Andrea Benedetti and from Dr. Sophie Huddart. I interpreted the data, prepared figures and tables, and wrote the manuscript. Co-authors include principal investigator (PI) for the R2D2 Network who were involved in data collection within-country. All co-authors provided critical feedback and approved the final manuscript.

Manuscript IV: Cough acoustics for COVID-19 detection: a comparative study of patient cohorts from Lima, Peru and Montreal, Canada

This study was conceptualized by Dr. Madhukar Pai, Dr. Simon Grandjean Lapierre, and me. Dr. Pai, Dr. Grandjean Lapierre and I wrote the Canadian Institutes of Health Research (CIHR) grant application, with input from my committee member Dr. Samira Rahimi, and Dr. César Ugarte-Gil. I was the primary grant writer for the CIHR project. I oversaw all primary data collection activities in Montreal, Canada and Lima, Peru. Dr. Grandjean Lapierre was the local PI in Montreal and Dr. Ugarte-Gil was the local PI in Lima. Ms. Patricia Espinoza was the in-country study coordinator in Peru. I designed all data collection tools and performed the data cleaning tasks. Model development and analyses of the cough audio recordings were performed by Dr. Vijay Ravi, Dr. George Kafentzis, and me. Dr. Ravi worked on the neural network approach while I worked on the machine learning approach with support from Dr. Kafentzis. Dr. Rahimi and Dr. Mirco Ravanelli oversaw the model development. I drafted the manuscript and generated all tables and figures. All co-authors provided critical feedback and approved the final manuscript.

Manuscript V: External validation of cough-based algorithms for pulmonary tuberculosis screening from the CODA TB DREAM Challenge using cough data from Peru

This study used the Peru cough data that was collected from our CIHR grant, as detailed in Manuscript IV above. The same individuals were involved in grant writing, data collection, and data cleaning as was presented for Manuscript IV. The analysis of this model was done using AI algorithms that were externally developed as part of the COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM Challenge. Dr. Solly Sieberts helped me obtain the algorithms from the Sage Bionetworks repository. I performed all external validation analyses using the Peru audio dataset. I also drafted the manuscript and generated all tables and figures. All co-authors provided critical feedback and approved the final manuscript.

Statement of originality

I hereby declare that all manuscripts presented in this thesis constitute original scholarship and make significant contributions to knowledge in the field. The specific contribution of each manuscript is detailed below.

Manuscript I is a perspective manuscript that presents a unique synthesis of current knowledge and future potential of digital cough monitoring in TB management. Our perspective introduces the emerging field of 'acoustic epidemiology' and explores its relevance to TB care, offering novel insights into how cough sounds can be leveraged to improve various aspects of TB diagnosis, treatment monitoring, post-care support, and public health surveillance.

Manuscript II is a scoping review that provides the first comprehensive synthesis of digital cough counting tools in clinical practice and public health. While numerous cough counting tools have been developed, our review uniquely maps their clinical utility and public health impact, addressing a critical gap between technological innovation and practical implementation.

Manuscript III is an original research study that provides the first comprehensive analysis of the association between baseline TB bacterial burden and longitudinal cough patterns during the initial phase of treatment. Our work uniquely combines advanced statistical modeling with continuous cough monitoring data from a diverse, multi-country cohort to offer novel insights into cough dynamics among TB patients. Our study addresses critical gaps in understanding how baseline bacterial burden influences cough trajectories, informing the development of future cough-based biomarkers for TB treatment response.

Manuscript IV is an original research study that provides a novel comparative analysis of cough acoustics for COVID-19 detection across two distinct populations in Lima, Peru and Montreal, Canada. Our work uniquely leverages high-quality cough recordings and gold-standard diagnostic data to investigate the external validity of cough classification models across populations. While numerous COVID-19 cough detection algorithms have been developed since the pandemic, our study is the first to comprehensively explore and demonstrate the population-specific nature of

cough acoustics and the limited transferability of these models across different geographical, demographic, and epidemiologic contexts.

Manuscript V presents the first independent external validation of artificial intelligence-based cough screening algorithms for TB developed as part of the CODA TB DREAM Challenge. Using high-quality cough recordings and reference standard testing that our team collected in Lima, Peru, we assessed the performance of these algorithms in a geographically distinct population. Our findings provide critical insights into the transferability of TB cough triage tools and highlight the challenges in developing globally applicable algorithms for TB detection.

List of abbreviations

ACF	Autocorrelation function
AI	Artificial intelligence
AIC	Akaike Information Criteria
AR1	Autoregressive 1
AUC	Area under the curve
BMI	Body mass index
CAD	Computer-aided detection
CayeCoM	Cayetano Cough Monitor
CI	Confidence interval
CODA TB	COugh Diagnostic Algorithm for Tuberculosis
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
cph	Coughs per hour
CSS	Cough symptom score
Ct	Cycle threshold
DOT	Directly observed therapy
ECAPA-TDNN	Emphasized Channel Attention, Propagation and Aggregation Time Delay Neural Network
EPTB	Extrapulmonary tuberculosis
Fbank	Filter bank
GLMM	Generalized linear mixed model
HIV	Human immunodeficiency virus
IRB	Institutional review board
IQR	Interquartile range
kNN	k-nearest neighbor
LCM	Leicester Cough Monitor
LCQ	Leicester Cough Questionnaire
LJ	Löwenstein–Jensen
LMIC	Low- and middle-income country

LOESS	Locally estimated scatterplot smoothing
MFCC	Mel-frequency cepstral coefficient
MICE	Multiple imputation by chained equation
ML	Machine learning
MUAC	Mid-upper arm circumference
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Mtb	*Mycobacterium tuberculosis*
NAAT	Nucleic acid amplification test
PC-QoL	Parental Cough-Specific Quality of Life
PCR	Polymerase chain reaction
PRO	Patient reported outcome
PTB	Pulmonary tuberculosis
R2D2	Rapid Research in Diagnostics Development for Tuberculosis Network
RCC	Refractory chronic cough
RIF	Rifampicin
RR	Rate ratio
RSV	Respiratory syncytial virus
SHAP	SHapley Additive exPlanation
SMOTE	Synthetic Minority Over-sampling Technique
SE	Standard error
SP	Standardized patients
STFT	Short-time Fourier transform
TB	Tuberculosis
TPP	Target product profile
VAS	Visual analog scale
XGBoost	eXtreme Gradient Boosting
Xpert	GeneXpert
WHO	World Health Organization
ZCR	Zero-crossing rate

List of tables

List of figures

Chapter 2. Literature review

Chapter 3. Manuscript I: Making cough count in tuberculosis care

Chapter 4. Manuscript II: Objective cough counting in clinical practice and public health: a scoping review

Chapter 5. Manuscript III: Baseline tuberculosis severity as a predictor of patient cough trajectory during the first two weeks of anti-tuberculosis therapy

Chapter 1. Introduction

1.1. Introduction

The COVID-19 pandemic transformed how society perceives cough, shifting it from a common, often overlooked symptom to a signal that could clear a room.[1] This heightened public awareness parallels a growth in researchers and clinicians beginning to explore cough's untapped potential in clinical care.[2]

While cough has traditionally been understood as both a vital protective reflex and a troublesome symptom,[3] it manifests across a spectrum of respiratory conditions, ranging from acute infectious diseases to chronic respiratory disorders. Acute cough is commonly associated with viral respiratory infections like influenza and COVID-19, bacterial infections such as tuberculosis (TB), and community-acquired pneumonia. In contrast, chronic cough, persisting beyond eight weeks, characterizes conditions like asthma, chronic obstructive pulmonary disease, and lung cancer.[4]

Despite the clinical significance of cough across these conditions, clinicians currently rely primarily on imperfect patient recall, subjective descriptions and basic characterizations of cough for assessment and monitoring.[5] However, modern technological advances have revealed another dimension in which cough serves as a rich source of diagnostic information. A single cough event contains complex acoustic signatures that could indicate underlying pathology,[6–8] and the temporal patterns of cough frequency, daily distribution of cough episodes, and overall duration of the symptom can reflect disease progression, treatment response, and recovery trajectories.[9,10] The widespread adoption of smartphones and increasing acceptance of digital health tools has made it increasingly possible and acceptable to capture data like cough-sounds in real-world settings.[11] When combined with artificial intelligence (AI) algorithms capable of analyzing acoustic features and identifying patterns in cough data, this multidimensional information could help clinicians differentiate between diseases, assess severity, and monitor treatment response.

These technological advances in cough analysis have given rise to a new field known as acoustic epidemiology, which leverages sound-based data to understand disease patterns and improve

health outcomes.[12,13] This emerging discipline combines digital health technologies, AI, and clinical medicine to transform traditionally subjective symptoms into objective, measurable biomarkers. However, realizing this potential requires careful consideration of technical capabilities, clinical validation, and real-world implementation challenges across diverse healthcare settings and populations.

1.2. Research gaps addressed by this thesis

The COVID-19 pandemic catalyzed rapid development of AI algorithms for cough classification and accelerated the availability of digital cough monitoring tools. Despite these technological advances, this data-rich symptom remains largely reduced to basic present/absent documentation in clinical practice. While numerous promising tools and algorithms have been developed, few have successfully transitioned from research to routine clinical use. This represents a missed opportunity in an era where AI and digital health tools are transforming other areas of medicine.[14] The persistent gap between cough's potential as a biomarker and its current clinical utilization suggests fundamental challenges in both analysis and implementation that must be addressed. This thesis investigates multiple dimensions of cough analysis—from acoustic signatures to longitudinal temporal patterns—to highlight the potential of cough and identify key barriers preventing the successful integration of cough-based tools into clinical practice.

I begin this exploration by examining cough monitoring through a disease-specific lens, focusing on TB, the leading infectious cause of death globally (**Manuscript I**). This perspective article identifies key opportunities and challenges for implementing cough-based tools in TB care, providing a roadmap for researchers and clinicians working in acoustic epidemiology. Building on this foundation, I conducted a comprehensive scoping review to map the landscape of existing cough counting tools across all respiratory conditions (**Manuscript II**). This systematic analysis allowed me to better understand what cough counting tools are available, how they are being used, and the existing challenges for implementing them in a clinical context.

In **Manuscript III**, I conducted the first comprehensive analysis examining the relationship between baseline TB disease bacterial burden and cough patterns during the initial phase of

treatment. This study not only revealed important insights into cough dynamics among TB patients but also highlighted previously overlooked analytical challenges in processing continuous cough data, establishing methodological considerations for future studies that analyze longitudinal cough data.

Both **Manuscript IV** and **Manuscript V** examine critical limitations in AI-based cough classification algorithms for COVID-19 and TB screening, respectively. Early development of these algorithms was driven by an assumption that cough acoustic features would be universal across populations, leading to efforts to create globally applicable screening tools. Through external validation studies, these manuscripts demonstrate that population-specific factors significantly influence cough acoustics, challenging the notion of universal cough classification models. The findings highlight the need to reconsider how we develop and validate cough-based screening tools, suggesting that targeted population-specific approaches may be more appropriate.

1.3. Thesis goal and objectives

The overall goal of this thesis is to advance the clinical utility of digital cough assessment tools by examining the potential of cough as objective biomarkers for respiratory disease screening and monitoring. This goal is addressed through five manuscripts with the following objectives:

Manuscript I:

1. To explore the potential of acoustic epidemiology and digital cough monitoring tools across the TB care cascade, from screening and diagnosis to post treatment care.

Manuscript II:

2. To identify and categorize the various digital cough counting tools and technologies used in clinical and public health contexts.
3. To analyze how digital cough counting tools are currently being used for the diagnosis, monitoring, and management of respiratory diseases, and their effectiveness in these roles.

4. To investigate the technological and logistical factors that affect the adoption and integration of digital cough counting tools into clinical practice and public health strategies.

Manuscript III:

5. To characterize the relationship between cough counts during the first 14 days of treatment and markers of baseline TB bacterial burden.

Manuscript IV:

6. To characterize population-level differences in cough acoustic features between distinct geographic, demographic, and epidemiologic cohorts.
7. To assess the external validity of AI and machine learning COVID-19 cough classification algorithms between two cohorts.

Manuscript V:

8. To conduct an external validation of COugh Diagnostic Algorithm for Tuberculosis Challenge AI cough classification models.

Chapter 2. Literature review

2.1. Epidemiology of cough and cough-related diseases

2.1.1. Cough epidemiology

The epidemiology of cough is complex, as it is associated with various respiratory diseases and conditions. Cough is typically categorized as either chronic or acute, with chronic cough often linked to conditions such as chronic obstructive pulmonary disease (COPD), while acute cough may be attributed to infections like COVID-19. Accurately estimating the prevalence of cough poses significant challenges due to the heterogeneity in cough types, definitions, and underlying aetiologies. A 2015 meta-analysis estimated the global prevalence of chronic cough at approximately 9.6%, which aligns closely with the prevalence of associated conditions such as asthma and COPD.[15] A more recent meta-analysis on cough of any etiology revealed a prevalence ranging from 3.8-4.2% in Western countries to 10.3-13.8% in Africa, Asia, and South America.[16] Demographic variations in cough prevalence and cough sensitivity have been observed, with women demonstrating a more sensitive cough reflex compared to men,[17,18] meaning that the level of stimuli needed to illicit a cough is lower compared to men. This biological difference is reflected in clinical presentations, as evidenced by a study of 10,000 consecutive patients presenting with chronic cough, where approximately 66% were female.[19]

The epidemiology of cough can be further elucidated by examining the burden of diseases associated with this symptom. The 2019 Global Burden of Disease study estimated that chronic respiratory diseases accounted for 103.5 million disability-adjusted life years (DALYs), representing 4.1% of global DALYs for all causes.[20] The geographic distribution of DALYs due to chronic respiratory diseases disproportionately affected sub-Saharan Africa, South Asia, and South-East Asia (**Figure 2.1**). Lower respiratory tract infections, including tuberculosis (TB), were responsible for 5.7% of global DALYs in 2019.[21]

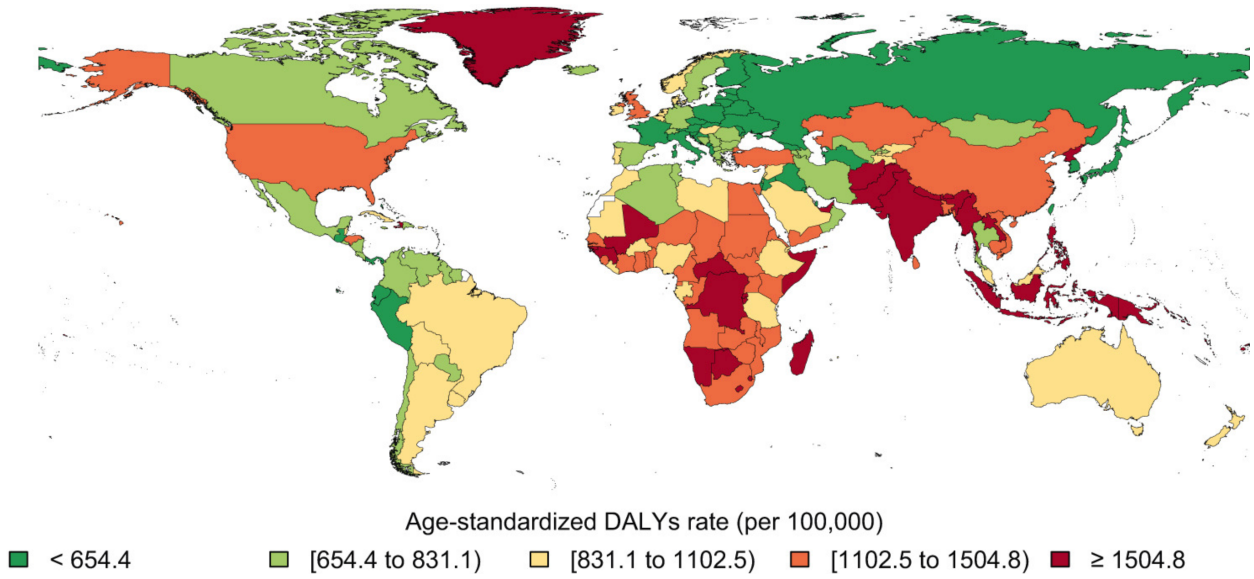


Figure 2.1. Age-standardized disability-adjusted life years (DALYs) due to chronic respiratory diseases. DALYs are highest in sub-Saharan Africa, South Asia, and South-East Asia. Main contributing diseases are chronic obstructive pulmonary disease (COPD), asthma, pulmonary sarcoidosis, and interstitial lung disease. (Source: Momtazmanesh et al. 2019.[20])

2.1.2. COVID-19 epidemiology

In November 2019, a highly infectious coronavirus (SARS-CoV-2) emerged, rapidly spreading globally and prompting the World Health Organization (WHO) to declare a pandemic in early 2020.[22] The resulting disease, COVID-19, has since caused widespread devastation, with an estimated 7 million cumulative deaths reported worldwide.[23] The burden of disease was particularly acute during the initial two years (2020-2021) due to limited population-level immunity and restricted vaccine availability and vaccine inequities, especially in low- and middle-income countries (LMICs).[24] The profound impact of the COVID-19 pandemic on global health is reflected in the 2021 Global Burden of Disease study, which ranked COVID-19 as the leading cause of DALYs, accounting for 212.0 million DALYs (**Figure 2.2**).[25]

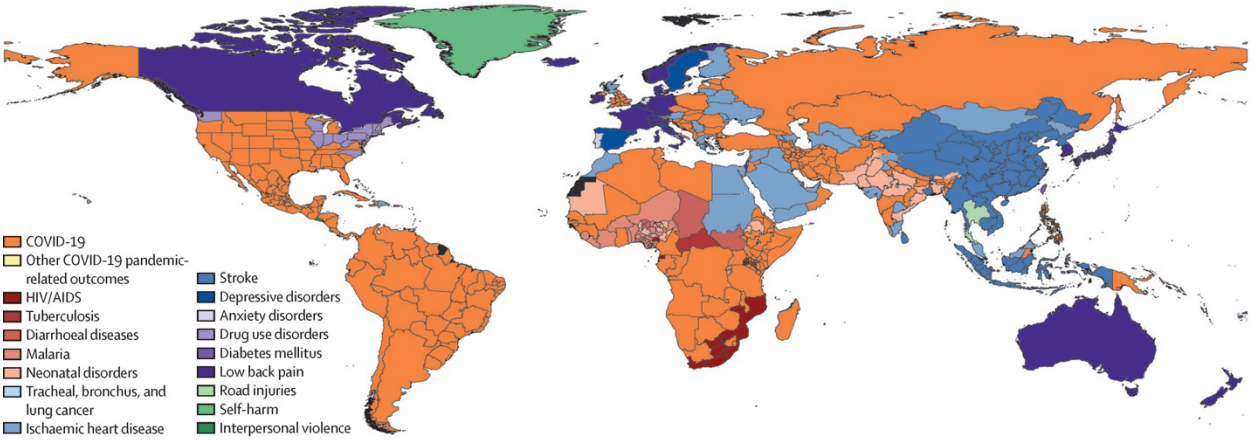


Figure 2.2. Global distribution of leading causes of disability-adjusted life years (DALYs) in 2021. This world map illustrates that COVID-19 (shown in orange) was the leading cause of DALYs across many regions, including North and South America, parts of Europe, and significant portions of Africa and Asia. (Source: Ferrari et al. 2021.[25])

As the pandemic progressed, the virus evolved through mutations, leading to the emergence of several variants. The WHO designated certain variants as "variants of concern," including Beta (B.1.351), Alpha (B.1.1.7), Delta (B.1.617.2), Gamma (P.1), and Omicron (B.1.1.529).[26] These new variants altered the virus's transmissibility, with Omicron becoming the dominant variant by the end of 2022.[27]

The evolution of SARS-CoV-2 variants also influenced the symptomatology of COVID-19. One study found that the proportion of Omicron-infected individuals reporting myalgia and sore throat was more than double compared to those infected with the initial variant.[28] Despite these shifts in symptom prevalence, cough remained a significant clinical feature across all variants, with studies consistently reporting cough in 40-60% of COVID-19 patients, regardless of the infecting variant.[28,29] Furthermore, cough has emerged as a prevalent symptom in post-acute sequelae of SARS-CoV-2 infection, commonly known as 'long COVID'. A comprehensive systematic review has revealed that COVID-associated cough can persist for up to 24 weeks in affected individuals, underscoring the long-term respiratory implications of the disease.[30]

2.1.3. Tuberculosis epidemiology

Prior to COVID-19, and again in 2023, TB was the leading cause of death due to an infectious pathogen. In 2023, it is estimated that over 10.8 million individuals had active TB and 1.25 million individuals died due to TB.[31] Despite being a global disease, the prevalence of TB is highly correlated to social inequities, with individuals living in poverty or with limited access to healthcare services being disproportionately affected by the disease.[31,32] The burden of TB disease is primarily concentrated within 30 high burden countries, all of which are LMICs (**Figure 2.3**).[31]

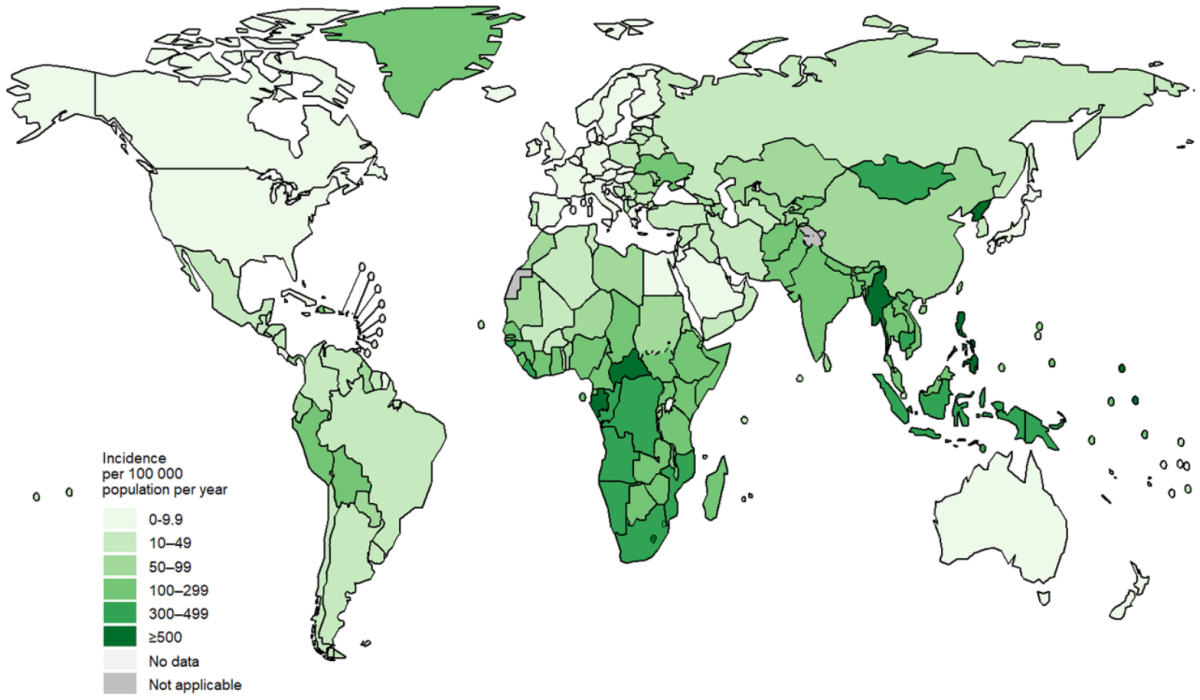


Figure 2.3. Estimated incidence of tuberculosis (TB) in 2023. TB is a global disease, present in all countries. The incidence and burden of disease is greatest in certain regions, including sub-Saharan Africa, South Asia, and South-East Asia. (Source: Global Tuberculosis Report, World Health Organization, 2024.[31])

TB is primarily a disease of the lungs, referred to as pulmonary TB, though it can also manifest in other parts of the body, known as extrapulmonary TB. The natural history of pulmonary TB is complex and is evolving as more is discovered about the disease.[33,34] A "continuum or spectrum of TB disease" model has been proposed, highlighting the different stages of disease (**Figure**

2.4).[35] Individuals with TB can fall along this continuum and progress forwards (towards active TB disease) or backwards (towards TB infection, or "latent" TB).[36] As people progress towards active TB, the bacterial burden in their lungs typically increases, leading to both greater potential infectiousness and the development of more severe symptoms. Cough is a prominent symptom of active pulmonary TB disease. Prolonged cough (≥14 days) is often used to triage patients for additional confirmatory microbiological testing.[37] Additional background information on the importance of cough and active pulmonary TB is presented in **Chapter 3, Manuscript I**. In recent years, research has focused on the subclinical presentation of TB, defined as individuals with bacteriological evidence of TB but do not report having symptoms.[36,38] Individuals with subclinical TB may be truly asymptomatic or they may have symptoms that are not recognized or acted upon, in terms of care seeking.[36] This also may extend to cough, where individuals with subclinical TB may not recognize an increase in their cough frequency or misattribute the cough to other causes.[39,40]

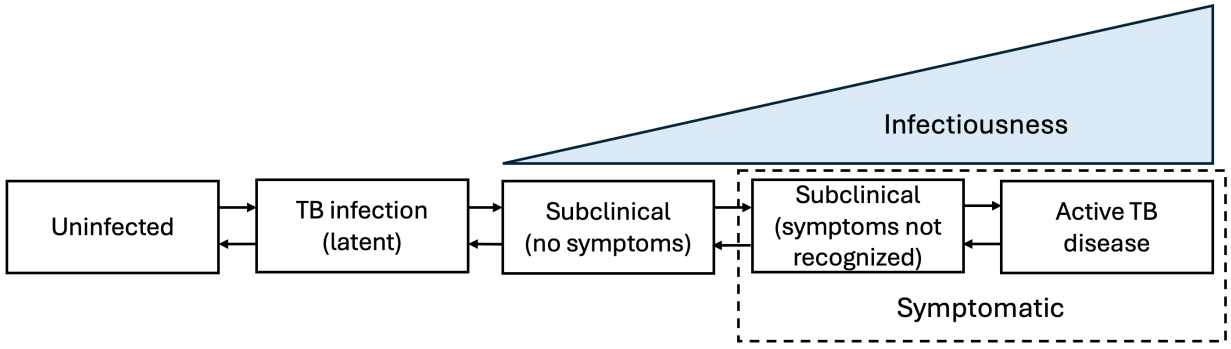


Figure 2.4. Continuum of tuberculosis (TB) disease. Disease severity progresses from Uninfected (left) to Active TB disease (right). Transition between disease states is bi-directional, suggesting that individuals may progress to more severe disease states as well as regress to milder disease. Subclinical TB can be truly asymptomatic or may be due to the individual not recognizing that their symptoms are related to TB. Infectiousness increases as individuals progress from Subclinical (no symptoms) to Active TB disease. (Figure adapted from Kendall et al.[36])

2.2. Cough as a clinically meaningful symptom

2.2.1. Importance for patients

A cough often acts as an indicator of underlying health issues and can significantly impact a patient's quality of life, prompting individuals to seek medical attention. In a comprehensive study of 13,902 patients with chronic and acute respiratory disorders, cough was reported as the most frequent symptom and primary reason for visiting medical facilities.[41] However, care-seeking behavior related to cough is more likely to occur if the cough is persistent, distressing, or associated with other symptoms.[42] A survey of 3,333 individuals in the United States revealed that 43% of respondents waited until their cough was 'bad enough' to seek care, while 20% waited for additional symptoms to present.[43] This hesitation in seeking care is even more pronounced in certain conditions; for instance, among people with TB, a study found that only 15% sought care based on cough alone, compared to 40% when hemoptysis was present.[44] Improved awareness of cough that persists beyond expected durations, particularly among high-risk populations and in TB-endemic regions, could improve appropriate care-seeking behavior and timely diagnosis

2.2.2. Importance for clinicians

For clinicians, the characteristics of a cough, including its type (e.g., wet or dry), duration, frequency, and associated symptoms (e.g., hemoptysis), can provide valuable insights into potential diseases or clinical states. Numerous clinical guidelines have been developed to assist healthcare professionals in evaluating the causes of both chronic and acute cough.[40,45–49]

The diagnostic significance of cough is evident across a spectrum of respiratory conditions. For instance, a persistent and productive cough, especially when accompanied by shortness of breath and chest tightness, can be predictive of a decline in lung function associated with COPD.[50,51] Asthma or cough variant asthma may present with a dry or minimally productive cough that worsens at night and is triggered by exercise or cold air.[52] In the case of *Bordetella pertussis* infections, the paroxysmal phase is characterized by bursts of rapid coughs followed by gasps, producing the distinctive 'whoop' sound.[53,54] In high-burden TB settings, a persistent cough lasting

more than two weeks is commonly used as a screening tool for pulmonary TB.[37] Conversely, acute coughs, defined as those present for three weeks or less, are often attributed to viral infections of the upper and lower respiratory tracts.[55] Beyond its diagnostic utility, cough can also serve as a prognostic indicator. A study involving patients with idiopathic pulmonary fibrosis demonstrated that individuals presenting with cough were more likely to experience adverse health outcomes (adjusted odds ratio [OR]: 4.97; 95% confidence interval [CI]: 1.25, 19.80).[56] In disease management, cough frequency has been observed to be higher among patients with uncontrolled asthma compared to those with controlled asthma.[57]

Despite its clinical significance, cough remains a non-specific symptom, as it is common across many respiratory diseases and conditions. This non-specificity poses a significant challenge for clinicians attempting to make diagnoses based solely on cough sounds. A study highlighted this difficulty, revealing that while healthcare professionals could correctly determine whether a cough produced mucus 76% of the time, their ability to identify the underlying clinical diagnosis based on cough alone was accurate only 34% of the time.[58]

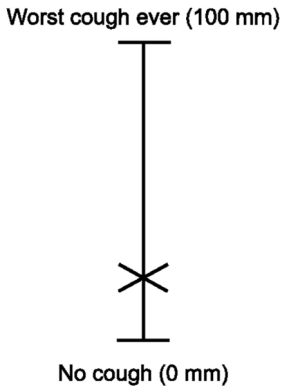
Despite these challenges, cough remains an important symptom to consider in a clinical context. Efforts to extract meaningful information from cough have led to the development of various assessment tools, both subjective and objective. These tools aim to standardize cough evaluation, quantify its characteristics, and potentially improve diagnostic accuracy. The following sections will explore these assessment methods, beginning with subjective measures that rely on patient-reported outcomes and clinician judgement, followed by a discussion of emerging objective technologies designed to capture and analyze cough sounds and patterns more precisely.

2.3. Subjective cough measurement and assessment tools

Subjective cough assessment tools have been developed to aid clinicians in evaluating the impact and severity of cough. These tools, typically in the form of questionnaires or scales, are designed to be easily administered in clinical settings. They primarily serve two purposes: assessing cough severity and understanding its effect on quality of life.[59]

Cough severity is commonly evaluated using the Visual Analog Scale (VAS) and the Cough Symptom Score (CSS). The VAS is a 100-mm linear scale where patients mark their perceived cough severity between "no cough" (0 mm) and "worst cough ever" (100 mm) **(Figure 2.5A**).[46] The CSS, ranging from 0 to 5, considers cough frequency, intensity, and influence during both day and night (**Figure 2.5B**).[60]

A) Cough VAS



B) Cough CSS

Score	Daytime	Night-time
0	No cough during the day	No cough during the night
1	Cough for one short period	Cough on waking only
2	Cough for more than two short periods	Wake once or early due to cough
3	Frequent coughing, which did not interfere with usual daytime activities	Frequent waking due to coughs
4	Frequent coughing, which did interfere with usual daytime activities	Frequent coughs most of the night
5	Distressing coughs most of the day	Distressing coughs preventing any sleep

Figure 2.5. Example of subjective cough assessment tools. A) Visual analog scale (VAS) for reporting cough severity. Patients are asked to make a mark along a 100 mm scale, with a higher mark indicative of a worse cough. B) Cough Symptom Score (CSS) scale for cough severity. Based on the patient's reported daytime and nighttime coughing patterns, the clinician can determine the severity of their cough on a scale from 0 to 5, with higher numbers indicating worse cough. (Figure A is adapted from Morice et al.[46]; Figure B is adapted from Wang et al.[60])

To capture the broader health impacts of cough, including psychological morbidity, health-related quality of life questionnaires (HRQLQ) are employed.[61] The most widely used are the Leicester Cough Questionnaire (LCQ) and the Cough-specific Quality of Life Questionnaire (CQLQ).[60,61] The LCQ, a 19-item questionnaire, covers physical, psychological, and social effects of cough. The CQLQ, with 28 items, assesses somatic symptoms, social psychology, functional ability, emotional state, extreme somatic symptoms, and personal safety fears.

While these subjective tools offer the advantages of easy administration both at time of consultation and repeatedly over time, concerns exist regarding their reliability and potential bias.

Recall bias has been found when patients are asked to retrospectively reflect on their cough, with one study finding that patients tend to underestimate their cough severity.[62] Various forms of response bias may also occur. For example, scale-based measurements like the VAS can introduce 'end-of-scale' bias, where respondents are either more or less likely to use the extreme ends of the scale.[63] Finally, the subjective nature of these assessments can be influenced by patients' perception of their symptoms and the manner in which healthcare providers administer the questionnaires. This subjectivity raises questions about the tools' accuracy in reflecting the true clinical picture and their effectiveness in guiding treatment decisions.

2.4. Objective cough measurement and assessment tools

The limitations of subjective cough assessment tools have led to the emergence of two distinct digital analysis modalities. The first encompasses digital cough counting tools, which facilitate the longitudinal tracking of cough counts. The second leverages advances in machine learning (ML) and artificial intelligence (AI) to develop sophisticated algorithms capable of extracting important audio features from cough sounds, with the intended clinical utility of identifying underlying pathologies. **Chapter 3, Manuscript II** presents a comprehensive scoping review of contemporary digital cough counting tools and their clinical applications. In the present section, I focus on the latter modality, offering a critical examination of the literature pertaining to cough-based AI technologies for disease screening.

2.4.1. Biological plausibility of disease-specific cough sounds

Radiological evidence suggests that biological differences in lung pathologies for various diseases can be discerned through imaging techniques.[64–67] COVID-19, for instance, exhibits unique pulmonary pathologies that could potentially influence the acoustic properties of cough sounds.[68–70] Similarly, TB presents distinct features in radiological images of affected lungs, which may impact the acoustic characteristics of TB-related coughing.[71–73] Further studies investigating cough acoustic differences have indicated that COVID-19 coughs and non-COVID-19 coughs possess distinct acoustic features.[74,75] A more recent study identified "spectral fingerprints" in cough sounds that correlate with different respiratory diseases.[76] In this study, asthmatic cough sounds

corresponded to frequency bands of 100-800 Hertz (Hz), compared to 1400-2100 Hz for pneumonia coughs. These differences have also been captured in cough spectrograms (**Figure 2.6**),[77] which provide a visual representation of the acoustic properties of coughs over time. A spectrogram presents as a heatmap with three dimensions: 1) time on the x-axis representing the duration of the cough, 2) frequency in Hz on the y-axis displaying the frequencies present in the sound, and 3) intensity as the colour, which indicates the energy of the signal at each time-frequency point. This is illustrated in **Figure 2.6**, where three cough spectrograms are presented, showing the cough from a healthy, COVID-19 positive, and TB-positive person. These spectrograms allow for the visualization and comparison of the unique acoustic signatures associated with different respiratory conditions.

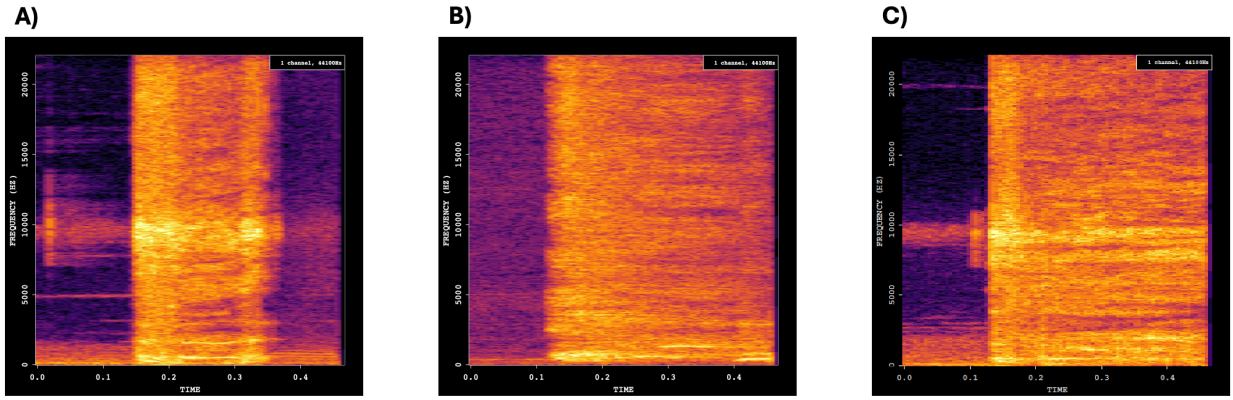


Figure 2.6. Spectrograms of cough sounds from A) a healthy individual, B) a COVID-19 positive patient, and C) a TB-positive patient. Time is represented on the x-axis (0-0.4 seconds), frequency on the y-axis (0-2000 Hz), and intensity by color (purple for low, yellow for high). A) The healthy cough shows a brief, intense burst of energy across a wide frequency range, followed by a quick return to baseline. B) The COVID-19 cough exhibits a more prolonged energy distribution with less distinct onset and offset, and a more uniform intensity across frequencies. C) The TB cough has strong initial burst of energy followed by a prolonged period of energy distribution, particularly in the lower to mid-frequency ranges. The energy appears to be more striated, with horizontal bands of higher intensity alternating with areas of lower intensity, especially in the lower half of the frequency range.

2.4.2 Digital cough datasets

The development of AI algorithms capable of detecting disease-specific signals in cough sounds necessitates large, annotated datasets of cough recordings. This requirement presents a fundamental challenge in AI development: the trade-off between dataset size and quality. In the context of cough sound analysis, most efforts have prioritized quantity, resulting in large-scale, crowdsourced cough datasets. **Table 2.1** provides a comprehensive summary of published cough datasets. These datasets predominantly emerged during the peak months of the pandemic, reflecting the urgent need for rapid diagnostic tools for COVID-19. A more recent dataset—COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM—is a publicly available cough dataset for TB.[78] CODA TB DREAM and the Virufy South Asia study were the only datasets that were not crowdsourced, but instead prospectively enrolled and recorded patient coughs in health facilities.

Among the COVID-19 datasets, the COUGHVID, Coswara, and University of Cambridge are the most widely utilized due to their extensive sample sizes.[74,79,80] The data collection methodology for these datasets involved participants recording their coughs via website applications, using personal smartphones, or computers in home environments. Importantly, the disease status (e.g., COVID-19 positive or negative) in these datasets was self-reported by participants.

The crowdsourcing approach to data collection, while facilitating the rapid accumulation of large sample sizes, raises significant concerns regarding data quality and reliability of the assigned diagnostic labels.[81] Another key issue is selection bias, a common challenge in crowdsourcing and big data methodologies.[82] The recruitment of participants predominantly through social media platforms resulted in datasets skewed towards younger demographics and individuals self-reporting as asymptomatic or COVID-19 negative. This selection bias potentially undermines the representativeness of the data, particularly among people who would be seeking care at health facilities for their cough. Geographical representativeness also becomes a concern. Many of these datasets position themselves for the generation of globally applicable AI algorithms, and yet they collected data from geographically narrow populations. The Coswara dataset obtained 91% of its cough sounds from individuals in India.[80] COUGHVID reported collected cough data from 125

countries, however the proportional distribution of individuals across these countries was not reported.[79]

Another critical concern is the misclassification of outcomes. While some participants self-reported PCR-confirmed results, the majority of studies relied on self-reported symptomatology for COVID-19 classification. The reliability of self-reported health status is often questionable, as individuals may misinterpret their symptoms or inaccurately assess their COVID-19 status. This subjectivity introduces a significant potential for error in the dataset label. Additionally, the narrow focus on COVID-19 in these datasets overlooked the potential presence of other circulating pathogens that may have been causing the cough. Consequently, it is highly probable that many of these datasets are merely distinguishing between "healthy" and "unhealthy" coughs, rather than specifically identifying COVID-19-related coughs.

Variability due to different recording devices used may also introduce a confounding effect. This issue was highlighted in the University of Cambridge dataset, where researchers uncovered a substantial bias related to inconsistent recording quality.[83] Notably, low sample rate recordings (below 12 kHz) were found to be disproportionately associated with COVID-19 positive cases. This unintended correlation led to artificially inflated classification scores, as the models were able to exploit the sample rate bias rather than relying solely on the acoustic characteristics of the coughs themselves.

Table 2.1. Overview of publicly available cough datasets.

Dataset	Collection period	Data collection	Recording platform	Disease	Label	N coughs	N people	Other data
CODA TB DREAM[78]	NR	Clinical study	Hyfe smartphone app.	TB	Composite microbiological reference (PCR, culture)	733,756	2,143	Age, sex, country, BMI, HIV, cough duration, prior TB, symptoms
Coswara[80]	Apr. 2020-Feb. 2022	Crowdsource	Web app.	COVID-19	Self-reported	23,700	2,635	Age, sex, country, smoking, symptoms, respiratory conditions, comorbidities
COUGHVID[79]	Apr. 2020-Dec. 2020	Crowdsource	Web app.	COVID-19	Self-reported & expert-labeled	>25,000	NR	Age, sex, fever, muscle pain, respiratory conditions
IATos[84]	Aug. 2020-Dec. 2020	Crowdsource	Smartphone	COVID-19	PCR	NR	2,281	Age, sex, symptoms
NoCoCoDa[85]	Mar. 2020-Apr. 2020*	Crowdsource	Online interviews*	COVID-19	Self-reported	73 (positive)	11 (positive)	Age (estimated), sex
smarty4covid[86]	Available as of Jan. 2022	Crowdsource	smarty4covid web app.	COVID-19	Self-reported & expert labeled	4,676	NR	Age, sex, COVID vaccination, medical history, vital signs, symptoms, smoking, emotional state, working conditions
University of Cambridge[74]	Up to May 2020	Crowdsource	Web/Android app.	COVID-19	Self-reported	9,986	2,261	Age, sex, symptoms, medical history
Virufy Latin America[87]	NR	Crowdsource	Virufy mobile app.	COVID-19	PCR, antibody testing	NR	31	Age, sex, smoking, symptoms, comorbidities
Virufy South Asia[87]	Apr. 2020-May 2020	Clinical study	Virufy mobile app.	COVID-19	PCR	NR	425	Age, sex, smoking, symptoms, comorbidities

*The authors analyzed cough sounds from COVID-19 positive patients through publicly available media interviews and recordings.
BMI, body mass index; NR, not reported; PCR, polymerase chain reaction; TB, tuberculosis

2.4.3 Artificial intelligence-based cough screening algorithms

AI algorithms for cough screening have gained significant attention in recent years. These algorithms, commonly referred to as "cough classification" models, aim to differentiate between "disease-positive" and "disease-negative" cough sounds. This approach is distinct from "cough detection" algorithms, which focus on distinguishing cough sounds from non-cough sounds.[88,89] The majority of cough classification algorithms have centered on COVID-19 detection, largely due to the availability of extensive, crowdsourced datasets labeled with COVID-19 status (**Table 2.1**).

Table 2.2 presents a summary of published COVID-19 cough classification algorithms. Initial interpretation of the elevated performance metrics suggests that these algorithms demonstrate remarkable efficacy in classifying COVID-19 coughs, with many studies reporting accuracy, sensitivity, and specificity values exceeding 90%. However, these seemingly impressive results warrant scrutiny as they may be artifacts of poor study design and data quality issues rather than true indicators of the algorithms' diagnostic capabilities.

Most of these studies utilized publicly available crowdsourced cough datasets, which are inherently biased, as previously discussed. These datasets were employed in public challenges, such as DiCOVA (using the Coswara dataset) and ComParE COVID-19 (using the University of Cambridge dataset), where researchers competed to develop the most effective COVID-19 cough classification algorithms.[83,90] However, few studies used primary cough data with robust reference standard testing. Notably, Bagad et al. conducted the only study to prospectively collect cough sounds from individuals who underwent polymerase chain reaction (PCR) testing.[91] While Imran et al. reported collecting their own cough data from confirmed COVID-19 cases, the confirmation method was not specified.[92] Ponomarchuk et al. and Tena et al. adopted a hybrid approach, supplementing crowdsourced cough sounds with a small sample of PCR-validated cough samples.[93,94]

Several methodological challenges inherent in the collected data and its utilization in algorithm development warrant consideration. A common issue in AI datasets is class imbalance. This

imbalance was observed across various crowdsourced datasets, manifesting in the underrepresentation of COVID-19 positive cough sounds and COVID-19 positive subjects. When datasets were split between training and test sets, or in different folds of cross-validation (CV), it was not always clear whether the split was performed at the individual or cough level. Best practices dictate that all cough sounds belonging to the same individual should remain in the same set or CV fold to ensure that samples from each individual are non-overlapping. However, some studies simply reported splitting cough sounds without indication that grouping effect was accounted for,[92,94–102] potentially leading to overestimation of performance metrics as the algorithm may recognize individual-level audio signatures and link them to their labels.

The comparison of algorithm performance across studies is further complicated by inconsistent reporting of performance metrics. While some studies provided a comprehensive evaluation including sensitivity, specificity, area under the curve (AUC), and F1-score, many reported only a selection of metrics. Often, the only metric provided was the accuracy of the model,[97,103–108] defined as the proportion of correct predictions among the total number of cases evaluated. However, accuracy alone can be misleading, particularly in the presence of class imbalance, as it can be artificially inflated by the model's performance on the majority class while masking poor detection of the minority class.[109] For instance, in a dataset where only 10% of samples are COVID-19 positive, a model that indiscriminately classifies all cases as negative would achieve 90% accuracy despite failing to identify any positive cases. It is therefore important for studies to report on different metrics, including those that address imbalance such as the F1-score, which provides a balanced measure of precision and recall by calculating their harmonic mean, thus accounting for both false positives and false negatives in imbalanced datasets.

Another notable limitation of these models is the infrequent evaluation using independent, external test sets. A truly external test set would represent a population whose cough sounds were not involved in the initial model training and validation. Only three studies reported testing their algorithms in this manner, although the datasets used were crowdsourced and relatively small.[94,106,110] This highlights the need for more robust external validation in future studies to ensure the generalizability and transferability of cough classification algorithms.

Table 2.2. Overview of COVID cough classification artificial intelligence algorithms.

			Sample size			Performance*	
Author (year)	**Training Dataset**	**Algorithm architecture***	**Training & validation**	**Internal test**	**External test**	**Internal**	**External**
Akgun (2021)[103]	Cambridge	CNN	NR	NR	Not done	Acc.: 86%	Not done
Akman (2022)[111]	Coswara	ResNet	1,040 coughs Pos.: 75	234 coughs Pos.: blind	Not done	AUC: 80%	Not done
Alaaeldein (2022)[108]	Coswara	CNN	NR	NR	Not done	Acc.: 82% AUC: 81%	Not done
Awais (2023)[112]	COUGHVID	Ensemble	NR	NR	Not done	Acc.: 98% Sens.: 98% Spec.: 98% Prec.: 98% F1: 98%	Not done
Bagad (2020)[91]	Own (clinical)	CNN	NR	Not done	Not done	(Validation) AUC: 72%	Not done
Banerjee (2021)[75]	Coswara + Coughvid	ResNet	21,112 coughs Pos.: 1,085	234 coughs Pos.: blind	Not done	AUC: 76% Sens.: 80% Spec.: 63%	Not done
Brown (2021)[74]	Cambridge	SVM	10-fold nested CV 86 coughs Pos.: 54 52 subjects Pos.: 23		Not done	AUC: 82% Sens.: 72% Precision: 80%	Not done
Celik (2023)[113]	COUGHVID	CNN	12,865 subjects Pos.: 2,998	3,217 subjects Pos.: 727	Not done	Acc.: 97% AUC: 95% Sens.: 95% Spec.: 95% Prec.: 96% F1: 95%	Not done
	Coswara	CNN	2,161 subjects Pos.: 546	541 subjects Pos.: 132	Not done	Acc.: 99% AUC: 98%	Not done

	COUGHVID + Coswara	CNN	15,026 subjects Pos.: 3,524	3,758 subjects Pos.: 859	Not done	Sens.: 98% Spec.: 98% Prec.: 100% F1: 99% Acc.: 97% AUC: 95% Sens.: 95% Spec.: 95% Prec.: 97% F1: 96%	Not done
Chang (2021)[114]	Coswara	CNN	1,040 coughs Pos.: 75	Not done	Not done	(Validation) AUC: 72%	Not done
	Cambridge	CNN	517 coughs Pos.: 119	208 coughs Pos.: 39	Not done	AUC: 70%	Not done
Chowdhury (2022)[95]	Cambridge + Coswara + Virufy + NoCoCoDa	Random forest	10-fold nested CV 1,599 coughs Pos.: 360		Not done	Acc.: 84% AUC: 83% Sens.: 74% Spec.: 100% Prec.: 100% F1: 85%	Not done
Despotovic (2021)[115]	Own (crowdsource)	Random forest	5-fold CV 496 coughs Pos.: 249 164 subjects Pos.: 92		Not done	Acc.: 89% Sens.: 87% Spec.: 90%	Not done
Elizalde (2021)[116]	Coswara	Random forest	1,090 coughs Pos.: 75	233 coughs Pos.: blind	Not done	AUC: 82% Sens.: 80% Spec.: 73%	Not done
Erdogan (2021)[117]	Virufy	SVM	5-fold CV 1,187 people Pos.: 595	Not done	Not done	Acc.: 98% Sens.: 99% Spec.: 97% Prec.: 97% F1: 98%	Not done

Hamdi (2022)[96]	COUGHVID	CNN-RNN	10-fold CV 11,624 coughs Pos.: 2,666	Not done	Not done	Acc.: 91% Sens.: 91% Spec.: 92% F1: 91% AUC: 91%	Not done
Haritaoglu (2022)[7]	COUGHVID + Coswara + Virufy + IATOS	CNN	12,823 coughs Pos.: 2,709	3,184 Pos.: 676	Not done	AUC: 78%	Not done
Hemdan (2023)[107]	Coswara	KNN	75%	25%	Not done	Acc.: 97%	Not done
Hoang (2022)[118]	Coswara	Gradient Boosting	965 coughs Pos.: 172	471 coughs Pos.: blind	Not done	Sens.: 48% Spec.: 95% Prec.: 59% F1: 53% AUC: 81%	Not done
Imran (2020)[92]	Own (clinical)	CNN	5-fold CV 543 coughs Pos.: 70		Not done	Acc.: 93% Sens.: 95% Spec.: 91% Prec.: 91% F1: 93%	Not done
Islam (2022)[95]	Virufy	DNN	5-fold CV 16 people Pos.: 7		Not done	Acc.: 94%	Not done
Kamble (2021)[119]	Coswara	Gradient Boosting	1,040 coughs Pos.: 75	233 coughs Pos.: blind	Not done	Sens. :80% Spec.: 54% AUC: 76%	Not done
Kapoor (2023)[97]	Virufy	MLP	60%	40%	Not done	Acc.: 97%	Not done
Laguarta (2020)[120]	Own (crowdsource)	CNN	4,256 subjects	1,064 subjects	Not done	AUC.:97%	Not done
Lella (2021)[121]	Cambridge	CNN	NR	NR	Not done	Acc.: 93% F1: 97%	Not done

Liu (2021)[122]	Cambridge	CNN	699 coughs Pos.: 262 218 subjects Pos.: 35	274 coughs Pos.: 109 54 subjects Pos.: 27	Not done	Acc.: 74% Sens.: 73% Prec.: 74% F1: 73%	Not done
Loey (2021)[123]	COUGHVID	CNN	NR	NR	Not done	Acc.: 95% Sens.: 94% Spec.: 95% F1: 95%	Not done
Manshouri (2022)[6]	Virufy	SVM	CV repeated 1000 times 121 coughs Pos.: 48 16 subjects Pos.: 7		Not done	Acc.: 96% Sens.: 99% Spec.: 91%	Not done
Melek (2021)[98]	Virufy + NoCoCoDa	KNN	Leave-one-out CV 180 coughs Pos.: 107		Not done	Acc.: 98% Sens.: 100% Spec.: 97% F1: 98% AUC: 99%	Not done
Mitrofanova (2021)[124]	Own (clinical)	CNN	300 coughs	48 coughs Pos.: 15	Not done	Acc.: 85% Sens.: 73% Prec.: 79% F1: 76%	Not done
Mohammed (2021)[125]	Virufy + Coswara	CNN	1,020 coughs Pos.: 510	256 coughs Pos.: 128	Not done	Sens.: 71% Prec.: 80% F1: 75% AUC: 77%	Not done
Mouawad (2021)[126]	Own (crowdsource)	XGBoost	NR	NR	Not done	Acc.: 97% Prec.: 78% F1: 91% AUC: 84%	Not done
Najaran (2023)[105]	COUGHVID	CNN	4-fold CV 8,405 coughs Pos.: 658		Not done	Acc.: 71%	Not done

Nasab (2023)[127]	Own (crowdsource)	RNN	NR	NR	Not done	Acc.: 95% Sens.: 96% Spec.: 94% Prec.: 95% F1: 95%	Not done
Nguyen (2023)[99]	Covid-19 Cough	CNN	5,738 coughs Pos.: 663	1,627 coughs	Not done	AUC: 93%	Not done
Nguyen-Trong (2023)[100]	Coswara	CNN	925 coughs	163 coughs	Not done	Acc.: 88% Sens.: 100% Prec.: 89% F1: 88% AUC: 93%	Not done
Pahar (2021)[110]	Coswara	RNN	Nested leave-*p*-out CV 1,171 subjects		Own (crowdsource) 44 people Pos.: 18	Acc.: 95% Sens.: 91% Spec.: 97% AUC: 94%	Acc.: 93% Sens.: 91% Spec.: 96% AUC: 94%
Ponomarchuk (2022)[94]	Covid19-Cough + Own (clinical)	CNN	10-fold CV 1,535 coughs Pos.: 910		Own (crowdsource)	AUC: 81%	AUC: 62%
Rahman (2022)[128]	Cambridge	CNN	Nested 5-fold CV 4,209 coughs Pos.: 1,996		Not done	Acc.: 97% Sens.: 96% Spec.: 95% Prec.: 96% F1: 96%	Not done
Ren (2022)[101]	COUGHVID	Logistic regression	5-fold CV 1,231 subjects Pos.: 210		Not done	Sens.: 63% AUC: 67%	Not done
Schuller (2020)[129]	Cambridge	CNN	Nested 3-fold CV 684 coughs	350 coughs	Not done	Sens.: 74%	Not done

Sobahi (2022)[130]	COUGHVID	CNN	5-fold CV NR		Not done	Acc.: 98% Sens.: 97% Spec.: 99% F1: 98%	Not done
	Virufy	CNN	5-fold CV 1,187 subjects Pos.: 595		Not done	Acc.: 98% Sens.: 97% Spec.: 98% F1: 98%	Not done
	Coswara	CNN	5-fold CV NR		Not done	Acc.: 98% Sens: 88% Spec.: 99% F1: 89%	Not done
Sunitha (2021)[102]	COUGHVID	CNN	2200 coughs Pos.: 1100	200 coughs Pos.: 100	Not done	Acc.: 78% Sens.: 77% Prec.: 76% F1: 76%	Not done
Tena (2022)[93]	Cambridge + Coswara + Virufy + Own (clinical)	Random forest	10-fold CV 813 subjects Pos.: 346		Not done	Acc.: 90% Sens.: 94% Spec.: 82% Prec.: 91% F1: 92% AUC: 96%	Not done
Ulukaya (2023)[106]	Coswara + COUGHVID	CNN	3,330 coughs	370 coughs	Virufy 16 subjects Pos.: 9 NoCoCoDa 10 subjects Pos.: 10	Acc.: 75% AUC: 80%	Virufy Acc.: 62% AUC: 73% NoCoCoDa Acc.: 90%

*Best performing model if multiple algorithms were tested.

Acc., accuracy; AUC, area under the curve; CNN, convolutional neural network; DNN, deep neural network; MLP, multi-layer precepton; NR, not reported; Prec., precision; ResNet, residual neural network; RNN, recurrent neural network; Sens., sensitivity; Spec., specificity; SVM, support vector machine; XGBoost, eXtreme Gradient Boosting

2.5. Machine learning and artificial intelligence approaches to cough classification

Detecting disease from cough sounds is typically approached as a binary classification problem (COVID-19 positive vs. COVID-19 negative). ML and AI methods are particularly valuable for this task due to the complex nature of cough audio data, which contains multiple features across time and frequency domains that must be analyzed simultaneously for accurate classification.

As illustrated in **Table 2.2**, different AI/ML approaches have been used for cough classification. These can be broadly categorized into four main groups: 1) linear ML, 2) non-linear ML, 3) ensemble methods, and 4) deep learning. Linear ML algorithms, such as Logistic Regression, assume that simple linear combinations of features can be used to predict the probability of an outcome.[131] These algorithms enforce a decision boundary that is a straight line in two dimensions or a hyperplane in higher dimensions. Linear ML methods typically offer high interpretability, though they may underperform when dealing with complex, non-linear patterns in data. Non-linear ML algorithms, such as k-Nearest Neighbors and Decision Trees, can capture more complex relationships in the data as they don't assume linear relationships between features and outcomes.[132] These algorithms create flexible decision boundaries that adapt to intricate data patterns. They offer greater flexibility and can identify subtle relationships, though at the cost of higher computational demands and potential overfitting.[133] Ensemble methods operate on the principle that combining multiple learning algorithms produces better predictive performance than could be obtained from any single learning algorithm alone.[134] These come in two main forms: Bagging (like Random Forests), which train multiple models on different subsets of the data and averaging their predictions, and Boosting (like XGBoost), which builds models sequentially, with each model trying to correct the errors of previous ones. Traditional ML approaches (linear, non-linear, and ensemble methods) typically rely on manually engineered features extracted from the audio signals and presented in tabular format.

Deep learning methods, particularly Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs), represent a fundamentally different approach.[135,136] These models are capable of automatically learning hierarchical features from raw or minimally processed data, such as spectrograms of cough sounds (**Figure 2.6**). This automatic feature learning ability makes them

particularly powerful when large amounts of training data are available. In cough classification, CNNs have become particularly popular as they are well-suited for processing spectrograms, effectively learning relevant patterns across both time and frequency domains.

While deep learning approaches have shown promising results, studies continue to demonstrate the effectiveness of traditional machine learning methods such as logistic regression, k-nearest neighbor (kNN), support vector machine (SVM), XGBoost, and Random Forest in cough classification tasks.[74,95,115–117,107,6,110,93]. The choice of method often depends on factors such as dataset size, available computational resources, and the specific requirements of the classification task.

2.6. External validation of artificial intelligence models

The development and validation of AI algorithms typically follows a structured process. Internal validation involves testing the model on data from the same population used for development, often using methods such as cross-validation or train-test splits. External validation (generalizability) assesses the model's performance on new, independent data from the same target population but collected from different sites or time periods than the development data. Another form of external validation (transferability) evaluates the model's performance using datasets from distinct populations to assess its applicability across different settings or contexts. The achievement of robust performance across external validation studies is often considered a crucial benchmark for algorithm deployment readiness, particularly in healthcare settings.[137]

This aspiration towards developing globally acceptable models is evident in various domains, including cough classification algorithms, where developers frequently imply that their algorithms possess universal applicability across diverse populations and settings despite being trained on a narrowly defined dataset.[91,92,113,126,128] The assumption of universality in cough sound features across populations underpins many of these algorithms. Yet, this presumption remains largely uninvestigated.

In recent years, there has been a shift in AI development across different fields, with experts increasingly questioning the pursuit of universally generalizable algorithms. Instead, there is growing advocacy for hyper-local algorithms that excel in specific populations but may not necessarily perform optimally across diverse groups.[138] An illustration of this shift is the Epic Sepsis Model, an AI prediction tool designed to alert clinicians to patients at risk of developing sepsis. Despite its ambitious goals, the model demonstrated poor performance in external validation across different hospitals, achieving a mere 33% sensitivity.[139] In response to these suboptimal results, the developers opted to fine-tune separate models tailored to each specific hospital's needs and characteristics.[140]

This evolving perspective in AI development highlights the complex interplay between algorithm performance, population diversity, and the practical implications of deployment in varied settings. In the context of cough classification, there is a need for a more nuanced approach to AI validation, one that acknowledges population-specific acoustic properties and focuses on optimizing performance for well-defined populations or contexts.

Chapter 3. Manuscript I: Making cough count in tuberculosis care

3.1. Preface

This perspective article, written at the outset of my PhD, marks the beginning of my exploration into the use of cough as a non-traditional biomarker in clinical decision-making for respiratory diseases. At the time of writing, the field of acoustic epidemiology was rapidly evolving. Longitudinal cough counting tools were available and were being investigated for their potential in disease monitoring. Simultaneously, the COVID-19 pandemic had accelerated research into artificial intelligence AI algorithms for cough-based respiratory disease screening.

Using TB as a case study, we summarized the current state of cough monitoring in TB care and explored the potential of cough as a biomarker for enhancing patient care and public health outcomes along the TB care cascade. We also discussed the necessary steps to advance the field of acoustic epidemiology in TB management.

The insights gained from this initial exploration laid the foundation for my subsequent PhD manuscripts.

This perspective article was published in Nature *Communications Medicine* in July 2022.

3.2. Title page

Making cough count in tuberculosis care

Alexandra J. Zimmer[1,2], César Ugarte-Gil[3,4], Rahul Pathri[5], Puneet Dewan[6], Devan Jaganath[7,8], Adithya Cattamanchi[7,8], Madhukar Pai[1,2], Simon Grandjean Lapierre[2,9,10]

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

[2] McGill International TB Centre, Montreal, Canada

[3] School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

[4] Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, Peru

[5] Docturnal Pvt. Ltd., R&D, Hyderabad, India

[6] Bill & Melinda Gates Foundation, Seattle, USA

[7] Department of Medicine, Division of Pulmonary & Critical Care Medicine, University of California, San Francisco, 1001 Potrero Avenue, San Francisco, CA 94110 USA

[8] Center for Tuberculosis, University of California, San Francisco, 1001 Potrero Avenue, San Francisco, CA 94110 USA

[9] Immunopathology Axis, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, 900 Rue Saint-Denis, Montréal, Québec, Canada

[10] Department of Microbiology, Infectious Diseases and Immunology, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec, Canada

3.3. Abstract

Cough assessment is central to clinical management of respiratory diseases, including tuberculosis (TB), but strategies to objectively and unobtrusively measure cough are lacking. Acoustic epidemiology is an emerging field, using technology to detect cough sounds and analyze cough patterns to improve health outcomes among people with respiratory conditions linked to cough. This field is increasingly exploring the potential of artificial intelligence (AI) for more advanced applications, such as analyzing cough sounds as a biomarker for disease screening. While much of the data are preliminary, objective cough assessment could potentially transform disease control programs, including for TB, and support individual patient management. Here, we present an overview of recent advances in this field and describe how cough assessment, if validated, could support public health programs at various stages of the TB care cascade.

3.4. Introduction

Prior to the COVID-19 pandemic, tuberculosis (TB) was the leading infectious cause of mortality, resulting in approximately 10.0 million new infections and 1.4 million deaths worldwide in 2019.[1] The COVID-19 pandemic and lockdowns have had a devastating impact on TB programs globally, as resources and tools used to diagnose and manage TB were diverted to COVID-19.[2] To restore progress and mitigate the impact of COVID-19 on TB management, it is essential to leverage new technologies and innovations to improve TB prevention and care.

TB is an infectious disease caused by the inhalation of droplets containing the bacteria *Mycobacterium tuberculosis* (*Mtb*).[3] TB varies in presentation, ranging from asymptomatic, non-transmissible TB infection (also known as latent TB infection) to symptomatic, contagious active TB disease.[4] Between these two extremes are subclinical forms of TB, where people are considered asymptomatic but may transmit TB to others.[4]

While active TB disease most commonly affects the lungs (pulmonary TB), approximately 15-20% of active TB occurs in other parts of the body, including lymph node TB, abdominal TB, TB meningitis, ocular TB, and neurological TB, to name a few.[5] The occurrence of TB in the body other than the lung is known as extrapulmonary TB (EPTB).[3] Active pulmonary TB is most commonly diagnosed by microbiological testing on mucus from the lung (sputum) samples. Sputum culture is the gold standard for TB testing. However, it is expensive, slow, and requires access to centralized biosafety laboratories.[6] Sputum smear microscopy is often used in primary care facilities in lower-resource settings as a cheaper alternative, but has low sensitivity and is not able to detect drug-resistance.[7] In recent years, more advanced molecular platforms (e.g., GeneXpert PCR machines) have been scaled up as smear-replacement tools that offer greater sensitivity and quicker turnaround times for TB diagnosis.[8,9] Culture, smear microscopy, and GeneXpert are commonly used as reference standards when evaluating the performance and accuracy of newer diagnostics. While active TB is curable, the long regimens (6 months for drug-susceptible TB) and adverse events caused by the antibiotics used, complicate treatment and increase the risk of drug-resistance emerging.[10,11]

As coughing is a common TB symptom, it can be used to screen for TB and assess effectiveness of treatment. This Perspective discusses advances in acoustic epidemiology and AI-based methods to assess cough and how these can be used during TB diagnosis and treatment.

3.5. Using cough as an objective biomarker for TB control and care

Cough is a complex physiological phenomenon as it is both a symptom of, and a defense mechanism against, respiratory diseases. Cough is a hallmark symptom of pulmonary TB and is clinically assessed throughout the cascade of TB care, for example, as a triage tool to trigger TB testing or to monitor response to therapy. Cough patterns vary depending on the amount of *Mtb* in the lungs, and cough tends to regress with successful TB therapy.[12-15]

While many TB screening programs use cough duration and symptoms to determine when TB testing is required, this symptom screening approach lacks sensitivity. In low-resource settings, peripheral health centers, and communities, triage tools such as chest X-rays are not available, thus symptom-based screening remains the only available strategy to identify people with TB. The World Health Organization (WHO) recommends testing people reporting symptoms compatible with TB, including prolonged cough (usually interpreted as a cough that lasts two weeks or longer).[16] According to the 2021 WHO TB screening guidelines, the sensitivity of prolonged cough alone is 42% among HIV-negative individuals, well below the WHO community-based triage test target product profile (TPP) of ≥90% sensitivity.[16,17]

It is difficult for people to describe their cough symptoms, and it is as challenging for clinicians to identify the cause. Individuals tend to have poor recall of the duration of their symptoms, and symptom severity is subjective.[18,19] Given our current inability to objectively detect and monitor cough sounds, patients and providers systematically reduce this data-rich symptom into subjective and dichotomous information (e.g., cough versus no cough, chronic versus acute, getting better versus getting worse), precluding rigorous understanding of cough data, and preventing the use of cough to its full clinical potential. By making cough an objective and measurable component of TB care, either by helping individuals recognize abnormal cough patterns, or by harnessing artificial intelligence (AI) technology (using computer systems to recognize and interpret the

implications of a cough sound)[20] to differentiate types of coughs, we can potentially improve patient management and clinical outcomes at different stages during the cascade of TB care.

3.6. Advances in acoustics for objective cough monitoring

Questionnaire-based tools and scales have been used to collect and evaluate the severity of coughs of varying etiology in an attempt to transform subjective cough reporting into objective data. Such tools include the visual analog scale (VAS), cough symptom score (CSS), and cough diaries.[21] Both the VAS and CSS attempt to quantify the severity of cough based on a patient's perception of their cough. Cough diaries can take various forms, but all depend on patients tracking the frequency and severity of their coughs over time. Other questionnaires expand their assessment of cough to incorporate questions on health-related quality of life.[21] For example, the Leicester Cough Questionnaire (LCQ) is a validated self-completed questionnaire that measures the quality of life of individuals with a chronic cough, and has previously been used to evaluate cohorts of people with TB undergoing anti-TB therapy.[22–24] While such tools are easy to use and implement in clinical settings, they remain subject to bias related to self-perception of health and attention to symptoms, ultimately limiting their clinical application.

Objectivity in cough analysis is improved when using recording devices and computer-assisted acoustic interpretation algorithms. As early as the 1960s, Loudon and Spohn used tape recorders to record and count the coughs of people with TB at night.[25] Other forms of early ambulatory cough meters involved the integration of audio recording devices and electromyogram electrodes,[26] which simultaneously recorded cough sounds and chest muscle contractions when the patient coughed. In 2006, Paul et al. developed and evaluated a self-contained cough monitor composed of an accelerometer (for measuring cough-related vibrations) that stored data on an attached CompactFlash memory card.[27] This device was attached to the patient's neck in the suprasternal notch (jugular notch) and demonstrated good agreement with coughing seen on video footage. Over the years, more advanced 24h recording devices have been developed. These devices typically have a microphone (e.g. free-field microphone necklace or one that attaches to the patient's lapel), which sends the cough sounds to a digital sound recorder, usually attached at the

hip of the patient.[28] Such recording devices include the Leicester Cough Monitor (LCM), the Cayetano Cough Monitor (CayeCoM), and the VitaloJak.[29–31]

Cough counts and patterns were the first objective markers used to analyze cough severity and variation over time. The LCM, CayeCoM and VitaloJak have all been validated for the measurement of cough frequency.[29–31] The LCM and VitaloJak are currently the most widely used cough monitoring tools, with reported cough detection sensitivities of 91% and >99%, respectively.[28] The LCM uses a largely automated algorithm for detecting cough sounds, requiring operator input for calibrating the device (approximately 5 minutes for every 24 hours of recording).[28] The LCM and the CayeCoM have been used to investigate cough among people with pulmonary TB. Turner et al. used the LCM as part of a cross-sectional survey of cough frequency among people with TB and their contacts.[32] Williams et al. used the LCM to correlate exhaled M. tuberculosis with cough frequency.[15] The CayeCoM has been used in various studies to measure cough frequency among cohorts of people with pulmonary TB undergoing treatment.[12–14,33,34] A summary of studies that use various tools for objective cough monitoring in the context of TB care can be found in **Table S3.1**.

While ambulatory recording devices have enabled continuous recording of cough, many of the devices used to date are bulky and obtrusive. Cough is an obvious and stigmatizing symptom, especially among people with TB, and the COVID-19 pandemic has dramatically heightened this stigmatization.[35] In order to efficiently monitor people with cough, recording strategies must be inconspicuous to avoid adding to the stigmatization of respiratory conditions. Smartphones with cough detection and recording applications provide a more discreet approach to monitoring TB coughs. Several cough recording applications have already been developed, including Hyfe Research, AI4COVID-19, and ResAppDx.[36–38]

3.7. Developments in artificial intelligence allow for rigorous assessment of cough

Advances in machine learning, a subset of AI that enables machines to apply algorithms on available data to automatically "learn" and make autonomous decisions,[20] has given rise to a variety of algorithms for cough monitoring that can be deployed on digital recording devices,

including smartphones (see **Table S3.1** for examples of the types of algorithms used for cough detection and cough classification). This new technology allows the analysis of both the frequency and the nature of cough sounds. For example, some algorithms first transform sound recordings into spectrograms—a visual representation of the frequency, amplitude, and time characteristics of sounds—before running an algorithm on the spectrogram to visually analyze the cough's features (**Figure 3.1**).

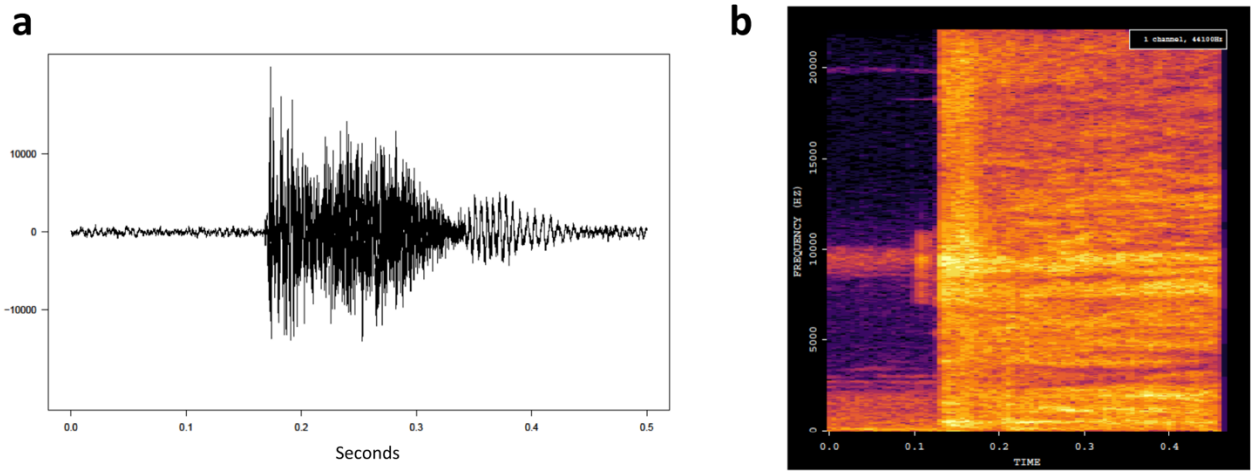


Figure 3.1. Digital cough spectrograms for artificial intelligence algorithm analysis. (a) Waveform image of a pulmonary TB cough. (b) Spectrogram conversion of the waveform cough. On the spectrogram, acoustic information is represented as frequency (y-axis) and amplitude (color) over time (x-axis).

These algorithms are being trained to identify human coughs from ambient sounds (cough detection), as well as to differentiate coughs from patients with distinct clinical conditions or at different stages of disease (cough classification), though the latter use case is yet to be validated.[39–43] Several preliminary cough classification algorithms have been developed for COVID-19 and TB. A classification algorithm was reported to detect COVID-19 infections among people with a cough with 98% sensitivity and 94% specificity, based on a sample of 5,320 individuals (half of whom were COVID-19 positive) and against a reference standard of an "official test" (laboratory method accepted as a diagnosis for COVID-19), doctor assessment, or personal assessment.[42] Another group reported that COVID-19 could be diagnosed using cough with 89% sensitivity and 97% specificity.[37] For TB, TimBre is a screening application that leverages machine learning to

detect TB coughs with a sensitivity of 80% and specificity of 92% against a composite reference standard of sputum smear microscopy, GeneXpert, and chest X-ray, from a sample of 5 bacteriologically-positive and 469 bacteriologically-negative individuals.[44] Another study developed a cough-based screening system that could discriminate cough sounds produced by 16 individuals with TB from those produced by 35 individuals with other lung diseases with 93% sensitivity and 95% specificity against a bacteriological (laboratory method not specified) reference standard, achieving the WHO's TPP requirements of 90% sensitivity and 70% specificity for a community-based TB triage test.[3,45] Botha et al. also developed an AI algorithm for TB cough classification from a sample of 17 people with TB and 21 healthy individuals, achieving an accuracy of 78% and a sensitivity of 95%, at a specificity of 72% against a sputum culture reference standard.[46] These early studies demonstrate that digital cough monitoring, including detection and classification of cough events, could potentially be used to assist TB screening (**Table S3.1**). However, further development and evaluation is critical to move the field forward.

The accuracy of these AI algorithms is contingent on the characteristics of the training dataset. To date, external validation of various AI algorithms has been limited, or has not yet been performed, and the sample sizes used to evaluate these algorithms have been relatively small.[47] Additionally, early diagnostic studies of novel tests, including AI algorithms, tend to overestimate the diagnostic accuracy, mainly because of the preferential exclusion of more complicated cases.[48] Until sufficient replication studies have been completed using large, and diverse cough datasets, representative of different populations, the clinical application of these AI algorithms will remain limited.

3.8. Using digital cough monitoring to change TB care

Digital cough monitoring has the potential to address multiple gaps in the TB cascade of care (**Figure 3.2**).[49] In this section, as an example of the breadth of the potential value of cough data, we outline hypothetical ways in which AI-based cough tools could be used.

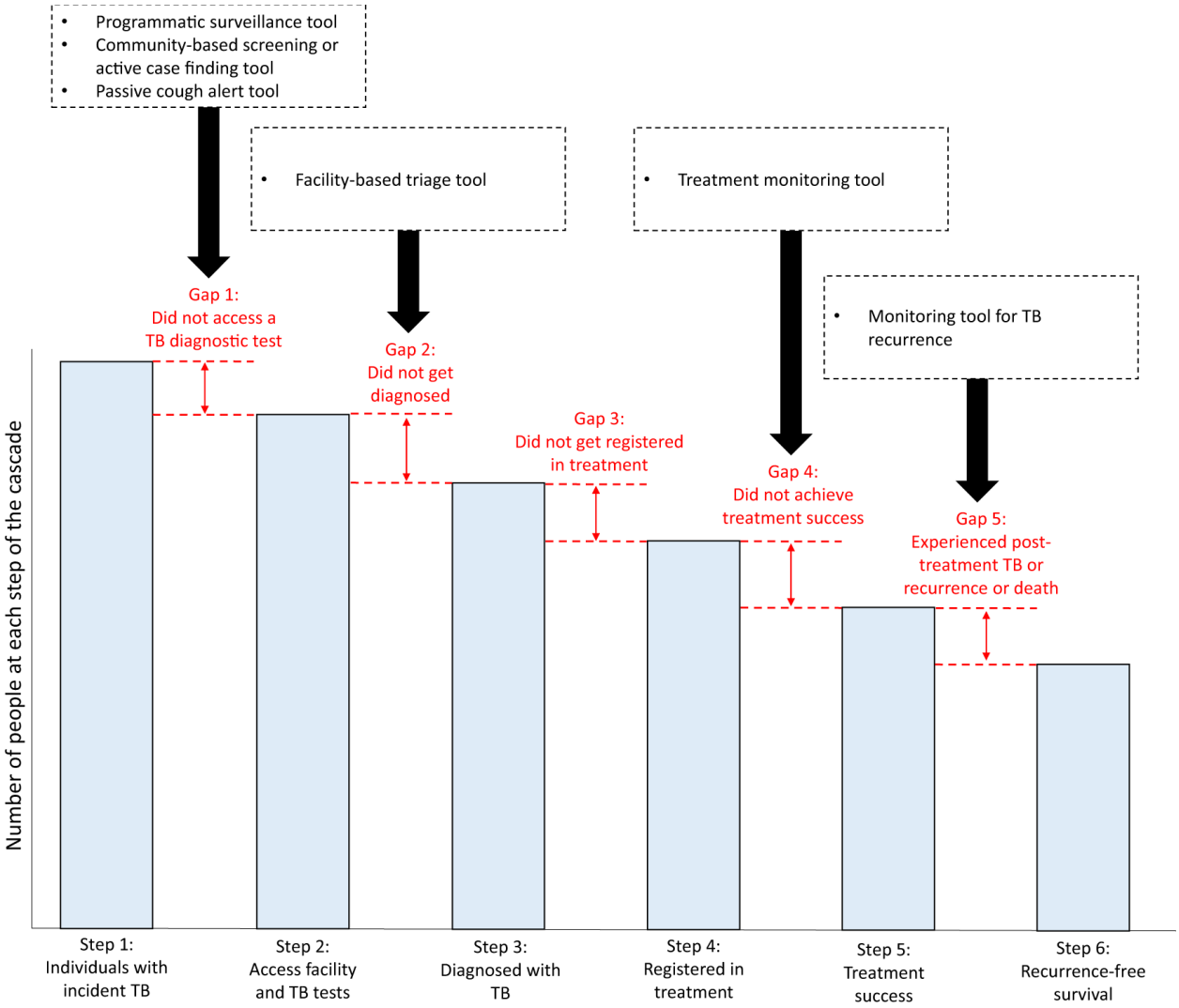


Figure 3.2. Potential use cases for digital cough in the tuberculosis cascade of care. Each step in TB care cascade is represented as a bar. The gaps in the cascade are in red between each step. Boxes pointing at the gaps represent possible digital cough-based solutions to address various gaps. The height of the bar graphs and the length of the gaps are not scaled to represent true values. They are intended to help illustrate the different steps of the care cascade and points at which people with TB may fail to benefit from care. (Cascade of care adapted from Subbaraman et al.[49])

3.8.1. Supporting TB program planning

Finding people with TB, or who have symptoms of TB, requires health systems and TB programs to strategically deploy limited resources. In a syndromic surveillance approach (i.e., detection and aggregation of individuals and populations' health indicators such as symptoms prior to

establishing a definitive diagnosis) both individuals at risk of developing TB, or people who previously had TB, could passively and prospectively monitor their cough. Temporal and geospatial aggregations of cough events could in turn be used to better target case-finding activities and identify high-risk settings. Spatiotemporal changes in cough frequency at population-level can be used as a proxy for incidence of COVID-19, TB or other respiratory diseases.[36] Whether specifically dedicating public health resources to investigate such cough clusters would accelerate the identification of additional prevalent cases and improve disease case notifications needs to be investigated. Restricting this cough surveillance analytic approach by monitoring people previously diagnosed active pulmonary TB could identify cough hotspots where the risk of TB transmission has been, and may still be, even higher.

3.8.2. Improving community-based monitoring and active case finding

Very preliminary data suggests that cough classification algorithms could be developed that meet WHO TPPs for a community-based TB screening test.[44,45] Further validation is needed using cohorts of large sample size and diverse populations before any definite conclusions can be made regarding their sensitivities and specificities. AI-based cough screening could complement other available community-based screening approaches, such as chest X-rays, increasing the number of people with presumed TB appropriately referred to facilities for confirmatory testing in a timely manner. Indeed, using cough to predict chest X-ray abnormalities could trigger radiology testing for which multiple automated interpretation algorithms have now been thoroughly validated.[50] If deployed on mobile devices, AI-based cough screening could allow for low-cost remote active case finding and self-screening, with subsequent referral to a health facility for confirmatory TB testing and linkage to care. The vignette in **Figure 3.3** illustrates how a cough tool may help refer people with a cough to a physician.

Figure 3.3. Example use of smartphone-based cough screening application for community-based monitoring. In this vignette, a female is experiencing symptoms of disease, including

cough. Using a phone with the example Health App (not a real app), she is prompted to cough and report any other symptoms she is experiencing. The AI algorithm in the Heath App uses the information to provide likely causes of disease (in this case, COVID-19 or TB) and refers her to consult a physician for confirmatory testing. (Vignette originally created for The Lancet Citizens' Commission on Reimagining India's Health System, by Raghu Dharmaraju, Vijay Chandru,Umakant Soni, and Shubraneel Ghosh, AI & Robotics Technology Park (ARTPARK) at Indian Institute of Science. "A vignette from 2030 in rural India: How might technology enable citizen-centered health journeys?".)

For individuals at higher risk of developing active pulmonary TB, such as household contacts, cough detection and longitudinal monitoring could objectively document an increase in TB-compatible symptoms, prompting early care-seeking and limiting transmission. This approach could also help address subclinical pulmonary TB.[51] Individuals who have mild symptoms, but do not recognize them as being significant, are also considered subclinical.[51] In such cases, digital cough monitoring could be used to identify the presence or significance of cough that would otherwise have gone unrecognized or unreported. However, digital cough monitoring would not extend to truly asymptomatic individuals with subclinical TB, limiting its application as an active case finding tool in this sub-group. A study of 24 people with TB found that cough frequency may not be associated with *Mtb* output collected on face masks.[15] That is, some participants who did not cough very often still expelled a lot of *Mtb* (and vice versa). While further investigations are needed, this raises potential limitations of relying on cough monitoring for evaluating active case-finding and reducing TB transmission.

3.8.3. Enhancing the performance of diagnostic algorithms

Even when people with presumed TB reach the health facility, it is not guaranteed that they will access proper confirmatory testing. One reason for this is a lack of awareness and training among healthcare workers to recognize key TB symptoms. This problem has been demonstrated by studies involving standardized patients (SPs), healthy persons trained to visit health facilities with fake TB symptoms, without the healthcare providers being aware that these symptoms are not real.[52] A systematic review on SPs in India found that only half of healthcare providers knew that prolonged

cough (>2 weeks) may be associated with TB.[52] Another study in India found that SPs presenting with TB symptoms were severely under-tested.[53]

Similar to community-based screening and active case-finding, health providers may potentially use AI-based cough classification applications to help triage people with presumed TB, complementing less sensitive symptom-based triage methods and increasing the proportion of individuals with presumed TB that undergo confirmatory testing. Because symptom screening is also non-specific, it may also help reduce the proportion of people without TB who unnecessarily undergo TB testing.

3.8.4. Monitoring the effect of treatment

Smartphones are globally available and can act as recording devices. They are already used for TB treatment-adherence monitoring with video Directly Observed Therapy (vDOT), which allows people with TB to send videos of themselves ingesting anti-TB treatment to their health provider, instead of having to travel to the clinic to take their anti-TB treatment in front of a health provider, as required under traditional DOT methods.[54] Given that cough symptoms regress with successful treatment, cough detection applications could be used as a low-cost, person-centric approach for clinicians to remotely monitor people with TB's clinical response to treatment, or even for people to self-monitor their cough as treatment progresses.[13] Objectively-documented unfavorable cough evolution patterns could prompt patients and providers to investigate whether the treatment regimen being used is effective, allowing for early recognition of drug resistance or poor adherence.

3.8.5 Achieving relapse-free cures and minimizing long-term lung damage

A significant proportion of people who are successfully cured of TB are at risk of TB recurrence within the first year following completion of anti-TB treatment.[55] The prospective cough monitoring used during treatment could be continued during this high-risk period to identify early signs of TB recurrence. Even if people do not experience TB recurrence or relapse, they are at increased risk of experiencing post-TB lung damage, an aspect of TB care that is often overlooked

in TB management pathways.[56] Thus, cough monitoring, if validated, could also be useful as a starting point in identifying individuals with post-TB lung disease and related lung function decline.

3.8.6. Supporting drug development and TB research

AI-based cough detection technology could also play a role in TB research and development. Digital cough monitoring could be used as a secondary endpoint in clinical drug development trials. Drug development trials have so far relied on evaluating whether sputum culture test results from the person with TB change from positive to negative during the first 8 weeks of therapy as a proxy for anti-TB treatment efficacy.[57] Such culture methods are resource- and time-intensive, and do not allow the monitoring of intermediate outcomes, including patient symptoms. In addition, regulatory agencies may request data on patient-reported improvement in cough, though again this is subjective and can have variable accuracy.[58,59] Similar to symptom-based screening, self-assessment of cough in the context of experimental therapy-efficacy measurement is unlikely to be fully accurate. Objective monitoring of cough may allow for more nuanced monitoring of intermediate endpoints by acting as a complement to conventional culture-based endpoints and patient-reported outcomes.

3.9. Furthering the clinical use of digital cough monitoring

The recent progress in acoustics and cough analysis, combined with the urgent need to improve respiratory disease detection and tracking methods in the context of COVID-19, have accelerated applications of acoustic epidemiology in clinical research.[37,42,60] This emerging field depends on increasingly less obtrusive ways to collect cough data as well as more sophisticated analytics that go beyond cough detection to infer clinical etiology based on cough patterns and spectral characteristics.

The development, validation, and roll-out of digital cough monitoring tools for TB will require global coordinated data collection, curation, and analysis effort. Training and validation cough datasets need to be collected from people in the intended use population and settings. They must

include large numbers of people with different demographic characteristics (e.g. age, sex, ethnicity) as well as different forms of pulmonary TB in clinical settings with variable background epidemiology of respiratory diseases. This 'big data' approach is mandatory for the development and refinement of AI algorithms to achieve high external validity. Since cough is not specific to TB, such datasets should not be limited to the development of AI algorithms for TB but should also be used to develop and refine cough algorithms for other respiratory diseases and conditions that are linked to cough. To accelerate this endeavour, we must avoid the multiplication of isolated algorithm development efforts that use data from homogeneous patient populations.[47]

Collective efforts to aggregate and annotate cough data may accelerate research and tool development. For example, Global Health Labs, the Bill and Melinda Gates Foundation, and the Patrick J. McGovern Foundation are currently supporting efforts to collect cough data and are investing in infrastructure to build an extensive database of cough sounds. Researchers interested in cough and acoustic epidemiology—in the context of TB or any other respiratory disease or condition linked to cough—can contribute to this growing anonymized database and use the existing data to develop and refine AI. While this effort is an important step towards integrating cough into TB care, there is still a need for a broader recognition of the potential advantages of integrated AI-based cough tools into TB care. As more AI-based cough detection tools and applications become available, increased effort should be made to routinely collect cough data within TB programs, prevalence surveys, and clinical studies in order to contribute to the growing field of acoustic epidemiology. Such efforts will help characterize the natural evolution of TB cough, objectively describe the impact of specific interventions on TB symptoms, and iteratively improve operational and performance characteristics of cough-based TB solutions. Like other biomarkers, collected cough data must be anonymized, annotated with clinical metadata, and shared in open-source repositories. TB cough data must also be made available in the same way that digital chest X-ray libraries are available for the validation of electronic interpretation algorithms, or that TB genomic sequences are available to support novel drug development and validation of drug resistance assays.[61,62] Through such collective efforts, we can accelerate algorithm development and the roll-out of cough-based clinical tools. This data sharing approach should also improve partnerships between academia and industry by allowing faster hypothesis-testing as well as rapid product design and translation into user-friendly tools that can be deployed

at scale in TB care. In conclusion, AI and acoustic epidemiology has the potential to revolutionize the fight against TB.

3.10. Acknowledgements

We would like to thank all people with TB, providers and researchers who have and will contribute acoustic data to the tuberculosis scientific community to make cough count in tuberculosis care. We would also like to thank Raghu Dharmaraju, Vijay Chandru, Umakant Soni, and Shubraneel Ghosh from ARTPARK (AI & Robotics Technology Park) at the Indian Institute of Science for allowing us to use the vignette they developed that illustrates the use of smartphone-based cough detection for community-based screening.

3.11. Contributions

Conceptualization (A.J.Z., S.G.L.) literature article review (A.J.Z., D.J., S.G.L.), cascade of TB care design and assessment of clinical applications (A.J.Z., C.U.G., R.P., P.D., D.J., A.C., M.P., S.G.L.), machine learning perspective and tuberculosis digital cough analysis (R.P.), collaborative cough consortium design and funding acquisition (P.D., A.C., S.G.L.), writing – original draft (A.J.Z., S.G.L.), writing – review and editing (C.U.G., R.P., P.D., D.J., A.C., M.P.), visualization (A.J.Z.).

3.12. References

1. World Health Organization. Global Tuberculosis Report 2020. World Health Organization (WHO); 2020. https://www.who.int/publications/i/item/9789240013131
2. Zimmer AJ, Klinton JS, Oga-Omenka C, et al. Tuberculosis in times of COVID-19. J Epidemiol Community Health. 2022;76(3):310-316. doi:10.1136/jech-2021-217529
3. Pai M, Behr MA, Dowdy D, et al. Tuberculosis. Nat Rev Dis Primers. 2016;2(1):16076. doi:10.1038/nrdp.2016.76
4. Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. Breathe. 2021;17(3):210079. doi:10.1183/20734735.0079-2021
5. Sharma SK, Mohan A, Kohli M. Extrapulmonary tuberculosis. Expert Review of Respiratory Medicine. 2021;15(7):931-948. doi:10.1080/17476348.2021.1927718

6. Vongthilath-Moeung R, Poncet A, Renzi G, Schrenzel J, Janssens JP. Time to Detection of Growth for Mycobacterium tuberculosis in a Low Incidence Area. Front Cell Infect Microbiol. 2021;11:704169. doi:10.3389/fcimb.2021.704169
7. Kik SV, Denkinger CM, Chedore P, Pai M. Replacing smear microscopy for the diagnosis of tuberculosis: what is the market potential? Eur Respir J. 2014;43(6):1793-1796. doi:10.1183/09031936.00217313
8. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Infectious Diseases Group, ed. Cochrane Database of Systematic Reviews. Published online January 21, 2014. doi:10.1002/14651858.CD009593.pub3
9. Acharya B, Acharya A, Gautam S, et al. Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. Mol Biol Rep. 2020;47(5):4065-4075. doi:10.1007/s11033-020-05413-7
10. Nahid P, Dorman SE, Alipanah N, et al. Executive Summary: Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. Clinical Infectious Diseases. 2016;63(7):853-867. doi:10.1093/cid/ciw566
11. Forget EJ, Menzies D. Adverse reactions to first-line antituberculosis drugs. Expert Opinion on Drug Safety. 2006;5(2):231-249. doi:10.1517/14740338.5.2.231
12. Proaño A, Bui DP, López JW, et al. Cough Frequency During Treatment Associated With Baseline Cavitary Volume and Proximity to the Airway in Pulmonary TB. Chest. 2018;153(6):1358-1367. doi:10.1016/j.chest.2018.03.006
13. Proaño A, Bravard MA, López JW, et al. Dynamics of Cough Frequency in Adults Undergoing Treatment for Pulmonary Tuberculosis. Clinical Infectious Diseases. 2017;64(9):1174-1181. doi:10.1093/cid/cix039
14. Lee GO, Comina G, Hernandez-Cordova G, et al. Cough dynamics in adults receiving tuberculosis treatment. Ehtesham HS, ed. PLoS ONE. 2020;15(6):e0231167. doi:10.1371/journal.pone.0231167
15. Williams CM, Abdulwhhab M, Birring SS, et al. Exhaled Mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: prospective observational studies. The Lancet Infectious Diseases. 2020;20(5):607-617. doi:10.1016/S1473-3099(19)30707-8
16. World Health Organization. WHO Operational Handbook on Tuberculosis - Module 2: Systematic Screening for Tuberculosis Disease. WHO; 2022. https://www.who.int/publications/i/item/9789240022614
17. World Health Organization. High-Priority Target Product Profiles for New Tuberculosis Diagnostics.; 2014. https://www.who.int/publications/i/item/WHO-HTM-TB-2014.18
18. Cho PSP, Birring SS, Fletcher HV, Turner RD. Methods of Cough Assessment. The Journal of Allergy and Clinical Immunology: In Practice. 2019;7(6):1715-1723. doi:10.1016/j.jaip.2019.01.049
19. Feikin DR, Audi A, Olack B, et al. Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. International Journal of Epidemiology. 2010;39(2):450-458. doi:10.1093/ije/dyp374

20. Helm JM, Swiergosz AM, Haeberle HS, et al. Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. Curr Rev Musculoskelet Med. 2020;13(1):69-76. doi:10.1007/s12178-020-09600-8
21. Wang Z, Wang M, Wen S, Yu L, Xu X. Types and applications of cough-related questionnaires. J Thorac Dis. 2019;11(10):4379-4388. doi:10.21037/jtd.2019.09.62
22. Suzuki T, Shirai T, Hirai K, et al. Improved cough- and sputum-related quality of life after initiation of treatment in pulmonary tuberculosis. Respiratory Investigation. 2019;57(3):252-259. doi:10.1016/j.resinv.2019.01.002
23. Turner R, Bothamley G, Birring S. P240 Validation of the Leicester Cough Questionnaire in pulmonary tuberculosis. Thorax. 2015;70(Suppl 3):A197.2-A198. doi:10.1136/thoraxjnl-2015-207770.376
24. Birring SS. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). Thorax. 2003;58(4):339-343. doi:10.1136/thorax.58.4.339
25. Loudon RG, Spohn SK. Cough frequency and infectivity in patients with pulmonary tuberculosis. Am Rev Respir Dis. 1969;99(1):109-111. doi:10.1164/arrd.1969.99.1.109
26. Hsu J, Stone R, Logan-Sinclair R, Worsdell M, Busst C, Chung K. Coughing frequency in patients with persistent cough: assessment using a 24 hour ambulatory recorder. Eur Respir J. 1994;7(7):1246-1253. doi:10.1183/09031936.94.07071246
27. Paul IM, Wai K, Jewell SJ, Shaffer ML, Varadan VV. Evaluation of a new self-contained, ambulatory, objective cough monitor. Cough. 2006;2:7. doi:10.1186/1745-9974-2-7
28. Hall JI, Lozano M, Estrada-Petrocelli L, Birring S, Turner R. The present and future of cough counting tools. J Thorac Dis. 2020;12(9):5207-5223. doi:10.21037/jtd-2020-icc-003
29. Smith J, Woodcock A. New Developments in the Objective Assessment of Cough. Lung. 2008;186(S1):48-54. doi:10.1007/s00408-007-9059-1
30. Matos S, Birring SS, Pavord ID, Evans DH. An Automated System for 24-h Monitoring of Cough Frequency: The Leicester Cough Monitor. IEEE Trans Biomed Eng. 2007;54(8):1472-1479. doi:10.1109/TBME.2007.900811
31. Birring SS, Fleming T, Matos S, Raj AA, Evans DH, Pavord ID. The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough. European Respiratory Journal. 2008;31(5):1013-1018. doi:10.1183/09031936.00057407
32. Turner R, Repossi A, Matos S, Birring S, Bothamley G. S79 Cough Prevalence And Frequency In Pulmonary Tuberculosis. Thorax. 2014;69(Suppl 2):A43-A44. doi:10.1136/thoraxjnl-2014-206260.85
33. Tracey BH, Comina G, Larson S, Bravard M, Lopez JW, Gilman RH. Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2011:6017-6020. doi:10.1109/IEMBS.2011.6091487
34. Larson S, Comina G, Gilman RH, Tracey BH, Bravard M, López JW. Validation of an Automated Cough Detection Algorithm for Tracking Recovery of Pulmonary Tuberculosis Patients. Pai M, ed. PLoS ONE. 2012;7(10):e46229. doi:10.1371/journal.pone.0046229

35. Cremers AL, De Laat MM, Kapata N, Gerrets R, Klipstein-Grobusch K, Grobusch MP. Assessing the Consequences of Stigma for Tuberculosis Patients in Urban Zambia. Kumar A, ed. PLoS ONE. 2015;10(3):e0119861. doi:10.1371/journal.pone.0119861
36. Gabaldon-Figueira JC, Brew J, Doré DH, et al. Digital acoustic surveillance for early detection of respiratory disease outbreaks in Spain: a protocol for an observational study. BMJ Open. 2021;11(7):e051278. doi:10.1136/bmjopen-2021-051278
37. Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Informatics in Medicine Unlocked. 2020;20:100378. doi:10.1016/j.imu.2020.100378
38. Moschovis PP, Sampayo EM, Cook A, et al. The diagnosis of respiratory disease in children using a phone-based cough and symptom analysis algorithm: The smartphone recordings of cough sounds 2 (SMARTCOUGH-C 2) trial design. Contemporary Clinical Trials. 2021;101:106278. doi:10.1016/j.cct.2021.106278
39. Bales C, Nabeel M, John CN, et al. Can Machine Learning Be Used to Recognize and Diagnose Coughs? Published online 2020. doi:10.48550/ARXIV.2004.01495
40. Coppock H, Gaskell A, Tzirakis P, Baird A, Jones L, Schuller B. End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. BMJ Innov. 2021;7(2):356-362. doi:10.1136/bmjinnov-2021-000668
41. Nessiem MA, Mohamed MM, Coppock H, Gaskell A, Schuller BW. Detecting COVID-19 from Breathing and Coughing Sounds using Deep Neural Networks. In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2021:183-188. doi:10.1109/CBMS52027.2021.00069
42. Laguarta J, Hueto F, Subirana B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open J Eng Med Biol. 2020;1:275-281. doi:10.1109/OJEMB.2020.3026928
43. Cesnakova L, Boza V, Dubec P, et al. Continuous Sound Collection Using Smartphones and Machine Learning to Measure Cough. Digit Biomark. 2019;3(3):166-175. doi:10.1159/000504666
44. Pathri R, Jha S, Tandon S, GangaShetty S. Acoustic Epidemiology of Pulmonary Tuberculosis (TB) & Covid19 leveraging AI/ML. Published online February 8, 2022. doi:10.1101/2022.02.05.22269707
45. Pahar M, Klopper M, Reeve B, Warren R, Theron G, Niesler T. Automatic cough classification for tuberculosis screening in a real-world environment. Physiol Meas. 2021;42(10):105014. doi:10.1088/1361-6579/ac2fb8
46. Botha GHR, Theron G, Warren RM, et al. Detection of tuberculosis by automatic cough sound analysis. Physiol Meas. 2018;39(4):045005. doi:10.1088/1361-6579/aab6d0
47. Topol EJ. Is my cough COVID-19? The Lancet. 2020;396(10266):1874. doi:10.1016/S0140-6736(20)32589-7
48. Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, Van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ. 2006;174(4):469-476. doi:10.1503/cmaj.050090

49. Subbaraman R, Nathavitharana RR, Mayer KH, et al. Constructing care cascades for active tuberculosis: A strategy for program monitoring and identifying gaps in quality of care. PLoS Med. 2019;16(2):e1002754. doi:10.1371/journal.pmed.1002754
50. Qin ZZ, Ahmed S, Sarker MS, et al. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. The Lancet Digital Health. 2021;3(9):e543-e554. doi:10.1016/S2589-7500(21)00116-3
51. Kendall EA, Shrestha S, Dowdy DW. The Epidemiological Importance of Subclinical Tuberculosis. A Critical Reappraisal. Am J Respir Crit Care Med. 2021;203(2):168-174. doi:10.1164/rccm.202006-2394PP
52. Satyanarayana S, Subbaraman R, Shete P, et al. Quality of tuberculosis care in India: a systematic review. Int J Tuberc Lung Dis. 2015;19(7):751-763. doi:10.5588/ijtld.15.0186
53. Das J, Kwan A, Daniels B, et al. Use of standardised patients to assess quality of tuberculosis care: a pilot, cross-sectional study. The Lancet Infectious Diseases. 2015;15(11):1305-1313. doi:10.1016/S1473-3099(15)00077-8
54. Nguyen TA, Pham MT, Nguyen TL, et al. Video Directly Observed Therapy to support adherence with treatment for tuberculosis in Vietnam: A prospective cohort study. International Journal of Infectious Diseases. 2017;65:85-89. doi:10.1016/j.ijid.2017.09.029
55. Velayutham B, Chadha VK, Singla N, et al. Recurrence of tuberculosis among newly diagnosed sputum positive pulmonary tuberculosis patients treated under the Revised National Tuberculosis Control Programme, India: A multi-centric prospective study. Cardona PJ, ed. PLoS ONE. 2018;13(7):e0200150. doi:10.1371/journal.pone.0200150
56. Meghji J, Simpson H, Squire SB, Mortimer K. A Systematic Review of the Prevalence and Pattern of Imaging Defined Post-TB Lung Disease. Hill PC, ed. PLoS ONE. 2016;11(8):e0161176. doi:10.1371/journal.pone.0161176
57. Davies G, Boeree M, Hermann D, Hoelscher M. Accelerating the transition of new tuberculosis drug combinations from Phase II to Phase III trials: New technologies and innovative designs. PLoS Med. 2019;16(7):e1002851. doi:10.1371/journal.pmed.1002851
58. Cruz Rivera S, McMullan C, Jones L, Kyte D, Slade A, Calvert M. The impact of patient-reported outcome data from clinical trials: perspectives from international stakeholders. J Patient Rep Outcomes. 2020;4(1):51. doi:10.1186/s41687-020-00219-4
59. Insmed Incorporated. A Randomized, Double-Blind, Placebo-Controlled, Active Comparator, Multicenter Study to Validate Patient-Reported Outcome Instruments in Adult Subjects With Newly Diagnosed Nontuberculous Mycobacterial (NTM) Lung Infection Caused by Mycobacterium Avium Complex (MAC). clinicaltrials.gov; 2021. https://clinicaltrials.gov/study/NCT04677543
60. Bagad P, Dalmia A, Doshi J, et al. Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds. Published online 2020. doi:10.48550/ARXIV.2009.08790
61. Jaeger S, Candemir S, Antani S, Wáng YXJ, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg. 2014;4(6):475-477. doi:10.3978/j.issn.2223-4292.2014.11.20
62. The CRyPTIC Consortium and the 100,000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med. 2018;379(15):1403-1415. doi:10.1056/NEJMoa1800474

3.13. Supplementary information

Table S3.1. Summary table of studies investigating cough in the context of TB care, cough detection, and cough classification

Cough recording tool	Author (Year)	Ref.	Sample size	Setting	Reference standard	Study design and objectives	Algorithm description
Objective cough monitoring for TB							
Tape recorder	Loudon (1969)	1	63 TB patients	TB patients hospitalized in Dallas, USA	N/A	Prospective cohort demonstrating that nighttime cough frequency is associated with disease severity	N/A
Leicester Cough Monitor and Visual Analog Scale (VAS)	Turner (2014)	2	108	N/A	Sputum culture, sputum smear microscopy	Retrospective review of medical records of TB patients' cough 24h prior to commencing treatment	N/A
Leicester Cough Questionnaire (LCQ) and Cough and Sputum Assessment Questionnaire (CASA-Q)	Suzuki (2019)	3	85	TB patients hospitalized in Shizuoka, Japan	Sputum culture, sputum smear microscopy	Prospective observational cohort comparing LCQ and CASA-Q score on admission and at discharge	N/A
Leicester Cough Monitor	Turner (2018)	4	44	TB clinics and in-patient facilities in the United Kingdom	Sputum culture, sputum smear microscopy, chest X-ray	Cross-sectional survey of cough frequency in patients with TB and their contacts	N/A

Leicester Cough Monitor	Williams (2020)	5	24	Inpatients admitted to one of three hospitals in Pretoria, South Africa	Liquid sputum culture (BACTEC MGIT 960), sputum smear microscopy, Xpert MTB/RIF, chest X-ray	Prospective cohort over a 24h period to correlate exhaled TB bacillary output with cough frequency	N/A
Cayetano Cough Monitor*	Tracey (2011)	6	62 TB patients	Public national tertiary referral hospital in Lima, Peru	MODS culture, sputum smear microscopy, clinical symptoms	Prospective cohort examining cough pattern and frequency among TB patients prior to treatment and during the first 60 days of treatment.	TB cough detection algorithm using sequential minimal optimization Sensitivity = 81% Specificity = N/A Overall accuracy = 86.4%
Cayetano Cough Monitor*	Larson (2012)	7	15 TB patients	Tertiary referral hospital in Lima, Peru	N/A	Prospective cohort examining the change in cough frequency during the first 2 weeks of TB therapy	TB cough detection algorithm using sequential minimal optimization Sensitivity = 75.5% Specificity = 99.6%
Cayetano Cough Monitor*	Proaño (2017)	8	64 TB patients	Two reference tertiary academic hospitals in Lima, Peru	MODS culture, auramine-stained smear	Prospective cohort study evaluating cough patterns among TB patients prior to treatment and during the first 62 days of treatment.	Same algorithm as Larson (2012)[7]
Cayetano Cough Monitor*	Proaño (2018)	9	41	Two tertiary hospitals in Lima, Peru	MODS culture	Prospective cohort study examining the relationship between cough frequency and cavitary lung disease throughout the first 60 days of TB therapy	N/A

Cayetano Cough Monitor*	Lee (2020)	10	71 TB patients	Two tertiary hospitals in Lima, Peru	MODS culture	Prospective cohort examining the change in cough frequency among patients with TB during the first 60 days of TB therapy	Same algorithm as Larson (2012)[7]
Examples of cough detection AI algorithms							
Hyfe Cough Tracker smartphone app	Gabaldon-Figueira (2021)	11	57 participants	Community members located within 5km of the University of Navarra, Spain	N/A	Prospective observational study to assess the value of digital acoustic surveillance in predicting respiratory disease incidence (including COVID-19)	Cough detection algorithm using Convolutional Neural Network model Sensitivity = 96.34% Specificity 96.54%
AI4COVID-19 smartphone app.	Imran (2020)	12	543 coughs	N/A	N/A	Development of COVID-19 cough detection AI model	Sensitivity = 96.01% Specificity = 95.19%
HealthMode Cough smartphone application	Kvapilova (2019)	13	20 people	Online material (including YouTube videos and SoundSnap website)	Manual cough counting	Development of cough detection AI model	Sensitivity = 90% at 99.5% specificity Sensitivity = 75% at 99.9% specificity
Examples of TB and COVID-19 cough classification AI algorithms							
TimBre smartphone app.	Pathri (2022)	14	# people TB: 5 Non-TB: 469	Tertiary hospital in Bangalore, India	Sputum smear microscopy, Xpert (unspecified), chest X-ray	Development of TB cough classification AI model	TB cough classification algorithm using RUS Boosted Algorithm Sensitivity = 80% Specificity = 92%
Tascam DR-44WL hand-held audio recorder and a Rhode M3 microphone	Botha (2018)	15	# people TB: 17 Non-TB: 21	Recording done in a “specially designed facility”	Sputum culture	Development of TB cough classification AI model	TB cough classification algorithms using fusion by logistic regression Sensitivity = 95% at 72% specificity

ZOOM F8N field recorder and a RØDE M3 condenser microphone	Pahar (2021)	16	# people TB: 16 Control: 35	Primary health care clinic in Cape Town, South Africa	"Bacteriological TB diagnosis"	Development of TB cough classification AI models	TB cough classification algorithms using logistic regression (LR), support vector machines (SMV), k-nearest neighbor (KNN), multilayer perceptron's (ML), and convolutional neural networks (CNN) LR performed best: Sensitivity = 93% Specificity = 95%
AI4COVID-19 smartphone app.	Imran (2020)	12	# coughs COVID: 70 Non-COVID: 473	N/A	N/A	Development of COVID-19 cough classification AI model	COVID-19 cough classification algorithms using Deep Transfer Learning-based Multi-Class (DTL-MC), Classical Machin Learning-based Multi Class (CML-MC), Deep Transfer Learning-based Binary Class (DTL-BC) classifiers DTL-MC Sensitivity = 89.14% Specificity = 96.67% CML-MC Sensitivity = 91.71% Specificity = 95.27% DTL-BC Sensitivity = 94.57% Specificity = 91.14%

MIT Open Voice Initiative website (opensigma.mit.edu)	Laguarta (2020)	17	# people COVID: 2660 Non-COVID: 2660	Global cough collection through online platform:	"Official test", doctor assessment, or personal assessment	Development of COVID-19 cough classification AI model	COVID-19 cough classification algorithm using a Convolutional Neural Network model Sensitivity = 98.5% Specificity = 94.2%
Android phones and web apps by University of Cambridge (covid-19-sounds.org./en/)	Coppock (2021)	18	# people COVID: 62 Non-COVID: 293	Crowdsourced participants	Self-reporting	Development of COVID-19 cough classification AI model	COVID-19 cough classification algorithm using a Deep Neural Network model AUC = 0.846

*Cayetano Cough Monitor is a Marantz PMD 620 handheld recorder with an Audio-Technica AT899 sub-mini microphone attached at the patient's lapel
TB, tuberculosis; AI, artificial intelligence; Ctl, control; MODS, microscopic-observation drug-susceptibility; N/A, not available

3.14. Supplementary references

1. Loudon, R. G. & Spohn, S. K. Cough frequency and infectivity in patients with pulmonary tuberculosis. Am. Rev. Respir. Dis. 99, 109–111 (1969).
2. Turner, R., Repossi, A., Matos, S., Birring, S. & Bothamley, G. S79 Cough Prevalence And Frequency In Pulmonary Tuberculosis. Thorax 69, A43–A44 (2014).
3. Suzuki, T. et al. Improved cough- and sputum-related quality of life after initiation of treatment in pulmonary tuberculosis. Respir. Investig. 57, 252–259 (2019).
4. Turner, R. D. et al. Daily cough frequency in tuberculosis and association with household infection. Int. J. Tuberc. Lung Dis. 22, 863–870 (2018).
5. Williams, C. M. et al. Exhaled Mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: prospective observational studies. Lancet. Infect. Dis. 20, 607–617 (2020).
6. Tracey, B. H. et al. Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis. Annu Int Conf IEEE Eng Med Biol Soc 2011, 6017–6020 (2011).
7. Larson, S. et al. Validation of an Automated Cough Detection Algorithm for Tracking Recovery of Pulmonary Tuberculosis Patients. PLoS One 7, e46229 (2012).
8. Proaño, A. et al. Dynamics of cough frequency in adults undergoing treatment for pulmonary tuberculosis. Clin. Infect. Dis. 64, 1174–1181 (2017).
9. Proaño, A. et al. Cough Frequency During Treatment Associated With Baseline Cavitary Volume and Proximity to the Airway in Pulmonary TB. Chest 153, 1358–1367 (2018).
10. Lee, G. O. et al. Cough dynamics in adults receiving tuberculosis treatment. PLoS One 15, e0231167 (2020).
11. Carlos Gabaldon-Figueira, J. et al. Digital acoustic surveillance for early detection of respiratory disease outbreaks in Spain: a protocol for an observational study. BMJ Open 11, e051278 (2021).
12. Imran, A. et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Informatics Med. Unlocked 20, 100378 (2020).
13. Kvapilova, L. et al. Continuous Sound Collection Using Smartphones and Machine Learning to Measure Cough. Digit. Biomarkers 3, 166–175 (2019).
14. Pathri, R., Jha, S., Tandon, S. & GangaShetty, S. Acoustic Epidemiology of Pulmonary Tuberculosis (TB) & Covid19 leveraging AI/ML. medRxiv 2022.02.05.22269707 (2022) doi:10.1101/2022.02.05.22269707.
15. Botha, G. et al. Detection of tuberculosis by automatic cough sound analysis. Physiol. Meas. 39, 045005 (2018).
16. Pahar, M. et al. Automatic Cough Classification for Tuberculosis Screening in a Real-World Environment. Physiol. Meas. 42, 105014 (2021).
17. Laguarta, J., Hueto, F. & Subirana, B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open J. Eng. Med. Biol. 1, 275–281 (2020).
18. Coppock, H. et al. End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. BMJ Innov. 7, 356–362 (2021).

Chapter 4. Manuscript II: Objective cough counting in clinical practice and public health: a scoping review

4.1. Preface

Through my work on the previous manuscript, I identified two distinct approaches to digital cough monitoring. The first involves autonomous or semi-autonomous cough counting tools that track changes in cough frequency over time. The second utilizes advanced AI algorithms to analyze cough sounds and identify acoustic features indicative of underlying disease. These approaches serve different purposes: cough counting requires longitudinal monitoring to identify meaningful changes in cough patterns, while acoustic analysis can be used as a point-of-care screening or triage tool with a single measurement.

Between these approaches, cough counting technology has a longer history and more extensive evidence base in the literature. Several cough counting tools have achieved commercial status over the past decade, though their integration into clinical practice remains unknown. Given the maturity and availability of cough counting technology, we sought to understand its current applications in clinical and public health contexts. Through a scoping review of the literature, we examined the how these tools are used for patient care and explored cross-cutting themes such as technological feasibility, user acceptance, and privacy considerations that affect implementation across all contexts.

This work is currently under review for publication in *The Lancet Digital Health*.

4.2. Title page

Objective cough counting in clinical practice and public health: a scoping review

Alexandra J. Zimmer[1,2], Rishav Das[1,2], Patricia Espinoza Lopez[3,4], Vaidehi Nafade[2,5], Genevieve Gore[6], César Ugarte-Gil[3,4,7], Kian Fan Chung[8], Woo-Jung Song[9], Madhukar Pai[1,2], Simon Grandjean Lapierre[2,10,11]

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

2 McGill International TB Centre, Montreal, Canada

3 School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

4 Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, Peru

5 School of Medicine, McGill University, Montreal, Canada

6 Schulich Library of Physical Sciences, Life Sciences, and Engineering, McGill University, Montreal, Canada

7 Department of Epidemiology, School of Public and Population Health, University of Texas Medical Branch, Galveston, USA

8 National Heart and Lung Institute, Imperial College London, London, UK

9 Department of Allergy and Clinical Immunology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

10 Immunopathology Axis, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, 900 Rue Saint-Denis, Montréal, Canada

11 Department of Microbiology, Infectious Diseases and Immunology, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Canada

4.3. Abstract

Cough counts are a marker for respiratory disease diagnosis and monitoring. Traditionally, patient-reported outcomes have provided subjective insights into symptoms. Novel digital cough counting tools now enable objective assessments, yet their integration into clinical practice is limited. This scoping review aims to bridge this gap by examining automated and semi-automated cough counting tools in patient care and public health. Four clinical use cases were identified from the literature: disease diagnosis and severity assessment, treatment monitoring, health outcomes prediction, and syndromic surveillance. Moderate correlations between objective cough frequencies and patient reported outcomes indicate a complex relationship between quantifiable measures and patient experiences. Feasibility challenges include device obtrusiveness, monitoring adherence, and addressing patient privacy concerns. Comprehensive studies are critically needed to validate these technologies in real-world settings and demonstrate their clinical value, and early feasibility and acceptability assessments are essential for successful integration.

4.4. Introduction

Cough has frequently signaled the need for care-seeking, clinical evaluation, and diagnostic testing. The clinical evaluation of cough typically relies on patient reported outcomes (PROs), which involves the subjective assessment of cough using tools such as the Leicester Cough Questionnaire (LCQ) for cough-related quality of life or the Visual Analog Scale (VAS) for cough severity.[1,2] These PROs, while straightforward and practical,[3] frequently fall short in accurately capturing the frequency and nature of cough.

Digital cough monitoring tools and technologies strive to mitigate these shortcomings by using autonomous methods for cough detection and tracking changes in cough frequency over time. Several cough counting technologies have been developed.[4] Prominent among these tools have been the Leicester Cough Monitor (LCM) and VitaloJAKTM, two semi-autonomous devices which have significantly contributed to the field of cough monitoring over the past 20 years for 24-hour cough counting.[5–7] In recent years, advances in artificial intelligence (AI) as well as the heightened focus on respiratory symptoms during the COVID-19 pandemic have accelerated the field of digital cough detection and monitoring.[8–10] This field of research has recently been coined acoustic epidemiology in reference to the use of technology to detect and analyze sounds produced by the body (coughing, sneezing, etc.) in order to better understand and predict health outcomes for patients.[11,12]

Despite these advances, digital cough tools are predominantly confined to clinical research trials to track cough counts as an outcome measurement for new cough-suppressant treatments.[13–15] While objective cough frequency is considered a critical outcome in cough guideline recommendations,[16] its use in clinical practice is rare. This discrepancy underscores a significant gap between ideal standards and actual practice. While existing commentaries have summarised technical aspects of these tools and their conceptual applications,[3,4,17–19] a systematic consolidation of their practical and clinical applications remains essential to advance the field of cough monitoring. Therefore, this scoping review seeks to synthesize the applications of objective cough monitoring tools in the clinical and public health management of respiratory diseases, with an emphasis on their real-world applicability and potential to transform patient care.

Due to the wide range of applications for cough monitoring across various diseases and the use of diverse cough monitoring tools, a scoping review methodology was chosen. The specific aims of this review are: 1) identify and categorize the various objective cough counting tools and technologies used in clinical and public health contexts; 2) analyze how these tools are currently being used for the diagnosis, monitoring, and management of respiratory diseases, and their effectiveness in these roles; and 3) investigate the technological and logistical factors that affect the adoption and integration of digital cough tools into clinical practice and public health strategies.

4.5. Methods

We conducted a scoping review to identify how digital cough counting tools can be used to support public health and clinical care. This review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines (**Table S4.1**).[20]

4.5.1. Definitions

We defined "digital cough counting tools" as devices or systems that utilized automated or semi-automated methods to detect, quantify, and record cough sounds or events continuously over a duration ranging from hours to days. These tools are distinct from subjective measurements, which predominantly rely on PROs and personal recollections (e.g., cough questionnaires).

4.5.2. Data sources and search strategy

To guide our scoping review, we developed a "Population, Concept, Context"[21] driven research question: "How do digital cough counting tools complement and/or enhance standard care for the diagnosis, treatment, and monitoring of patients with respiratory diseases?" This question steered our search across the following databases: MEDLINE (Ovid), EMBASE (Ovid), CENTRAL (Cochrane Library), Web of Science Core Collection, IEEE Xplore, and preprints indexed in Europe PMC (e.g., bioRxiv, MedRxiv). Search terms were carefully selected to include a

combination of keywords related to 'cough' or ‘tussis’ as well as various descriptors of objective measurement, such as 'automated,' biomarker,' 'monitor’, and ‘frequency’, among others. The detailed search strategies (**Table S4.2**) were developed with input from the research team and a skilled research librarian (GG). The scope of the search was limited to articles that have been published in the last 10 years (January 2013- September 2022) with an updated search performed in September 2023. The search was restricted to publications in English, French, and Spanish. During the review process, certain references used to the same dataset, such as when a conference proceeding was later followed by a peer-reviewed article. In such cases, data was extracted from one of the sources, prioritizing peer-reviewed publications when available.

4.5.3. Inclusion and exclusion criteria

Studies were included if they utilized digital cough counting tools to assess cough patterns and frequencies with an explicit clinical intent. While studies with a defined clinical purpose was the primary focus, studies that addressed feasibility and acceptability issues associated with implementing these tools in clinical practice were also included. A broad spectrum of study designs were included, such as observational studies, case reports, and case series, all without a minimum sample size requirement. Randomized trials were included when the primary investigation was the utility of digital cough monitors as an intervention for clinical care. The search was unrestricted with respect to patient demographics, settings, geographic locations, or cough treatment.

Conversely, studies were excluded if the aim was the technical development of the tool (e.g., developing a cough detection algorithm), rather than the application or validation in a clinical or public health context. Studies were also excluded if they used cough counting tools exclusively among healthy individuals, as these do not contribute to our understanding of how best those tools can be used for decision making in a clinical or public health. Experimental studies or clinical trials in controlled settings using cough monitoring as an outcome to evaluate the efficacy of a novel therapeutic were also excluded. Studies that exclusively employed subjective cough assessment tools such as questionnaires, surveys, or scores were excluded. Publications such as reviews, commentaries, editorials, or those lacking primary data or complete publication

information were omitted but scanned for relevant citations. Finally, studies not carried out on human subjects were excluded.

4.5.3. Study screening and selection

The screening process was conducted in two stages and by two reviewers independently to ensure thoroughness and accuracy. During title and abstract screening, all papers were primarily reviewed by author AJZ, with RD, PE, and VN serving as secondary screeners. Full-text articles were screened in the same way. Throughout both screening stages, any instances of discordance or uncertainty were addressed by consulting with SGL for a final decision.

4.5.4. Data extraction

A structured data extraction form was used to systematically gather relevant study information. This form captured a range of information, including basic publication data, study design, country, patient demographics, and the specific respiratory disease(s) being investigated. Additionally, the cough counting tool used was cataloged, as well as the specific clinical application in which it was used in each study, taking note of barriers or facilitators. Studies that reported correlation coefficients between objective cough frequency and subjective cough scores or questionnaires were also extracted. The primary extraction was carried out by AJZ, with a subsequent verification process overseen by SGL.

4.5.5. Qualitative synthesis

To systematically categorize the identified literature, an inductive thematic analysis was used to group the studies based on their clinical and public health applications. This process resulted in the organization of the literature into four main categories: 1) diagnosis and assessment of disease severity and activity; 2) monitoring the treatment of diagnosed conditions; 3) predicting health outcomes; and 4) conducting syndromic surveillance. A fifth category was designated to papers that primarily explored cross-cutting themes, such as the feasibility and acceptability of using

cough monitors. Beyond this categorization, no additional qualitative coding methods were employed.

4.6. Results

4.6.1. Search results

The search strategy identified 7,631 abstracts. After removing duplicates, 5,175 abstracts were screened, of which 534 were selected for full-text assessment (**Figure 4.1**). From this subset, 68 met the inclusion criteria.

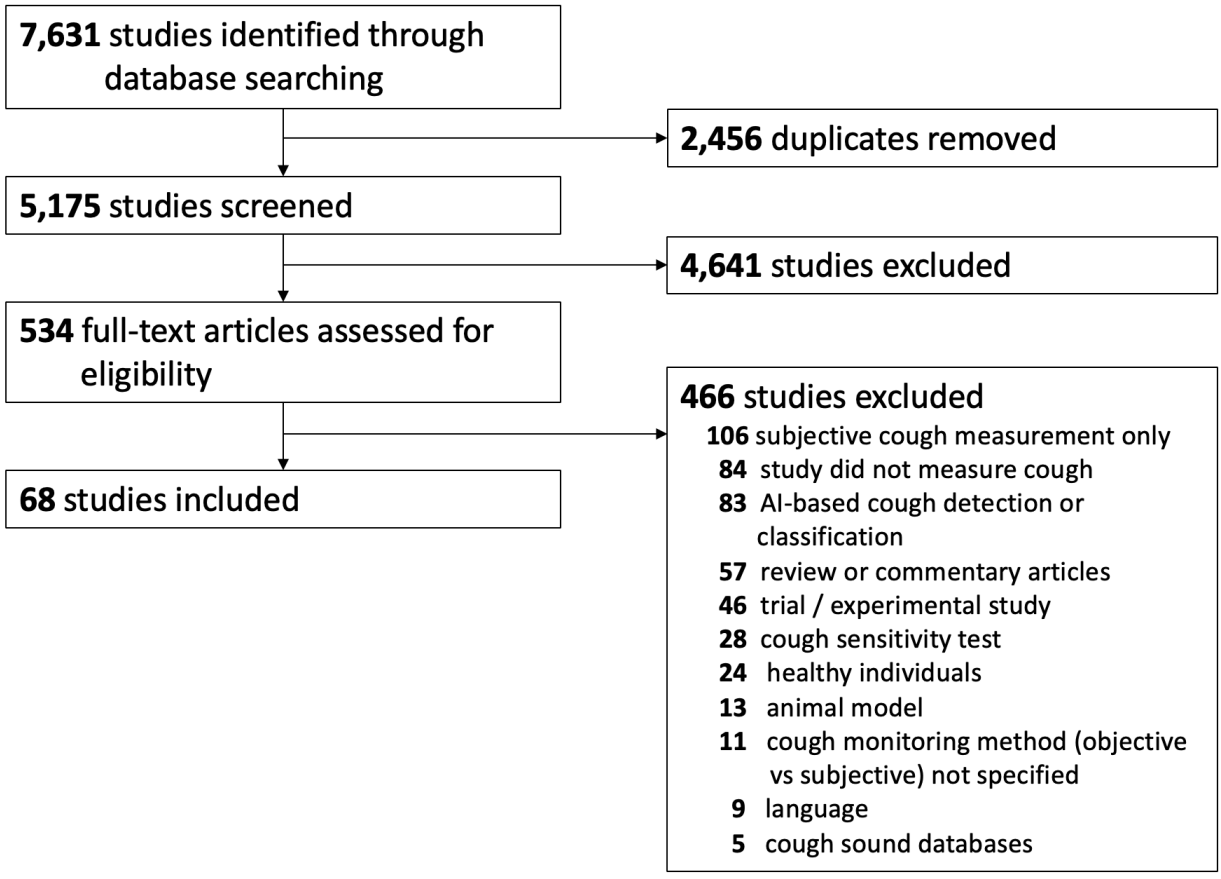


Figure 4.1. PRISMA flow diagram for literature search and study inclusion.

Most publications were excluded because they either solely used subjective cough assessment tools or did not measure cough at all. Eighty-three studies focused on developing AI algorithms

for cough detection and classification without considering longitudinal cough counts as a marker of interest. Additionally, 46 studies used objective cough tools in experimental or trial settings to validate anti-tussive therapeutics, 28 studies investigated cough sensitivity reflex, and 24 studies examined the application among healthy participants.

4.6.2. Study characteristics

Table 4.1 presents a summary of the included studies. Most studies recruited a small sample of individuals with respiratory conditions, with a median of 44 participants, ranging from 1 (case report) to 616 participants. Geographically, 63 (93%) of the studies examining the clinical utility of digital cough tools were conducted in high-income countries. The five studies (7%) in low- and middle-income countries exclusively centered on monitoring outcomes related to tuberculosis (TB).[22–26]

Table 4.1. Summary of studies examining the use of digital cough monitoring tools in a clinical or public health context

Author (Ref.)	Country	Setting	Age group	Disease(s)	N*	Recording time	Cough monitor	Unit of analysis
Clinical application: Diagnosis and assessing disease severity								
Bisballe-Müller 2021[27]	Australia	In-patient	Pediatric	Asthma, bronchiolitis, pneumonia, ARIs	118	24h	Sony ICD-PX470 Digital Voice Recorder	Cough cluster (explosive sounds with no more than 5s between each sound) per hour
Fletcher 2017[28]	N/A	Ambulatory	N/A	Asthma, RCC, GORD, Rhinosinusitis	320	24h	LCM	Cough count per hour
Grosse-Onnebrink 2016[29]	Germany	In-patient	Adolescents, Adults	Cystic fibrosis	21	7h (night)	LEOSound	Cough count per hour
Hirai 2019[30]	Japan	In-patient	Pediatric	Asthma, RSV-bronchiolitis	36	8h (night)	Custom	Cough count per 30min
Hirai 2022[31]	Japan	In-patient	Pediatric	Asthma with and without PND	8	8h (night)	Custom	Cough count per 30min
Imai 2017[32]	Japan	In-patient	Pediatric	Psychogenic cough	2	8h (night)	Custom	Cough count per night
Key 2019[33]	N/A	Ambulatory	Adults	Bronchiectasis	6	24h	N/A	Cough counts per day/night
Lindenhofer 2020[34]	Austria	Ambulatory, In-patient	Pediatric, Adolescents	Asthma, Cystic fibrosis, Pneumonia, Habit cough, Chronic cough	39	8-10h (night)	LEOSound	Cough episodes (30s period in which at least one cough was registered) per night
Radine 2019[35]	Germany	Ambulatory	Pediatric, Adolescents, Adults	Cystic fibrosis, Primary ciliary disease	49	2 nights (consecutive)	LEOSound	Cough counts and cough epochs (continuous coughing without a 2s pause) per hour
Sinha 2016[36]	UK	Ambulatory	Adults	Sarcoidosis	32	24h	LCM	Cough counts per 24h
Spinou 2017[37]	UK	Ambulatory	Adults	Bronchiectasis	54	24h	LCM	Cough counts per 24h
Sumner 2013[38]	N/A	Ambulatory	Adults	COPD	68	24h	VitaloJAK	Cough counts per hour
Turner 2013[39]	N/A	In-patient	Adults	Asthma, COPD, LRTI	40	24h	LCM	Cough counts per hour and cough counts (undefined) per 24h
Vertigan 2018[40]	Australia	Ambulatory	Adults	RCC, Muscle tension	57	24h	LCM	Cough counts per hour

				dysphonia, Vocal cord dysfunction				
Vertigan 2020[41]	Australia	Ambulatory	Adults	Vocal cord dysfunction, chronic cough	90	24h	LCM	Cough counts per hour
Wang 2014[42]	UK	Ambulatory	Pediatric	Pertussis	6	24h	LCM	Cough counts per 24h
Yousaf 2013[43]	UK	Ambulatory	Adults	Asthma, Bronchitis, Chronic cough, COPD	78	24h	LCM	Cough counts per 24h
Clinical application: Disease severity assessment								
Crooks 2016[44]	N/A	Ambulatory	Adults	COPD	18	24h: Baseline, Day 5, Day 20, Day 45	Hull Automated Cough Counter	Cough count per hour
Doenges 2020[45]	Germany	Ambulatory	Adults	Asthma	55	5-9h (night)	LEOSound	Cough count per hour
Elghamoudi 2017[46]	N/A	N/A	Pediatric, Adolescents	Asthma	26	24h: During exacerbation, When stable	VitaloJAK	Cough count per 24h/day/night
Fischer 2018[47]	Germany	Ambulatory	Adults	COPD	30	2 nights (consecutive)	LEOSound	Cough epochs (continuous coughing without a 2 s pause) per recording period
Harle 2019[48]	N/A	Ambulatory	Adults	Lung cancer	39	24h: Baseline, Day 60	VitaloJAK	Cough count per hour
Hirai 2016[49]	Japan	In-patient	Pediatric	Asthma	34	8h (night)	Custom	Cough count per 30min
Holmes 2022[50]	UK	Ambulatory	Adults	Asthma	61	24h	LCM	Cough count per hour
Koehler 2019[51]	Germany	Ambulatory, In-patient	Pediatric	Acute bronchitis	36	10h: Night 1, Night 5, Night 9	LEOSound	Cough epochs (continuous coughing without a 2s pause) per 10-min
Krönig 2017[52]	N/A	Ambulatory	Adults	COPD	48	2 nights (consecutive)	LEOSound	Cough period (cough events within an interval shorter than 15 seconds) per hour
Kulnik 2015[53]	N/A	In-patient	Adults	Stroke	21	24h: Baseline, Week 1, Week 4	LCM	Cough counts per 24h
Lodhi 2019[54]	N/A	N/A	Adults	Asthma	92	24h	VitaloJAK	Cough counts per hour/4h/24h
Mackay 2015[55]	UK	Ambulatory	Adults	COPD	64	24h	VitaloJAK	Cough counts per 24h

Marsden 2016[56]	UK	Ambulatory	Adolescents, Adults	Asthma	89	24h	VitaloJAK	Cough counts per hour
Ovsyannikov 2019[57]	Russia	In-patient	Adults	COPD	110	12h: Baseline, Day 10	Custom	Cough counts per 12h
Proaño 2018[23]	Peru	Ambulatory	Adults	Tuberculosis	41	2 weeks (consecutive), Day 21, Day 30, Day 60	CayeCoM	Cough episode (continuous coughing without a 2s pause) per hour
Rassouli 2020[58]	Switzerland	Ambulatory	Adults	Asthma	79	29 nights (consecutive)	Clara smartphone app.	Cough counts and cluster (a series of at least two coughs with a maximum time of 2s in between their expulsive phases) per night & hour
Rhee 2015[59]	USA	Ambulatory	Adolescents	Asthma	42	7 days (consecutive)	ADAM device	Cough counts per 6s
Schwarz 2021[60]	Germany	Ambulatory	Adults	COPD	40	24h	LEOSound	Cough epochs (among of coughing during 30s) per day/night
Turner 2014[61]	N/A	N/A	N/A	Tuberculosis	30	24h	LCM	Cough counts per 24h
Turner 2015[62]	N/A	Ambulatory	N/A	Chronic cough, COPD, Tuberculosis	69	24h	LCM	Cough episodes (a lone cough or bout of multiple coughs separated by 2s of no cough) per 24h
Turner 2018[63]	UK	Ambulatory In-patient	Adolescents, Adults	Tuberculosis	61	24h	LCM	Cough counts per hour & 24h
Weisser 2022[64]	N/A	N/A	Pediatric	Asthma	94	2 nights (consecutive)	LEOSound	Cough episodes (undefined) per night
Winders 2023[65]	N/A	Ambulatory	Adults	Asthma	108	9h (night)	Tablet with CurieAI algorithm	Cough episode (all coughs within 10 min after an initial cough) per 2h
Clinical application: Treatment monitoring								
Faruqi 2016[66]	N/A	Ambulatory	Adults	Cystic fibrosis	2	24h: Pre-treatment, Post-treatment	Hull Automated Cough Counter	Cough counts per hour
Faruqi 2020[67]	UK	Ambulatory	Adults	Severe eosinophilic asthma	11	24h: Baseline, 1 month, 3 months, 6 months	Hull Automated Cough Counter	Cough counts per 24h
Fukuhara 2019[68]	Japan	Ambulatory	Adults	Asthma	73	24h: Pre-treatment, Post-treatment	MP3 sound recorder with a free-field microphone	Cough counts per hour

Huddart 2023[22]	India, the Philippines, South Africa, Uganda, Vietnam	Ambulatory	Adults	Tuberculosis	144	14 days (consecutive)	Hyfe Cough Tracker app.	Median cough counts per hour
Jung 2023[69]	South Korea	Ambulatory	N/A	Asthma, Interstitial lung disease	45	Asthma: 7 days (consecutive) 2h during daytime & 5h during nighttime ILD: Baseline, 3 months 2h during daytime & 5h during nighttime	Coughy app.	Cough epochs (undefined) and cough counts per recording period
Kang 2023[70]	South Korea	Ambulatory	Adults	Long-COVID-19 refractory cough	1	42 days (consecutive)	Hyfe Cough Tracker app.	Cough counts per 24h
Lee 2020[25]	Peru	Ambulatory	Adults	Tuberculosis	69	4h: Baseline, Day 3, Day 7, Day 30, Day 60	CayeCoM	Cough episodes (series of coughs separated by <2s between each cough) per hour
Lee 2023[71]	South Korea	Ambulatory	Adults	Chronic cough	43	Up to 14 days (consecutive)	Hyfe Cough Tracker app.	Cough counts per hour
Proaño 2017[24]	Peru	Ambulatory	Adults	Tuberculosis	64	24h: Pre-treatment, Day 3, Day 7, Day 14, Day 21, Day 30, Day 60	CayeCoM	Cough episodes (series of coughs separated by <2s between each cough) per hour
Shim 2023[72]	South Korea	Ambulatory	N/A	Asthma	24	7 days (consecutive) 2h during daytime & 5h during nighttime	Coughy app.	Cough counts per hour
Turner 2015[73]	N/A	Ambulatory, In-patient	N/A	Tuberculosis	44	24h: Baseline, Day 1, Day 2-4, Day 5-8, Week 2, Week 8, Week 26	LCM	Cough counts per hour
Vertigan 2021[74]	Australia	Ambulatory	Adults	Asthma, RCC, Laryngeal obstruction	174	24h: Pre-intervention, post-intervention	LCM	Cough counts per 24h, daytime & nighttime
Zhang 2022[75]	UK	Ambulatory	Adults	Cystic fibrosis	16	24h: Pre-treatment, 1 month	Philips Digital Record	Cough counts per 24h, daytime & nighttime
Clinical application: Predicting health outcome								

Altshuler 2023[76]	Canada, USA	In-patient	Adults	COVID-19	123	Until discharge/death	Hyfe Cough Tracker app.	Cough counts per hour
Boesch 2023[77]	Switzerland	In-patient	Adults	COVID-19, Pneumonia	46	Until discharge/death	Cough detection app. (unspecified)	Cough counts per hour and 6h; Mean coughs per hour
Crooks 2021[78]	UK	Ambulatory	Adults	COPD	25	8 days before & 8 days after AE-COPD (consecutive)	Stationary microphone with laptop	Cough counts per 24h
den Brinker 2021[79]	UK	Ambulatory	Adults	COPD	28	90 days (consecutive)	Custom	Cough counts per hour
Pekacka-Egli 2021[80]	Switzerland	In-patient	Adults	Post-stroke pneumonia	30	8h (night)	LEOSound	Cough counts per hour
Tinschert 2020[81]	Switzerland	Ambulatory	Adults	Asthma	79	29 nights (consecutive)	Smartphone app.	Cough counts during first 30 min of sleep; Cough counts per night
Public health application: Syndromic surveillance								
Al-Hossain 2020[82]	USA	Ambulatory	All ages	Influenza, Influenza-like illness	N/A	Daytime	FluSense	Average cough counts per hour per day; Total cough epoch (coughs occurring with less than 3s difference)
Gabaldon-Figueira 2022[83]	Spain	Ambulatory	Adolescents, Adults	COVID-19	616	24h: Until discontinuation	Hyfe Cough Tracker app.	Cough counts per hour
Rahman 2023[84]	USA	Waiting room	Adults	COVID-19, Influenza, RSV	N/A	4 months	Syndromic Logger	Cough counts per 24h
Zürcher 2022[26]	South Africa	Waiting room	Adolescents, Adults	Tuberculosis	N/A	Daytime	CoughSense	Cough counts per 24h
Cross-cutting and feasibility studies								
Elghamoudi 2015[85]	N/A	Ambulatory, In-patient	Pediatric, Adolescents	Acute cough, chronic cough	40	24h	VitaloJAK	N/A
Gross 2015[86]	N/A	Ambulatory	Pediatric, Adolescents	Asthma	40	8h (night)	LEOSound	N/A
Huddart 2022[87]	India, the Philippines, South Africa, Uganda, Vietnam	Ambulatory	Adults	Tuberculosis	144	14 days (consecutive)	Hyfe Cough Tracker app.	Cough counts per hour, Maximum cough counts per hour per day, median cough counts per hour per day
Rhee 2014[88]	USA	Ambulatory	Adolescents	Asthma	42	7 days (consecutive)	ADAM device	N/A

Urban 2022[89]	Germany	In-patient	Pediatric, Adolescents	Asthma, Bronchitis, Pneumonia	86	8h (night)	LEOSound	Cough epochs (coughs during 30s) per hour

*N does not include healthy individuals.

ADAM, automated device for asthma monitoring; CayeCoM, Cayetano Cough Monitor; COPD, chronic obstructive respiratory disease; GORD, gastro-oesophageal reflux disease; LCM, Leicester Cough Monitor; LRTI, lower respiratory tract infection; N/A, not available; PND, post-nasal drip; RCC, refractory chronic chough; RSV, respiratory syncytial virus; UK, United Kingdom; USA, United States of America

Most studies (N=47, 69%) focused on conditions linked to chronic cough. Five studies (7%) investigated both chronic and acute respiratory ailments. Regarding specific diseases, asthma was the most frequently investigated, accounting for 26 (38%) of studies. This was followed by chronic obstructive pulmonary disease (COPD) (N=11, 16%), TB (N=10, 15%), and cystic fibrosis (N=5, 7%).

Digital cough counting tools predominantly find their application in ambulatory settings, enabling remote monitoring as patients take the devices home. Most of these ambulatory studies primarily focus on adults over 18 years old. Few studies (N=15, 22%) assessed objective cough counting in pediatric populations under 15 years of age. Only contact-based devices were compatible with pediatric populations, such as the LEOSound,[34,86,89] and a custom cough monitoring system developed by Hirai et al. for in-patient pediatric monitoring.[30,31,49] Non-contact (e.g., smartphone applications) tools were not investigated in pediatric populations.

Contact-based semi-automated systems—those requiring physical proximity or direct interaction with the patient—remain the most widely used. These include the LCM (N=15, 22%%), VitaloJAK (N=7, 10%), Hull Automatic Cough Monitor (N=3, HACC) (4%), and Cayetano Cough Monitor (CayeCoM) (N=3, 4%). The technical specifications of these devices have been well-documented.[3,4] These tools use commercially available digital recorders (e.g., MP3) with free-field microphones,[4] aside from the VitaloJAK which has a custom built-in microphone.[6] We also identified several studies that deployed non-contact, autonomous cough detection algorithms on smartphones, including Clara,[58] Coughy,[69,90] Hyfe Cough Tracker,[22,70,71,76,83,87] and CurieAI.[65] These algorithms represent a shift towards more accessible and user-friendly cough counting solutions, leveraging the widespread availability of smartphones.

The variability also extends to how authors define the unit of analysis, specifically, cough counts. Most studies (N=49, 72%) reported coughs counts per unit of time (e.g., per hour or per 24 hours) as the primary outcome for measuring cough frequency (**Table 4.1**). Other studies adopted the concept of a “cough epoch” (or “cough bout”) per hour, defined as a continuous coughing episode without a pause exceeding two seconds, proposed by the European Respiratory Society guidelines for cough evaluation.[16]

4.7. Qualitative synthesis: clinical and public health applications

In line with the thematic analysis, studies were categorized based on their clinical use of cough counting (**Figure 4.2**): 1) disease screening (N=17, 25%), 2) disease severity assessment (N=23, 34%); 2) treatment monitoring of previously diagnosed conditions (N=13, 19%); 3) health outcome prediction (N=6, 9%); 4) syndromic surveillance (N=4, 6%).

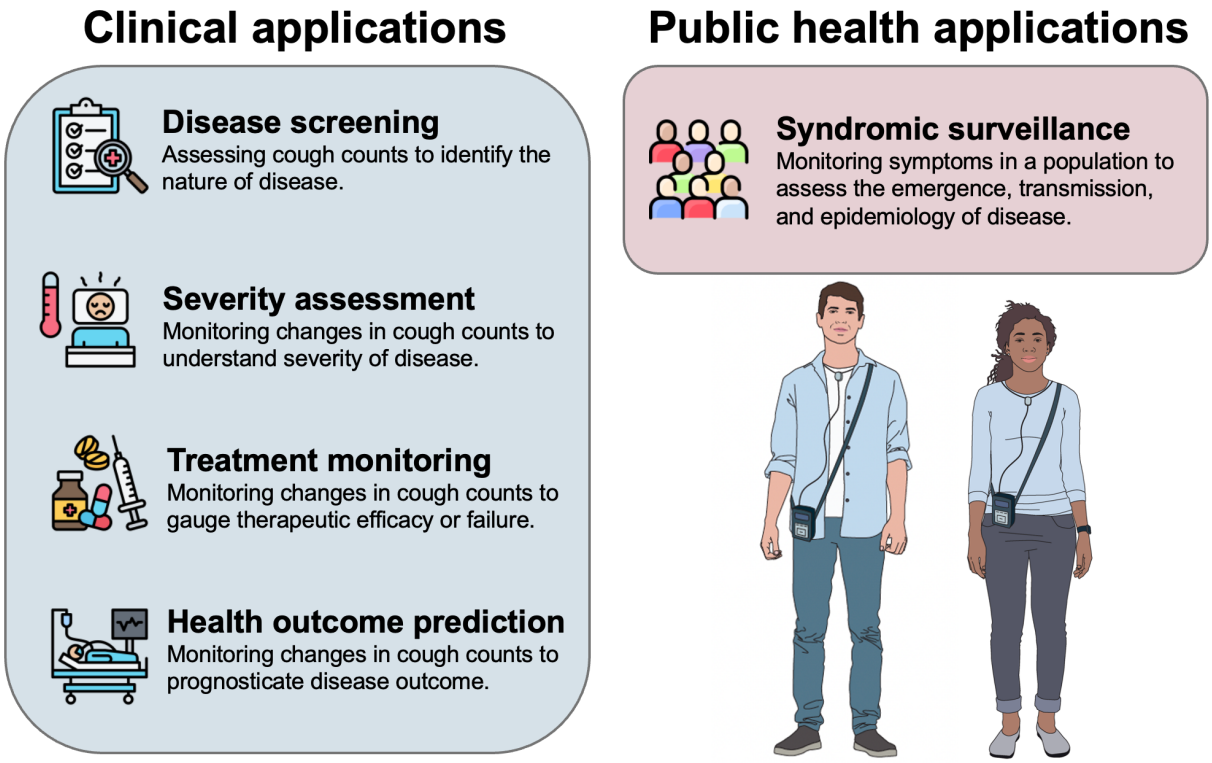


Figure 4.2. Potential clinical and public health applications of digital cough counting tools.

4.7.1. Disease screening

The potential of cough frequency as a diagnostic biomarker lies in its ability to reveal disease-specific cough patterns. These patterns can be examined over the course of a single day (circadian analysis) or over several days or weeks (longitudinal analysis).

Circadian cough frequency monitoring may offer a non-specific indication of underlying disease. Studies have shown differences in circadian cough frequency changes between healthy individuals

and those with pulmonary diseases.[29,35–37,40,43,54,74] However, distinguishing rates and patterns between different diseases is more challenging. While some diseases exhibit different absolute cough counts (e.g., cystic fibrosis vs primary ciliary dyskinesia)[35] or unique cough patterns (e.g., asthma vs acute bronchiolitis),30 other do not.[28,39,41,43,62,74] Given the many illnesses that present with coughing and the variability in cough patterns, using cough counts alone for diagnosis is challenging. Therefore, while increased cough counts may indicate the presence of disease, they are not consistently reliable for differential diagnosis.

4.7.2. Disease severity assessment

Longitudinal cough monitoring, which often involved repeated 24-hour intervals, is used to observe trends and understand the evolving nature of cough patterns and severity over time.[23,44,48,51,53,57–59] For example, Crooks et al. documented a decline in cough frequency from baseline to day 45 among patients recovering from acute exacerbations of COPD, highlighting the natural trajectory of recovery.[44] Similarly, Koelher et al. examined nocturnal cough frequency in children with bronchitis over three nights (baseline, night 5, night 9) and demonstrated how objective cough monitoring can track the evolution of acute respiratory diseases.[51]

Both circadian and longitudinal cough trends provide valuable information on disease severity in patients with established diagnoses. Asthma is the most extensively researched condition in this context. Reports suggest that nocturnal coughing among asthmatic individuals can indicate more severe asthma.[56] Several studies on single-night nocturnal cough patterns discerned variations in cough frequency among individuals with varying levels of asthma severity. Notably, those with more severe, less controlled asthma demonstrated heightened nocturnal cough episodes.[45,46,49,50,54,56,64,65]

Studies have also examined circadian and nocturnal cough patterns in COPD patients. Unlike asthma, no significant differences in circadian and nocturnal cough frequency was observed when comparing stable COPD with acute exacerbations of COPD (AECOPD).[55,60] Additionally, patients in advanced stages (COPD IV) coughed less frequently than those in the moderate stage (COPD

III) and exhibited a more uniform pattern of nocturnal coughing, showing less variation throughout the night compared to patients in earlier stages.[47,52]

In the realm of respiratory infections, cough trends in TB have been explored.[23,61–63] A correlation between TB severity and cough frequency has been observed, with patients having positive sputum smear results (indicative of a higher bacterial load) showing increased cough frequency.[61,63] Proaño et al. further supported this by finding a positive association between the severity of pulmonary cavitation and a heightened cough rate.[23] This underscores cough counts as a potential marker for assessing the severity, activity, and progression of TB.

4.7.3. Treatment monitoring of previously diagnosed conditions

Frequent assessments throughout treatment ensure that therapeutic interventions are effective. Typically, treatment response uses various indicators, from PROs to objective clinical markers like laboratory results or imaging. Amid these different tools and biomarkers, objective cough monitoring could offer real-time, non-invasive insights into treatment efficacy. Instead of assessing novel treatment efficacy in controlled trials, the studies included in this section primarily evaluated patient outcomes in routine clinical practice settings.

Most research has explored cough as a biomarker for monitoring treatment in chronic respiratory diseases such as asthma,[67–69,74,90] interstitial lung disease,[69] cystic fibrosis,[66,75] chronic cough,[71] and long-COVID chronic cough.[70] These studies consistently observed reductions in cough counts during treatment, enabling a more personalized approach to monitoring based on real-time feedback from the patient's cough pattern and frequency.

For infectious respiratory diseases, cough monitoring has primarily been examined as a biomarker for TB treatment response.[22,24,25,73] The current standard for TB treatment monitoring relies on sputum production, which can be challenging for certain populations and during later stages of treatment.[91] Therefore, objective cough monitoring offers the potential to improve TB treatment monitoring by promoting a more patient-centric approach. That said, cough counts are contingent on various demographic and clinical factors and thus may not consistently manifest across

individuals. For instance, Proaño et al. found that some TB-diagnosed participants did not cough,[24] while Lee et al. observed that 21% of TB patients had pre-treatment cough rates similar to healthy individuals.[25] Lee et al. also highlighted the difficulty of applying a universal threshold for assessing TB treatment response using cough counts.

While reduced cough frequency during treatment appears to correlate with positive therapeutic outcomes for many diseases, its use as a solitary measure of treatment success remains uncertain due to this variability across patient populations. The current evidence, although promising, is not robust enough to support cough counts as the sole indicator of treatment efficacy. Therefore, it is more prudent to use objective cough monitoring as an adjunct to a suite of clinical indicators (e.g., weight changes) or in combination with cough-specific PROs, particularly when cough is a prominent symptom of the disease.

Addressing the heterogeneity of cough may require a personalized approach to monitoring. Various demographic and clinical factors influence an individual's cough rate, including smoking habits, sputum production, and airway inflammation.[38] Instead of establishing a universal threshold, it may be more effective to analyze individual changes in cough frequency relative to a pre-treatment baseline.

4.7.4. Health outcome prediction

Telemonitoring, a component of telemedicine that involves the remote collection and transmission of patient health data to healthcare providers using digital technologies and telecommunications systems, enables real-time tracking of patient health metrics and timely interventions. However, evidence of its role in reducing hospitalizations is inconsistent,[92] partly due to the generation of "false alerts".[93] A growing body of research is exploring the potential of monitoring cough counts and patterns as a more precise tool.[76–81]

Passive cough alert systems have been evaluated for monitoring exacerbations in chronic pulmonary diseases such as COPD and asthma.[78–80] These studies found that cough monitoring for exacerbations is more specific than sensitive. For instance, Crooks et al. discovered that symptom

questionnaires predicted 88% of acute COPD exacerbations, while a cough monitor only identified 45%, though the monitor generated fewer false alerts.[78] Similarly, Tinschert et al. observed that nocturnal cough events were more specific than sensitive in forecasting asthma attacks.[81] They proposed that integrating cough monitoring with additional markers, such as sleep quality, could enhance predictive accuracy.

In hospitals, there's interest in leveraging cough counts for immediate clinical decision-making. This method could avoid more invasive techniques and enable quick responses to disease fluctuations. Several studies have investigated the relationship between cough rates and health outcomes. For instance, Pekacka-Egli et al. linked frequent coughs to a higher risk of post-stroke pneumonia.[80] Boesch et al. found a positive correlations between cough frequency and markers indicative of disease progression among COVID-19 and pneumonia patients, suggesting that increased cough indicates heightened disease activity.[77] In contrast, Altshuler et al. found an inverse relationship between cough frequency and severe COVID-19 outcomes, implying that a higher cough rate was linked to decreased adverse outcomes.[76] These findings, when compared to Boesch's, highlight the complexities of interpreting cough dynamics in acute care settings. The variability of cough frequency at different disease stages adds another layer of complexity. Altshuler's study highlighted disparities in cough frequency between patients in Florida and Montreal, attributing these differences to the disease stage: Montreal patients were typically at an earlier stage of COVID-19, leading to more frequent coughing than those in Florida.[76]

4.7.5. Syndromic surveillance

In public health and epidemiology, syndromic surveillance is crucial for early detection and response to outbreaks by monitoring symptom patterns within a population. Traditionally, this involves using healthcare data like emergency department visits or pharmacy sales to predict disease activity. By monitoring population-level changes in cough frequency, researchers are beginning to uncover how this unobtrusive measure might serve as a proxy for the incidence of infectious respiratory diseases.

Gabaldon-Figueira et al. leveraged the widespread availability of smartphones to deploy the Hyfe Cough Tracker application to 616 residents in Northern Spain over a 10-month period in 2021.[83] The portability and ubiquity of smartphones streamlined the community-wide deployment. However, the study faced challenges with consistent usage among participants, with retention gradually declining. The data suggested higher adoption rates among individuals with chronic respiratory diseases, who naturally cough more frequently and may bias the results. Additionally, the study covered only 1.7% of the intended population, which may not be sufficient to reflect population-wide cough dynamics.

Other studies explored stationary approaches to monitoring respiratory health. Al-Hossain et al. introduced FluSense, a multi-modal platform (including cough detection) designed to assess the influenza burden within a university setting.[82] Installed in four clinic waiting rooms, FluSense found that cough counts correlated with the number of positive flu tests, with a correlation coefficient (ρ) of 0.40. Similarly, Rahman et al. deployed the Syndromic Logger in a tertiary hospital's emergency department waiting room to track respiratory pathogens, including COVID-19, influenza, and RSV.[84] The daily cough counts showed moderate correlation with positive COVID-19 (ρ=0.40) and RSV (ρ=0.27) cases, but no association was found for influenza.

Stationary systems like FluSense and the Syndromic Logger offer insights in healthcare environments, where understanding the disease burden is crucial. However, a fundamental limitation of stationary systems is their confinement to specific settings. These studies primarily captured cough sounds from patients seeking care, which might exaggerate the perceived severity of respiratory conditions in the broader community. Mobile systems like the Hyfe Cough Tracker can potentially capture cough dynamics in the broader community, provided challenges with uptake and retention are addressed.

4.7.6. Cross-cutting themes

Continuous cough recording tools, which constantly monitors and records cough sounds, raise privacy concerns and may lead to user distrust. Al-Hossain et al. highlighted these privacy

concerns in the context of cough surveillance.[82] A potential solution is to refine cough detection algorithms to selectively record cough sounds, minimizing the risk of recording conversations.[78,83]

Portable cough monitors can be obtrusive, often requiring both a recorder and an external microphone, which can be burdensome for patients in ambulatory settings.[22,51,75,85,89] There is a trend toward developing more user-friendly, less obtrusive solutions like smartphone-based applications. These applications could overcome limitations of wearable devices but introduce new challenges, such as battery consumption, potential interference from other applications, and user compliance. Patients may need to actively engage with the app, leading to gaps in data collection. Lee et al. found that only 66.2% of participants adhered to a smartphone-based cough monitoring over one week, with lower adherence observed among older subjects.[71] This study further highlighted issues affecting ease of use, including the inconvenience of having to carry the device. Further studies are needed to understand these emerging challenges related to smartphone-based solutions, and how they might impact the widespread the adoption of smartphone-based solutions.

Environmental factors influencing cough detection accuracy remains a pertinent issue, as current algorithms do not differentiate between individuals. In crowded spaces or shared living environments, cough sounds from others may be inadvertently recorded, potentially skewing the patient's cough data and limiting its clinical utility. This problem is particularly relevant in shared domestic settings or public spaces, where multiple cough sources coexist. [24,34,44,77,79,83,89] The development of more sophisticated AI models capable of recognizing individual cough signatures, referred to as "diarization" in audio data analysis,[94] promises to enhance the precision and personalization of cough tracking. Until such advancements are realized, the indiscriminate nature of cough recording remains a limitation in the field of objective cough monitoring, underscoring the need for further refinement in cough detection technology.

Finally, several studies evaluated the relationship between objective cough counts and PROs. Studies reported moderate correlations with cough severity PROs, with median unweighted correlation coefficients (ρ) of 0.44 (interquartile range [IQR]: 0.36 to 0.55, n=24) for the VAS and 0.43 (IQR: 0.34 to 0.46, n=10) for the Verbal Category Descriptive scale. These correlations were maintained across day and night assessments (**Table S4.3**). In terms of quality-of-life scales, the

LCQ similarly displayed a moderate negative correlation with objective cough counts, with a median ρ of -0.49 (IQR: -0.61 to -0.44, n=10). Notably, the association between cough frequency and the Parental Cough-Specific Quality of Life (PC-QoL)—a standardized tool for assessing cough in young children, with parents serving as proxy assessors—was weaker, with a median correlation coefficient of -0.26 (IQR: -0.29 to -0.08, n=7).[27,34,51]

The moderate associations observed across different studies, each investigating different diseases and utilizing different cough measuring tools (**Table S4.3**), underscore the complex relation between an individual's subjective perception and empirically quantified cough frequency. One hypothesis advocates for the primacy of the subjective experience, suggesting that cough frequency is only one aspect contributing to an individual's experience and quality of life and that it is important to consider the patient experience as a whole.[17] Another hypothesis highlights the importance of objective measurements, proposing that individuals may not fully recognize the frequency of their coughing and thus underestimate its severity. This is particularly true for proxy tools such as the PC-QoL, where reporting is based on a third-party observer. Conversely, as shown by Ovsyannikov et al., patients with anxiety might exaggerate their symptoms exhibiting an inverse correlation between VAS and cough (r=-0.38 for normal weight individuals with anxiety and r=-0.40 for obese individuals with anxiety).[57] These hypotheses underscore the need to comprehensively document various dimensions of coughs to ensure a holistic understanding and management of this symptom within the greater context of the patient's health profile.

4.8. Discussion

While digital cough counting tools are commonly used in clinical trials to evaluate cough suppressant therapies, their integration into everyday clinical practice and public health remains underexplored. This review aimed to bridge the gap between rapid technological innovation in digital cough counting tools and their practical implementation clinical settings.

Across all identified clinical applications, there is insufficient evidence to support using cough as a standalone marker. While cough patterns may change with disease etiology, severity, or treatment response, relying solely on this marker could lead to inaccuracies due to the variability in cough

patterns between and within individuals. Monitoring over several days or weeks offers a better understanding of cough status or treatment response compared to short (24-hour) or intermittent monitoring, though this poses feasibility challenges. Despite these challenges, cough counts and patterns have potential to support clinical management when used within a broader clinical context. Specifically, they should be considered as part of a comprehensive suite of biomarkers that provide a complete picture of an individual's disease status. Personalized approaches can enhance the diagnostic value of cough monitoring, making it most effective when correlated with additional patient-specific health indicators, such as respiratory sounds, cardiac rate, and blood pressure. Additionally, PROs are crucial in capturing subjective experiences of cough and its impact on quality of life, which may not align with objective cough frequency. Measuring PROs acknowledges that a reduction in cough frequency does not necessarily equate to an improved patient experience.

Most studies focused on using cough monitoring to identify disease etiology and assess disease severity. While cough patterns can consistently signal the presence of disease, their diagnostic utility is limited. Combining cough patterns with other cough characteristics, such as type of cough (e.g., wet vs dry), could enhance diagnostic accuracy. The development of AI algorithms for detecting underlying disease from cough sounds ("cough classification") takes this application one step further.[95] It remains important for future studies to characterize cough patterns associated with different diseases and across diverse populations. Emphasis should be placed on long-term cough monitoring to track fluctuations in chronic disease activity or severity, such as chronic cough, asthma, or COPD cough-related exacerbations.

Cough monitors for public health surveillance represent a new yet underexplored opportunity to improve disease outbreak management. These tools have the potential to enhance our grasp of disease transmission dynamics. An expanding field of research is examining the role of various physiological markers gathered from wearable devices to support outbreak management efforts,[96] though the reliability and accuracy of this approach requires further validation. Integrating personalized data, including cough monitoring data, into efficient outbreak response strategies is still developing, with data privacy and ownership remaining critical concerns. Evolving regulations on AI in healthcare, which vary across regions, also affect the adoption of AI systems,

including cough monitors. Addressing these issues is essential for leveraging cough monitors and other wearable devices as effective tools in disease surveillance and outbreak management.

While clinical performance data for cough monitoring tools are limited, insights into their implementation, feasibility, and societal acceptance are even scarcer. Technologies that record patient data raise social and ethical concerns, influencing adoption regardless of performance. Few studies addressed the patient's perspective on these issues. While some discussed privacy and comfort, comprehensive exploration of these concerns was lacking. As clinical validation progresses, it is imperative to conduct qualitative research encompassing patient experiences, ethical considerations, data governance, and balance of societal values with innovation.[97] This will ensure that these tools are clinically effective, socially responsible, and aligned with patient expectations and norms.

This scoping review has limitations. The clinical application of digital cough counting tools remains speculative, primarily inferred from clinical studies rather than grounded in routine healthcare practice. Additionally, there is considerable diversity in the studied diseases, evaluated technologies, and their clinical applications across studies, which limits direct comparison between studies.

This scoping review marks an initial effort to summarize how digital cough monitoring tools could enhance clinical and public health outcomes for respiratory conditions. Currently, these tools are not integrated into standard clinical practice. While innovation is crucial for addressing challenges like accurate cough detection and device ergonomics, there is a pressing need for more extensive evaluation of these tools' effectiveness in clinical environments. This review underscores the scarcity of research focused on the clinical application, indicating that technological innovation has outpaced clinical validation. Reliable and available digital cough data is not currently guiding decision making and clinical management. For effective adoption in routine practice, a deeper understanding of end-user acceptability and rigorous clinical validation studies are essential, providing clinicians with actionable information.

4.9. Acknowledgements

We thank Patrick Lane (ScEYEnce Studios, USA) for the graphic design contributions to Figure 4.2. The authors received no financial support for this research.

4.10. Contributions

Conceptualization (A.J.Z., M.P., S.G.L.), protocol development (A.J.Z., R.D., P.E.L., V.N., G.G., C.U.G., M.P., S.G.L.), search strategy development (A.J.Z., G.G.), database searching (G.G.), screening (A.J.Z., R.D., P.E.L. V.N.), data extraction (A.J.Z., S.G.L.), data synthesis (A.J.Z.), writing original draft (A.J.Z.), writing – review and editing (A.J.Z., R.D., P.E.L., V.N., G.G., C.U.G., K.F.C., W.J.S., M.P., S.G.L.).

4.11. References

1. Birring SS. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). Thorax. 2003;58(4):339-343. doi:10.1136/thorax.58.4.339
2. Martin Nguyen A, Bacci ED, Vernon M, et al. Validation of a visual analog scale for assessing cough severity in patients with chronic cough. Ther Adv Respir Dis. 2021;15:175346662110497. doi:10.1177/17534666211049743
3. Spinou A, Birring SS. An update on measurement and monitoring of cough: what are the important study endpoints? J Thorac Dis. 2014;6(Suppl 7):S728-734. doi:10.3978/j.issn.2072-1439.2014.10.08
4. Hall JI, Lozano M, Estrada-Petrocelli L, Birring S, Turner R. The present and future of cough counting tools. J Thorac Dis. 2020;12(9):5207-5223. doi:10.21037/jtd-2020-icc-003
5. Birring SS, Fleming T, Matos S, Raj AA, Evans DH, Pavord ID. The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough. European Respiratory Journal. 2008;31(5):1013-1018. doi:10.1183/09031936.00057407
6. McGuinness K, Holt K, Dockry R, Smith J. P159 Validation of the VitaloJAKTM 24 Hour Ambulatory Cough Monitor: Abstract P159 Table 1. Thorax. 2012;67(Suppl 2):A131.1-A131. doi:10.1136/thoraxjnl-2012-202678.220
7. Smith J, Woodcock A. New Developments in the Objective Assessment of Cough. Lung. 2008;186(S1):48-54. doi:10.1007/s00408-007-9059-1
8. Rashid HA, Mazumder AN, Niyogi UPK, Mohsenin T. CoughNet: A Flexible Low Power CNN-LSTM Processor for Cough Sound Detection. In: 2021 IEEE 3rd International

Conference on Artificial Intelligence Circuits and Systems (AICAS). IEEE; 2021:1-4. doi:10.1109/AICAS51828.2021.9458509

9. Sharan RV. Cough sound detection from raw waveform using SincNet and bidirectional GRU. Biomedical Signal Processing and Control. 2023;82:104580. doi:10.1016/j.bspc.2023.104580
10. Alqudaihi KS, Aslam N, Khan IU, et al. Cough Sound Detection and Diagnosis Using Artificial Intelligence Techniques: Challenges and Opportunities. IEEE Access. 2021;9:102327-102344. doi:10.1109/ACCESS.2021.3097559
11. Keen EM, True EJ, Summers AR, Smith EC, Brew J, Grandjean Lapierre S. High-throughput digital cough recording on a university campus: A SARS-CoV-2-negative curated open database and operational template for acoustic screening of respiratory diseases. DIGITAL HEALTH. 2022;8:205520762210975. doi:10.1177/20552076221097513
12. Zimmer AJ, Ugarte-Gil C, Pathri R, et al. Making cough count in tuberculosis care. Commun Med. 2022;2(1):83. doi:10.1038/s43856-022-00149-w
13. Friedrich C, Francke K, Birring SS, et al. The P2X3 receptor antagonist filapixant in patients with refractory chronic cough: a randomized controlled trial. Respir Res. 2023;24(1):109. doi:10.1186/s12931-023-02384-8
14. Dutta P, Funston W, Mossop H, et al. Randomised, double-blind, placebo-controlled pilot trial of omeprazole in idiopathic pulmonary fibrosis. Thorax. 2019;74(4):346-353. doi:10.1136/thoraxjnl-2018-212102
15. Ludbrook VJ, Hanrott KE, Kreindler JL, et al. Adaptive study design to assess effect of TRPV4 inhibition in patients with chronic cough. ERJ Open Res. 2021;7(3):00269-02021. doi:10.1183/23120541.00269-2021
16. Morice AH, Fontana GA, Belvisi MG, et al. ERS guidelines on the assessment of cough. European Respiratory Journal. 2007;29(6):1256-1276. doi:10.1183/09031936.00101006
17. Turner RD, Birring SS. Measuring cough: what really matters? J Thorac Dis. 2023;15(4):2288-2299. doi:10.21037/jtd-23-230
18. Schmit KM, Coeytaux RR, Goode AP, et al. Evaluating Cough Assessment Tools. Chest. 2013;144(6):1819-1826. doi:10.1378/chest.13-0310
19. Cho PSP, Birring SS, Fletcher HV, Turner RD. Methods of Cough Assessment. The Journal of Allergy and Clinical Immunology: In Practice. 2019;7(6):1715-1723. doi:10.1016/j.jaip.2019.01.049
20. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169(7):467-473. doi:10.7326/M18-0850
21. The Joanna Briggs Institute. The Joanna Briggs Institute Reviewers' Manual - Methodology for JBI Scoping Review.; 2015. https://reben.com.br/revista/wp-content/uploads/2020/10/Scoping.pdf
22. Huddart S, Asege L, Jaganath D, et al. Continuous cough monitoring: a novel digital biomarker for TB diagnosis and treatment response monitoring. int j tuberc lung dis. 2023;27(3):221-222. doi:10.5588/ijtld.22.0511

23. Proaño A, Bui DP, López JW, et al. Cough Frequency During Treatment Associated With Baseline Cavitary Volume and Proximity to the Airway in Pulmonary TB. Chest. 2018;153(6):1358-1367. doi:10.1016/j.chest.2018.03.006
24. Proaño A, Bravard MA, López JW, et al. Dynamics of Cough Frequency in Adults Undergoing Treatment for Pulmonary Tuberculosis. Clinical Infectious Diseases. 2017;64(9):1174-1181. doi:10.1093/cid/cix039
25. Lee GO, Comina G, Hernandez-Cordova G, et al. Cough dynamics in adults receiving tuberculosis treatment. Ehtesham HS, ed. PLoS ONE. 2020;15(6):e0231167. doi:10.1371/journal.pone.0231167
26. Zürcher K, Riou J, Morrow C, et al. Estimating Tuberculosis Transmission Risks in a Primary Care Clinic in South Africa: Modeling of Environmental and Clinical Data. The Journal of Infectious Diseases. 2022;225(9):1642-1652. doi:10.1093/infdis/jiab534
27. Bisballe-Müller N, Chang AB, Plumb EJ, Oguoma VM, Halken S, McCallum GB. Can Acute Cough Characteristics From Sound Recordings Differentiate Common Respiratory Illnesses in Children? Chest. 2021;159(1):259-269. doi:10.1016/j.chest.2020.06.067
28. Fletcher H, Cho PSP, Matos S, et al. Chronic cough: objective cough frequency in a large cough clinic cohort. In: Clinical Problems Asthma. European Respiratory Society; 2017:OA4677. doi:10.1183/1393003.congress-2017.OA4677
29. Grosse-Onnebrink J, Radine A, Werner C, Omran H. ePS05.7 Objectification of cough in cystic fibrosis. Journal of Cystic Fibrosis. 2016;15:S47. doi:10.1016/S1569-1993(16)30230-2
30. Hirai K, Nukaga M, Tabata H, et al. Objective measurement of nocturnal cough in infants with acute bronchiolitis. Respiratory Investigation. 2019;57(6):605-610. doi:10.1016/j.resinv.2019.06.005
31. Hirai K, Kaiga C, Otomo T, et al. Effect of Post-nasal Drip on Overnight-cough Frequency and Cough Pattern in Children with Asthma. Tokai J Exp Clin Med. 2022;47(2):79-84.
32. Imai E, Hirai K, Mikami Y, et al. An objective evaluation of nocturnal cough count and cough pattern in children with psychogenic cough. Respiratory Investigation. 2017;55(6):334-337. doi:10.1016/j.resinv.2017.08.008
33. Key A, Holt K, Smith J, Walker P. Objective cough count in bronchiectasis patients is reproducible and relates to quality of life. In: Monitoring Airway Disease. European Respiratory Society; 2019:PA2672. doi:10.1183/13993003.congress-2019.PA2672
34. Lindenhofer M, Roth L, Mädel C, et al. Wheeze and cough measurements at night in children with respiratory symptoms. BMC Pediatr. 2020;20(1):556. doi:10.1186/s12887-020-02455-5
35. Radine A, Werner C, Raidt J, et al. Comparison of Nocturnal Cough Analysis in Healthy Subjects and in Patients with Cystic Fibrosis and Primary Ciliary Dyskinesia: A Prospective Observational Study. Respiration. 2019;97(1):60-69. doi:10.1159/000493323
36. Sinha A, Lee KK, Rafferty GF, et al. Predictors of objective cough frequency in pulmonary sarcoidosis. Eur Respir J. 2016;47(5):1461-1471. doi:10.1183/13993003.01369-2015
37. Spinou A, Lee KK, Sinha A, et al. The Objective Assessment of Cough Frequency in Bronchiectasis. Lung. 2017;195(5):575-585. doi:10.1007/s00408-017-0038-x

38. Sumner H, Woodcock A, Kolsum U, et al. Predictors of Objective Cough Frequency in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2013;187(9):943-949. doi:10.1164/rccm.201211-2000OC
39. Turner R, Birring S, Matos S, Bothamley G. S29 Cough frequency and morbidity in inpatients with acute respiratory disease. Thorax. 2013;68(Suppl 3):A17.2-A18. doi:10.1136/thoraxjnl-2013-204457.36
40. Vertigan AE, Kapela SM, Kearney EK, Gibson PG. Laryngeal Dysfunction in Cough Hypersensitivity Syndrome: A Cross-Sectional Observational Study. The Journal of Allergy and Clinical Immunology: In Practice. 2018;6(6):2087-2095. doi:10.1016/j.jaip.2018.04.015
41. Vertigan AE, Kapela SL, Gibson PG. Chronic cough in Vocal Cord Dysfunction: Description of a clinical entity. Respiratory Medicine. 2020;168:105990. doi:10.1016/j.rmed.2020.105990
42. Wang K, Bettiol S, Thompson MJ, et al. Symptomatic treatment of the cough in whooping cough. Cochrane Acute Respiratory Infections Group, ed. Cochrane Database of Systematic Reviews. Published online September 22, 2014. doi:10.1002/14651858.CD003257.pub5
43. Yousaf N, Monteiro W, Matos S, Birring SS, Pavord ID. Cough frequency in health and disease. European Respiratory Journal. 2013;41(1):241-243. doi:10.1183/09031936.00089312
44. Crooks MG, Hayman Y, Innes A, Williamson J, Wright CE, Morice AH. Objective Measurement of Cough Frequency During COPD Exacerbation Convalescence. Lung. 2016;194(1):117-120. doi:10.1007/s00408-015-9782-y
45. Doenges J, Kuckuck E, Cassel W, et al. Disease control in patients with asthma and respiratory symptoms (wheezing, cough) during sleep. asthma res and pract. 2020;6(1):9. doi:10.1186/s40733-020-00062-w
46. Elghamoudi D, McGuinness K, Smith J, Murray C. S73 Cough frequency and diurnal patterns in children with asthma. In: New Approaches to Characterising Paediatric Respiratory Diseases. BMJ Publishing Group Ltd and British Thoracic Society; 2017:A46-A47. doi:10.1136/thoraxjnl-2017-210983.79
47. Fischer P, Gross V, Kroenig J, et al. Description of nighttime cough epochs in patients with stable COPD GOLD II–IV. COPD. 2018;Volume 13:1071-1078. doi:10.2147/COPD.S154539
48. Harle ASM, Blackhall FH, Molassiotis A, et al. Cough in Patients With Lung Cancer: A Longitudinal Observational Study of Characterization and Clinical Associations. Chest. 2019;155(1):103-113. doi:10.1016/j.chest.2018.10.003
49. Hirai K, Enseki M, Tabata H, et al. Objective measurement of frequency and pattern of nocturnal cough in children with asthma exacerbation. Annals of Allergy, Asthma & Immunology. 2016;117(2):169-174. doi:10.1016/j.anai.2016.06.007
50. Holmes J, McGarvey LPA, Birring SS, Fletcher H, Heaney LG. An observational study to determine the relationship between cough frequency and markers of inflammation in severe asthma. Eur Respir J. 2022;60(6):2103205. doi:10.1183/13993003.03205-2021

51. Koehler U, Hildebrandt O, Fischer P, et al. Time course of nocturnal cough and wheezing in children with acute bronchitis monitored by lung sound analysis. Eur J Pediatr. 2019;178(9):1385-1394. doi:10.1007/s00431-019-03426-4
52. Krönig J, Hildebrandt O, Weissflog A, et al. Long-term Recording of Night-Time Respiratory Symptoms in Patients with Stable COPD II–IV. COPD: Journal of Chronic Obstructive Pulmonary Disease. 2017;14(5):498-503. doi:10.1080/15412555.2017.1338681
53. Kulnik S, Birring S, Rafferty G, Moxham J, Kalra L. P237 Cough frequency in acute stroke: Abstract P237 Table 1. Thorax. 2015;70(Suppl 3):A196.1-A196. doi:10.1136/thoraxjnl-2015-207770.373
54. Lodhi S, Smith JA, Satia I, Holt KJ, Maidstone RJ, Durrington HJ. Cough rhythms in asthma: Potential implication for management. The Journal of Allergy and Clinical Immunology: In Practice. 2019;7(6):2024-2027. doi:10.1016/j.jaip.2018.12.020
55. Mackay A, Donaldson G, Singh R, et al. Utility of Objective Cough Monitoring in Stable COPD and at Exacerbation. American Thoracic Society. Published online 2015:A2539-A2539.
56. Marsden PA, Satia I, Ibrahim B, et al. Objective Cough Frequency, Airway Inflammation, and Disease Control in Asthma. Chest. 2016;149(6):1460-1466. doi:10.1016/j.chest.2016.02.676
57. Ovsyannikov E, Avdeev S, Budnevsky A, Shkatova Y. Influence of Anxiety/Depression on the Subjective Evaluation of Cough in Patients with Chronic Obstructive Pulmonary Disease and Obesity. Medicina. 2019;55(5):134. doi:10.3390/medicina55050134
58. Rassouli F, Tinschert P, Barata F, et al. Characteristics of Asthma-related Nocturnal Cough: A Potential New Digital Biomarker. JAA. 2020;Volume 13:649-657. doi:10.2147/JAA.S278119
59. Rhee H, Belyea MJ, Sterling M, Bocko MF. Evaluating the Validity of an Automated Device for Asthma Monitoring for Adolescents: Correlational Design. J Med Internet Res. 2015;17(10):e234. doi:10.2196/jmir.4975
60. Schwarz SB, Windisch W, Majorski DS, Callegari J, Pläcking M, Magnet FS. Long-Term Auscultation in Chronic Obstructive Pulmonary Disease: Renaissance of an Ideograph of Medical Care. Respiration. 2021;100(3):201-208. doi:10.1159/000513439
61. Turner R, Repossi A, Matos S, Birring S, Bothamley G. S79 Cough Prevalence And Frequency In Pulmonary Tuberculosis. Thorax. 2014;69(Suppl 2):A43-A44. doi:10.1136/thoraxjnl-2014-206260.85
62. Turner R, Matos S, Birring S, Bothamley G. The temporal pattern of coughing in tuberculosis compared to other diseases. In: 10.2 Tuberculosis. European Respiratory Society; 2015:PA4533. doi:10.1183/13993003.congress-2015.PA4533
63. Turner RD, Birring SS, Darmalingam M, et al. Daily cough frequency in tuberculosis and association with household infection. int j tuberc lung dis. 2018;22(8):863-870. doi:10.5588/ijtld.17.0652
64. Weisser S, Lex C, Roth L, et al. Evaluation of nocturnal asthma control (NAC-study): Comparative multicenter analysis of subjective and objective methods. In: 07.02 - Paediatric Asthma and Allergy. European Respiratory Society; 2022:1588. doi:10.1183/13993003.congress-2022.1588

65. Winders T, Nemala S, De Gardner D, Stempel D. Novel Technology to Identify Times of Peak Nighttime Cough Frequency in Severe Asthma. Journal of Allergy and Clinical Immunology. 2023;151(2):AB334. doi:10.1016/j.jaci.2022.12.775
66. Faruqi S, Shiferaw D, Morice AH. Effect of Ivacaftor on Objective and Subjective Measures of Cough in Patients with Cystic Fibrosis. TORMJ. 2016;10(1):105-108. doi:10.2174/1874306401610010105
67. Faruqi S, Sykes DL, Crooks MG, Brindle K, Thompson J, Morice AH. Objective Assessment of Cough: An Early Marker of Response to Biological Therapies in Asthma? Lung. 2020;198(5):767-770. doi:10.1007/s00408-020-00391-w
68. Fukuhara A, Saito J, Birring SS, et al. Clinical Characteristics of Cough Frequency Patterns in Patients with and without Asthma. The Journal of Allergy and Clinical Immunology: In Practice. 2020;8(2):654-661. doi:10.1016/j.jaip.2019.08.053
69. Jung J, Lee K, Song C, Sun Park J. ACC2308. Exploration of Cough Counts and Epochs in Patients with Asthma and Interstitial Lung Disease. Lung. 2023;201(S1):1-6. doi:10.1007/s00408-023-00634-6
70. Kang YR, Oh JY, Lee JH, Small PM, Chung KF, Song WJ. Long-COVID severe refractory cough: discussion of a case with 6-week longitudinal cough characterization. Asia Pacific Allergy. 2022;12(2):e19. doi:10.5415/apallergy.2022.12.e19
71. Lee SE, Rudd M, Kim TH, et al. Feasibility and Utility of a Smartphone Application-Based Longitudinal Cough Monitoring in Chronic Cough Patients in a Real-World Setting. Lung. 2023;201(6):555-564. doi:10.1007/s00408-023-00647-1
72. Shim JS, Kim BK, Kim SH, et al. A smartphone-based application for cough counting in patients with acute asthma exacerbation. J Thorac Dis. 2023;15(7):4053-4065. doi:10.21037/jtd-22-1492
73. Turner R, Birring S, Bothamley G. S37 The response of objectively-measured cough to treatment in tuberculosis: an exploratory study. Thorax. 2015;70(Suppl 3):A25.2-A26. doi:10.1136/thoraxjnl-2015-207770.43
74. Vertigan AE, Kapela SL, Birring SS, Gibson PG. Feasibility and clinical utility of ambulatory cough monitoring in an outpatient clinical setting: a real-world retrospective evaluation. ERJ Open Res. 2021;7(4):00319-02021. doi:10.1183/23120541.00319-2021
75. Zhang M, Brindle K, Robinson M, Ingram D, Cavany T, Morice A. Chronic cough in cystic fibrosis: the effect of modulator therapy on objective 24-h cough monitoring. ERJ Open Res. 2022;8(2):00031-02022. doi:10.1183/23120541.00031-2022
76. Altshuler E, Tannir B, Jolicoeur G, et al. Digital cough monitoring – A potential predictive acoustic biomarker of clinical outcomes in hospitalized COVID-19 patients. Journal of Biomedical Informatics. 2023;138:104283. doi:10.1016/j.jbi.2023.104283
77. Boesch M, Rassouli F, Baty F, et al. Smartphone-based cough monitoring as a near real-time digital pneumonia biomarker. ERJ Open Res. 2023;9(3):00518-02022. doi:10.1183/23120541.00518-2022
78. Crooks MG, Den Brinker AC, Thackray-Nocera S, Van Dinther R, Wright CE, Morice AH. Domiciliary Cough Monitoring for the Prediction of COPD Exacerbations. Lung. 2021;199(2):131-137. doi:10.1007/s00408-021-00435-9

79. Den Brinker AC, Van Dinther R, Crooks MG, Thackray-Nocera S, Morice AH. Alert system design based on experimental findings from long-term unobtrusive monitoring in COPD. Biomedical Signal Processing and Control. 2021;63:102205. doi:10.1016/j.bspc.2020.102205
80. Pekacka-Egli AM, Kazmierski R, Lutz D, et al. Predictive Value of Cough Frequency in Addition to Aspiration Risk for Increased Risk of Pneumonia in Dysphagic Stroke Survivors: A Clinical Pilot Study. Brain Sciences. 2021;11(7):847. doi:10.3390/brainsci11070847
81. Tinschert P, Rassouli F, Barata F, et al. Nocturnal Cough and Sleep Quality to Assess Asthma Control and Predict Attacks. JAA. 2020;Volume 13:669-678. doi:10.2147/JAA.S278155
82. Al Hossain F, Lover AA, Corey GA, Reich NG, Rahman T. FluSense: A Contactless Syndromic Surveillance Platform for Influenza-Like Illness in Hospital Waiting Areas. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2020;4(1):1-28. doi:10.1145/3381014
83. Gabaldón-Figueira JC, Keen E, Giménez G, et al. Acoustic surveillance of cough for detecting respiratory disease using artificial intelligence. ERJ Open Res. 2022;8(2):00053-02022. doi:10.1183/23120541.00053-2022
84. Rahman T, Hossain F, Tonmoy TH, et al. Passive Monitoring of Crowd-Level Cough Counts in Waiting Areas Produces Reliable Syndromic Indicator for Total COVID-19 Burden in a Hospital Emergency Clinic. In Review; 2023. doi:10.21203/rs.3.rs-3084318/v1
85. Deblej Elghamoudi D, Sumner H, McGuiness K, Smith J, Murray C. P241 The feasibility and validity of objective cough monitoring in children using an adult cough detection system. Thorax. 2015;70(Suppl 3):A198.1-A198. doi:10.1136/thoraxjnl-2015-207770.377
86. Gross V, Scholtes M, Sohrabi K, Weissflog A. Validation of the LEOSound®-monitor for standardized detection of cough and wheezing in children. In: World Sleep. ; 2015. http://www.congrex-switzerland.com/fileadmin/files/2015/worldsleep2015/pdf/Worldsleep_2015_-_Online_Abstract_Book_by_session__date_and_time.pdf
87. Huddart S, Asege L, Dang H, et al. Feasibility and Analysis Considerations of Continuous Community-Based Cough Monitoring in Low- and Middle-Income Settings. Public and Global Health; 2022. doi:10.1101/2022.07.06.22277224
88. Rhee H, Miner S, Sterling M, Halterman JS, Fairbanks E. The Development of an Automated Device for Asthma Monitoring for Adolescents: Methodologic Approach and User Acceptability. JMIR mHealth uHealth. 2014;2(2):e27. doi:10.2196/mhealth.3118
89. Urban C, Kiefer A, Conradt R, et al. Validation of the LEOSound® monitor for standardized detection of wheezing and cough in children. Pediatric Pulmonology. 2022;57(2):551-559. doi:10.1002/ppul.25768
90. Shim J, Kim M, Lee SM, et al. An artificial intelligence algorithm-based smartphone application for daily cough monitoring. Allergy. 2023;78(5):1378-1380. doi:10.1111/all.15632
91. Zimmer AJ, Lainati F, Aguilera Vasquez N, et al. Biomarkers That Correlate with Active Pulmonary Tuberculosis Treatment Response: a Systematic Review and Meta-analysis. McAdam AJ, ed. J Clin Microbiol. 2022;60(2):e01859-21. doi:10.1128/jcm.01859-21

92. Ulrik CS, Ringbæk T, Laursen LC, Green A, Brøndum E, Frausing E. Effect of tele health care on exacerbations and hospital admissions in patients with chronic obstructive pulmonary disease: a randomized clinical trial. COPD. Published online September 2015:1801. doi:10.2147/COPD.S85596
93. Al Rajeh A, Steiner MC, Aldabayan Y, et al. Use, utility and methods of telehealth for patients with COPD in England and Wales: a healthcare provider survey. BMJ Open Resp Res. 2019;6(1):e000345. doi:10.1136/bmjresp-2018-000345
94. Anguera Miro X, Bozonnet S, Evans N, Fredouille C, Friedland G, Vinyals O. Speaker Diarization: A Review of Recent Research. IEEE Trans Audio Speech Lang Process. 2012;20(2):356-370. doi:10.1109/TASL.2011.2125954
95. Santosh K, Rasmussen N, Mamun M, Aryal S. A systematic review on cough sound analysis for Covid-19 diagnosis and screening: is my cough sound COVID-19? PeerJ Computer Science. 2022;8:e958. doi:10.7717/peerj-cs.958
96. Ming DK, Sangkaew S, Chanh HQ, et al. Continuous physiological monitoring using wearable technology to inform individual management of infectious diseases, public health and outbreak responses. International Journal of Infectious Diseases. 2020;96:648-654. doi:10.1016/j.ijid.2020.05.086
97. Gauthier-Beaupré A, Grosjean S. Understanding acceptability of digital health technologies among francophone-speaking communities across the world: a meta-ethnographic study. Front Commun. 2023;8:1230015. doi:10.3389/fcomm.2023.1230015

4.12. Supplementary information

Table S4.1. Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	57
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	58
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	59
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	60
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	-
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	61-62
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	60-61; 91-92
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	91-92
Selection of sources of evidence	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	62
Data charting process	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	62-63
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	62
Critical appraisal of individual sources of evidence	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	-
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	62-63
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	63
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	14-19
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	NA
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	65-70
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	72-79
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	709-82
Limitations	20	Discuss the limitations of the scoping review process.	81
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	81

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	-

Table S4.2. Search strategy and results.

			Original search*		Updated search*‡	
Platform	**Database(s)**	**Database coverage dates**	**Date**	**# Results**	**Date**	**# Results**
OvidSP	Ovid MEDLINE ALL(R)	1946 -	2022/09/09	1930	2023/09/19	2131
OvidSP	EMBASE	1996 -	2022/09/09	3191	2023/09/19	3519
Web of Science	e.g., SCI-EXP, CPCI-S, ESCI	1900 -	2022/09/09	3341	2023/09/19	3732
IEEE	IEEEE Xplore	Inception -	2022/09/09	645	2023/09/19	783
Cochrane Library	CENTRAL (Trials)	Inception -	2022/09/09	1160	2023/09/19	880
Europe PMC	Preprints limit	Inception -	2022/09/09	68	2023/09/19	77
TOTAL NUMBER OF RECORDS			**10,335**		**11,122**†	

Search was conducted by librarian Genevieve Gore.
*Limits or filters used: 2013-
†The update search strategy removed the search terms "feature*" and "classif*" to reduce the number of records that were exclusively focused on exploring the use of artificial intelligence algorithms to use cough sounds as a biomarker for disease screening ("cough classification").
‡Before removing duplicates from the original search.

Ovid MEDLINE(R) ALL <1946 to September 19, 2023>

Step	Search term	# Records
1	((cough* or tussis) adj2 (objective or signature* or sound* or biomarker* or marker* or count* or frequenc* or intensit* or pattern* or assess* or measur* or monitor* or record* or analy* or diagnos* or eval* or algorithm*))	3180
2	(*cough/ or (cough* or tussis).ti.) and (objective or signature* or sound* or biomarker* or marker* or count* or frequenc* or intensit* or pattern* or assess* or measur* or monitor* or record* or analy* or diagnos* or eval* or algorithm*).ti	2143
3	1 or 2	4483
4	limit 3 to yr="2013 -Current"	2131

Embase <1996 to 2023 Week 37>

Step	Search term	# Records
1	((cough* or tussis) adj2 (objective or signature* or sound* or biomarker* or marker* or count* or frequenc* or intensit* or pattern* or assess* or measur* or monitor* or record* or analy* or diagnos* or eval* or algorithm*)).mp	4563
2	(*exp coughing/ or (cough* or tussis).ti.) and (objective or signature* or sound* or biomarker* or marker* or count* or frequenc* or intensit* or pattern* or assess* or measur* or monitor* or record* or analy* or diagnos* or eval* or algorithm*).ti	2139

3	1 or 2	5657
4	limit 3 to yr="2013 -Current"	3519

Web of Science Core Collection: SCI-EXPANDED, ESCI, CPCI-S

Step	Search term	# Records
1	(TS=((cough* OR tussis) NEAR/2 (objective OR signature* OR sound* OR biomarker* OR marker* OR count* OR frequenc* OR intensit* OR pattern* OR assess* OR measur* OR monitor* OR record* OR analy* OR diagnos* OR eval* OR algorithm*))) OR ((AK=cough* OR KP=cough* OR TI=(cough* OR tussis)) AND TI=(objective OR signature* OR sound* OR biomarker* OR marker* OR count* OR frequenc* OR intensit* OR pattern* OR assess* OR measur* OR monitor* OR record* OR analy* OR diagnos* OR eval* OR algorithm*))	3732
2	Limit to 2013-	

IEEE (IEEE Xplore)

Step	Search term	# Records
1	("All Metadata":cough* OR "All Metadata":tussis)	783
2	Filters Applied: 2013-2023	

CENTRAL (Cochrane Library)

Step	Search term	# Records
1	((cough*:ti,ab,kw OR tussis:ti,ab,kw) NEAR/2 (objective:ti,ab,kw OR signature*:ti,ab,kw OR sound*:ti,ab,kw OR biomarker*:ti,ab,kw OR marker*:ti,ab,kw OR count*:ti,ab,kw OR frequenc*:ti,ab,kw OR intensit*:ti,ab,kw OR pattern*:ti,ab,kw OR assess*:ti,ab,kw OR measur*:ti,ab,kw OR monitor*:ti,ab,kw OR record*:ti,ab,kw OR analy*:ti,ab,kw OR diagnos*:ti,ab,kw OR eval*:ti,ab,kw OR algorithm*:ti,ab,kw))) OR (([mh cough] OR (cough*:ti OR tussis:ti)) AND (objective:ti OR signature*:ti OR sound*:ti OR biomarker*:ti OR marker*:ti OR count*:ti OR frequenc*:ti OR intensit*:ti OR pattern*:ti OR assess*:ti OR measur*:ti OR monitor*:ti OR record*:ti OR analy*:ti OR diagnos*:ti OR eval*:ti OR algorithm*:ti)) with Publication Year from 2013 to 2023, in Trials	880
2	#1 NOT "Trial registry record":pt	

Preprints (Europe PMC)

Step	Search term	# Records
1	(TITLE:(cough* OR tussis) AND TITLE:(objective OR signature* OR sound* OR biomarker* OR marker* OR count* OR frequenc* OR intensit* OR pattern* OR assess* OR measur* OR monitor* OR record* OR analy* OR diagnos* OR eval* OR algorithm*)) AND (SRC:PPR)	77

Table S4.3. Study details for correlation with patient-reported outcomes (PROs)

PRO	Author (Year)	Ref.	Cough tool	Patient group	Sample size	Rho (p-value)
VAS	Bisballe-Müller (2021)	1	Sony ICD-PX470 Digital Voice Recorder	Asthma	24	0.26 (0.16)
				ARIs	25	0.53 (0.003)
				Bronchiolitis	35	0.56 (<0.001)
				Pneumonia	30	0.52 (<0.001)
	Fletcher (2017)	2	LCM	Asthma, RCC, GORD, Rhinosinusitis	320	0.43 (0.001)
	Ovsyannikov (2019)	3	Custom	COPD, normal weight, no anxiety/depression	30	0.42 (0.0231)
				COPD, obese, no anxiety/depression	33	0.44 (0.003)
				COPD, normal weight, with anxiety/depression	23	-0.38 (0.005)

				COPD, obese, with anxiety/depression	24	-0.40 (0.012)
	Rassouli (2020)	4	Clara app.	Asthma	94	0.38 (<0.001)
	Rhee (2015)	5	ADAM device	Asthma (day)	42	0.38 (0.02)
				Asthma (night)	42	0.40 (0.02)
	Shim (2023)	6	Coughy app.	Asthma (day)	168	0.31 (<0.001)
				Asthma (night)	168	0.34 (<0.001)
	Sinha (2016)	7	LCM	Sarcoidosis	32	0.62 (0.001)
	Spinou (2017)	8	LCM	Bronchiectasis	54	0.54 (<0.001)
	Sumner (2013)	9	VitaloJAK	COPD (night)	89	0.48 (<0.001)
				COPD (day)	89	0.66 (<0.001)
	Turner (2013)	10	LCM	Asthma, COPD, LRTI	40	0.33 (0.05)
	Turner (2014)	11	LCM	Tuberculosis	108	0.60 (0.001)
	Yousaf (2013)	12	LCM	Asthma, Bronchitis, Chronic cough, COPD	78	0.66 (<0.001)
VCD	Bisballe- Müller (2021)	1	Sony ICD-PX470 Digital Voice Recorder	Asthma (day)	25	0.50 (0.001)
				Asthma (night)	11	-0.36 (0.26)
				ARIs (night)	16	0.39 (0.15)
				ARIs (day)	25	0.34 (0.054)
				Bronchiolitis (day)	34	0.33 (0.04)
				Bronchiolitis (night)	25	0.65 (<0.001)
				Pneumonia (day)	30	0.45 (0.002)
				Pneumonia (night)	26	0.44 (0.02)
	Lindenhofer (2020)	13	LEOSound	Asthma, Cystic fibrosis, Pneumonia, Habit cough, Chronic cough (night)	34	0.42 (<0.01)
				Asthma, Cystic fibrosis, Pneumonia, Habit cough, Chronic cough (day)	35	0.46 (0.01)
LCQ	Crooks (2016)	14	HACC	COPD	18	-0.44 (0.001)
	Fletcher (2017)	2	LCM	Asthma, RCC, GORD, Rhinosinusitis	320	-0.45 (0.001)
	Lee (2023)	15	Hyfe Cough Tracker app	Chronic cough	25	-0.57 (0.001)
	Sinha (2016)	7	LCM	Sarcoidosis	32	-0.61 (0.001)
	Spinou (2017)	8	LCM	Bronchiectasis	54	-0.52 (0.001)
	Sumner (2013)	9	VitaloJAK	COPD	89	-0.64 (0.001)
	Turner (2013)	10	LCM	Asthma, COPD, LRTI	40	-0.13 (0.48)
	Vertigan (2021)	16	LCM	Asthma, RCC, Laryngeal obstruction	112	-0.43 (<0.001)
	Yousaf (2013)	12	LCM	Asthma, Bronchitis, Chronic cough, COPD	78	-0.44 (0.001)
PC-QoL	Bisballe- Müller (2021)	1	Sony ICD-PX470 Digital Voice Recorder	Asthma	25	-0.08 (0.8)
				ARIs	25	-0.29 (0.16)
				Bronchiolitis	35	-0.17 (0.31)
				Pneumonia	29	-0.05 (0.73)
	Koehler (2019)	17	LEOSound	Acute bronchitis (cough epochs)	36	-0.43 (<0.001)
	Lindenhofer (2020)	13	LEOSound	Asthma, Cystic fibrosis, Pneumonia, Habit cough, Chronic cough	7	-0.26 (0.56)

ADAM, automated device for asthma monitoring; ARI, acute respiratory infection; COPD, chronic obstructive pulmonary disease; HAC, Hull Automated Cough Counter; LCM, GORD, gastro-oesophageal reflux disease; Leicester Cough Monitor; LRTI, lower respiratory tract infection; PRO, patient-reported outcome; RCC, refractory chronic cough

4.13. Supplementary references

1. Bisballe-Müller N, Chang AB, Plumb EJ, Oguoma VM, Halken S, McCallum GB. Can Acute Cough Characteristics From Sound Recordings Differentiate Common Respiratory Illnesses in Children? Chest. 2021;159(1):259-269. doi:10.1016/j.chest.2020.06.067
2. Fletcher H, Cho PSP, Matos S, et al. Chronic cough: objective cough frequency in a large cough clinic cohort. In: Clinical Problems Asthma. European Respiratory Society; 2017:OA4677. doi:10.1183/1393003.congress-2017.OA4677
3. Ovsyannikov E, Avdeev S, Budnevsky A, Shkatova Y. Influence of Anxiety/Depression on the Subjective Evaluation of Cough in Patients with Chronic Obstructive Pulmonary Disease and Obesity. Medicina. 2019;55(5):134. doi:10.3390/medicina55050134
4. Rassouli F, Tinschert P, Barata F, et al. Characteristics of Asthma-related Nocturnal Cough: A Potential New Digital Biomarker. JAA. 2020;Volume 13:649-657. doi:10.2147/JAA.S278119
5. Rhee H, Belyea MJ, Sterling M, Bocko MF. Evaluating the Validity of an Automated Device for Asthma Monitoring for Adolescents: Correlational Design. J Med Internet Res. 2015;17(10):e234. doi:10.2196/jmir.4975
6. Shim J, Kim M, Lee SM, et al. An artificial intelligence algorithm-based smartphone application for daily cough monitoring. Allergy. 2023;78(5):1378-1380. doi:10.1111/all.15632
7. Sinha A, Lee KK, Rafferty GF, et al. Predictors of objective cough frequency in pulmonary sarcoidosis. Eur Respir J. 2016;47(5):1461-1471. doi:10.1183/13993003.01369-2015
8. Spinou A, Lee KK, Sinha A, et al. The Objective Assessment of Cough Frequency in Bronchiectasis. Lung. 2017;195(5):575-585. doi:10.1007/s00408-017-0038-x
9. Sumner H, Woodcock A, Kolsum U, et al. Predictors of Objective Cough Frequency in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2013;187(9):943-949. doi:10.1164/rccm.201211-2000OC
10. Turner R, Birring S, Matos S, Bothamley G. S29 Cough frequency and morbidity in inpatients with acute respiratory disease. Thorax. 2013;68(Suppl 3):A17.2-A18. doi:10.1136/thoraxjnl-2013-204457.36
11. Turner R, Repossi A, Matos S, Birring S, Bothamley G. S79 Cough Prevalence And Frequency In Pulmonary Tuberculosis. Thorax. 2014;69(Suppl 2):A43-A44. doi:10.1136/thoraxjnl-2014-206260.85
12. Yousaf N, Monteiro W, Matos S, Birring SS, Pavord ID. Cough frequency in health and disease. Eur. Respir. J. 2013;41(1):241-243. doi:10.1183/09031936.00089312
13. Lindenhofer M, Roth L, Mädel C, et al. Wheeze and cough measurements at night in children with respiratory symptoms. BMC Pediatr. 2020;20(1):556. doi:10.1186/s12887-020-02455-5

14. Crooks MG, Hayman Y, Innes A, Williamson J, Wright CE, Morice AH. Objective Measurement of Cough Frequency During COPD Exacerbation Convalescence. Lung. 2016;194(1):117-120. doi:10.1007/s00408-015-9782-y
15. Lee SE, Rudd M, Kim TH, et al. Feasibility and Utility of a Smartphone Application-Based Longitudinal Cough Monitoring in Chronic Cough Patients in a Real-World Setting. Lung. 2023;201(6):555-564. doi:10.1007/s00408-023-00647-1
16. Vertigan AE, Kapela SL, Birring SS, Gibson PG. Feasibility and clinical utility of ambulatory cough monitoring in an outpatient clinical setting: a real-world retrospective evaluation. ERJ Open Res. 2021;7(4):00319-02021. doi:10.1183/23120541.00319-2021
17. Koehler U, Hildebrandt O, Fischer P, et al. Time course of nocturnal cough and wheezing in children with acute bronchitis monitored by lung sound analysis. Eur J Pediatr. 2019;178(9):1385-1394. doi:10.1007/s00431-019-03426-4

Chapter 5. Manuscript III: Baseline tuberculosis bacterial burden as a predictor of patient cough trajectory during the first two weeks of anti-tuberculosis therapy

5.1. Preface

Our scoping review demonstrated that while numerous technologies exist for longitudinal cough monitoring across various diseases, their implementation in clinical settings remains limited. A key barrier to clinical adoption is our incomplete understanding of longitudinal cough dynamics, coupled with insufficient analytical tools to meaningfully interpret the collected data. This challenge is further complicated by the substantial heterogeneity in cough patterns, which varies both between individuals and within the same individual over time.

Within the reviewed literature, several studies investigated longitudinal cough dynamics among TB patients. However, these studies were constrained by intermittent monitoring approaches and failed to adequately address the analytical complexities inherent in continuous cough monitoring data. More broadly, the field of digital cough monitoring lacks standardized methods for managing data quality issues that arise from continuous monitoring.

In this study, we aimed to enhance the understanding of cough dynamics during TB treatment through a comprehensive analysis of continuously collected cough data during the initial two weeks of anti-TB treatment. Utilizing prospectively collected data from multiple countries, we investigated the relationship between cough frequency and TB disease severity while addressing key measurement challenges, including missing data, zero-inflation, and potential over-recording of coughs from others. This analysis establishes a foundation for future studies seeking to validate cough as a biomarker for TB treatment response.

5.2. Title page

Baseline tuberculosis bacterial burden as a predictor of patient cough trajectory during the first two weeks of anti-tuberculosis therapy

Zimmer AJ[1,2], Raberahona M[3,4], Jaganath D[5,6], Rakotoarivelo R[3,4], Christopher DJ7, Nhung NV[8], Theron G[9], Worodria W[10], Yu C[11], Nahid P[5,12], Denkinger CM[13,14], Benedetti A[1,2,15], Pai M[1,2,16], Cattamanchi A[5,17], Grandjean Lapierre S[2,18,19], Huddart S[5,20]

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

2 McGill International TB Centre, Montreal, Canada

3 Infectious Diseases Department, University Hospital Joseph Raseta Befelatanana, Antananarivo, Madagascar

4 Centre d'Infectiologie Charles Mérieux, University of Antananarivo, Madagascar

5 Center for Tuberculosis, University of California San Francisco, San Francisco, USA

6 Division of Pediatric Infectious Diseases, University of California San Francisco, San Francisco, California, USA

7 Department of Pulmonary Medicine, Christian Medical College, Vellore (Ranipet Campus), India

8 Vietnam National Tuberculosis Control Program, Hanoi, Vietnam

9 DSI-NRF Centre of Excellence for Biomedical Tuberculosis Research, and SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

10 Department of Medicine, Mulago National Referral Hospital, Kampala, Uganda

11 De La Salle Medical and Health Sciences Institute, Center for Tuberculosis Research, City of Dasmariñas, Cavite, The Philippines

12 Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, University of California San Francisco, San Francisco, USA

13 Division of Infectious Diseases and Tropical Medicine, Heidelberg University Hospital, Heidelberg, Germany

[14] German Center for Infection Research (DZIF), Partner Site, Heidelberg, Germany

[15] Centre for Outcomes Research & Evaluation, Research Institute of the McGill University Health Centre, Montréal, Canada

[16] School of Population and Global Health, McGill University, Montreal, Canada

[17] Division of Pulmonary Diseases and Critical Care Medicine, University of California Irvine, Irvine, USA

[18] Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Canada

[19] Department of microbiology, Immunology and Infectious Diseases, Université de Montréal, Canada

[20] Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, California, USA

5.3. Abstract

Introduction

Cough is a common symptom among people with pulmonary tuberculosis (PTB). Cough counts are expected to decrease after the commencement of anti-TB treatment, sparking interest in using cough counts as an objective marker of treatment response. Given that cough may be influenced by various factors, including TB disease bacterial burden, understanding this relationship could help interpret cough dynamics throughout anti-TB treatment. This study aims to characterize the relationship between daily cough counts during the first 14 days of treatment and baseline PTB burden.

Methods

People with microbiologically confirmed PTB were enrolled from four countries (Madagascar, Uganda, Philippines, and Vietnam). Participants continuously recorded their cough counts using the Hyfe smartphone application during the first 14 days of anti-TB treatment. The association between cough counts and TB bacterial burden—measured by categorical GeneXpert MTB/RIF Ultra (Xpert) semi-quantitative results and continuous CAD4TB scores—was analyzed using a zero-inflated negative binomial model. Sensitivity analyses examined how measurement error from excess coughs recorded by non-participants may have influenced this association.

Results

Among the 209 included participants, baseline cough counts were highest among individuals with markers of elevated bacterial burden, as indicated by either a 'High' Xpert result or a high CAD4TB score. Cough counts declined over the 14-day period for all participants. Multivariable analyses of Xpert semi-quantitative results found a notable decrease in rate ratios (RR) of cough frequency as Xpert semi-quantitative results decreased from 'High' to ‘Xpert Trace/Negative but Culture Positive’. Similarly, individuals with lower CAD4TB scores coughed less frequently than those with higher scores (RR: 0.80; 95% confidence interval: 0.67, 0.95). Sensitivity analyses indicated

that accounting for potential measurement errors from excess coughs slightly strengthened the association between markers of bacterial burden and cough frequency.

Discussion

Despite its heterogenous presentation, cough appeared to be correlated with markers of baseline bacterial burden and declined with TB treatment. This study offers insights into the variability of cough counts with respect to important baseline variables, thereby contributing to the advancement of cough-based indicators for evaluating treatment response. Future studies should validate cough as a TB treatment response marker.

5.4. Introduction

Tuberculosis (TB), caused by the bacterium *Mycobacterium tuberculosis* (*Mtb*), remains the leading infectious disease killer worldwide. Despite diagnostic and treatment advancements, TB continues to impose a significant burden, with an estimated 10.8 million incident cases and 1.25 million deaths globally in 2023.[1]

Cough is a cardinal symptom of pulmonary TB (PTB). Prolonged cough (≥2 weeks) is commonly used as a clinical screening marker.[2] Coughing is also widely considered to facilitate the transmission of *Mtb* through aerosolization.[3] Moreover, early studies observed a decline in cough frequency during the course of anti-TB treatment.[4]

Current PTB treatment monitoring tools have a limited ability to identify individuals who are not responding to treatment. The World Health Organization (WHO) recommends sputum smear microscopy and sputum culture conversion as microbiological markers of treatment response,[5] but these are resource-intensive, require sputum production, and have poor sensitivity and specificity for informing treatment outcome.[6] There is a need for novel biomarkers that are low-cost, minimally invasive, and deployable at multiple time points and in community settings. The WHO has published a target product profile that can guide new tool development for this use case.[7] While many lab-based biomarkers have been evaluated,[8] there is growing interest in novel methods to objectively monitor clinical signs and symptoms, including cough.[9,10]

Subjective cough assessment tools, such as the Leicester Cough Questionnaire for assessing cough severity and the Cough Quality-of-Life Questionnaire for evaluating quality of life, have been used in the context of TB care.[4,11,12] The development of semi-autonomous ambulatory cough monitors enabled objective cough monitoring. These devices, composed of a microphone and digital recorder worn by the patient, include the Leicester Cough Monitor and the Cayetano Cough Monitor, both capable of continuous monitoring for up to 24 hours.[13–16] Less obtrusive solutions have emerged more recently, such as cough-counting smartphone applications.[17,18] With these advancements, both in terms of hardware and software accuracy and reliability, longer monitoring periods can be achieved with greater accuracy. This renewed interest in cough as a TB digital

biomarker has spurred artificial intelligence (AI)-based cough classifiers for screening and cough counting applications for treatment monitoring.[15–21]

Given the trend of cough remission observed during PTB treatment, there is substantial interest in using cough as a non-intrusive marker for monitoring treatment response. However, to fully understand how cough can serve this role, it is necessary to better understand the complexities and variability of cough data.[22] Cough counts are highly variable both between individuals and within the same individual over time, making it difficult to assess treatment response on an individual level. This heterogeneity may be influenced by several demographic and clinical factors, such as age, sex, comorbidities (e.g., HIV infection or chronic obstructive pulmonary disease), and smoking habits. Bacterial burden, referring to the concentration of *Mtb* bacteria in sputum, may correlate with cough frequency.[23] Given that bacterial burden influences treatment response and outcomes,[24] understanding these relationships is essential for optimizing the use of cough as a biomarker for treatment monitoring.

This study aims to characterize the relationship between cough frequency during the first 14 days of treatment and markers of baseline bacterial burden. Bacterial burden is assessed using microbiological (GeneXpert [Xpert] Ultra PCR semi-quantitative results) and radiological (digital chest X-rays computer-aided detection [CAD] prediction score) methods. While this study does not seek to validate cough as a biomarker for TB treatment monitoring, it aims to deepen the understanding of cough dynamics—including persistence, frequency, and temporal patterns—and their relationship to underlying disease severity.

5.5. Methods

5.5.1. Setting

Participants were enrolled from the Rapid Research in Diagnostic Development for TB Network (R2D2 TB Network) clinical study and an independent cough study in Madagascar. The R2D2 TB Network, a multinational study across ten countries, investigates novel TB diagnostic technologies, including digital cough monitoring. Study participants were often co-enrolled in

complementary diagnostic studies.[25] This analysis included data from R2D2 TB Network participants in Uganda, Philippines, and Vietnam, where sufficiently large cohorts underwent longitudinal cough monitoring. Study procedures were standardized across all sites.

5.5.2. Study population and inclusion criteria

The studies recruited adults 18 years or older from outpatient facilities experiencing a new or worsening cough for ≥2 weeks. Individuals who lived more than 20 kilometers from a study site, had taken anti-TB medication in the last year, or had taken any medication with anti-mycobacterial activity within the prior 2 weeks were excluded. For this analysis, only participants who had had microbiologically confirmed PTB were included, defined as a positive result from sputum Xpert MTB/RIF Ultra (Xpert, Cepheid, USA) and/or sputum culture (MGIT, 7H10 agar, or Löwenstein–Jensen [LJ]). All participants were initiated on anti-TB therapy and were treated as per national TB guidelines with the standard 6-month regimen (2 months of isoniazid, rifampicin, pyrazinamide and ethambutol, followed by continuation phase of 4 months of isoniazid and rifampicin) for drug susceptible TB. Drug-resistance for rifampicin (RIF) was evaluated using Xpert MTB/RIF Ultra. All individuals were RIF negative.

5.5.3. Participant enrollment procedures

Participants that met the inclusion criteria were enrolled in the R2D2 TB Network and Madagascar Cough study upon presentation to the health facility. Up to three sputum samples were collected for repeated Xpert and culture testing. Cough monitoring was initiated on the day of enrollment. For R2D2 sites, TB treatment was initiated a median of 1 (IQR: 0, 2.5) day after study enrollment.[17] For Madagascar, TB treatment was initiated a median of 2 (IQR: 2, 4) days after study enrollment. To ensure uniformity in time on treatment, the first day of follow-up was analytically set to the treatment initiation date.

5.5.4. TB severity assessment

Xpert is a nucleic acid amplification test (NAAT) used to diagnose TB that provides semi-quantitative estimate of bacterial load ('High', 'Medium', 'Low', and 'Very low').[26,27] These semi-quantitative categories correlate with cycle threshold (Ct) values from PCR testing, which detect the expression of specific genes in *Mtb*, including *IS6110*, *IS1081*, and at least two *rpoB* genes.[28] Previous studies have demonstrated that Ct values inversely correlate with semi-quantitative categories, with lower Ct values (indicating higher bacterial loads) corresponding to 'High' results, and progressively higher Ct values corresponding to 'Medium', 'Low', and 'Very Low' categories.[26,27,29] 'Trace' results occur when one or both probes for *IS6110* and *IS1081* targets are positive (Ct <37), and no more than one *rpoB* probe shows a Ct <40.[28] 'Trace' results indicate extremely low TB levels and are difficult to interpret.[30] Therefore, participants with a 'Trace' or negative Xpert result but with a positive TB culture result were categorized as 'Xpert Trace/Negative & Culture Positive'.

Participants had their digital chest X-rays taken at time of enrollment, prior to initiating treatment, which were analyzed using CAD4TB version 7 software (Delft Imaging Systems, Netherlands). CAD4TB is a commercial software that uses deep learning to screen for PTB from chest x-rays.[31] The results are interpreted as a score output, ranging from 0-100, with a higher score representing a higher likelihood of PTB. The accuracy of CAD4TB was recently evaluated among R2D2 participating countries, achieving a specificity of 74% (95% CI: 72%, 75%) at 90% sensitivity.[32] While CAD4TB is primarily a diagnostic tool, some evidence suggests its potential utility as a severity indicator. A prior study have shown that higher CAD4TB scores significantly correlate with bacterial load ($p<0.001$), as measured by sputum Xpert semi-quantitative scores.[33] Another study found that CAD4TB sensitivity improved with higher bacillary burden, indicating a positive correlation between TB scores and bacterial load.[34] However, it's important to note that these correlations do not necessarily establish CAD4TB scores as direct measures of disease severity, as radiographic findings may reflect cumulative lung damage rather than current bacterial activity. In our study, we explore these relationships further to assess whether these different measurements (Xpert semi-quantitative score, CAD4TB score) provide complementary information about TB severity and its association with cough frequency over time.

5.5.5. Cough monitoring

Participants received a study-provided smartphone with the Hyfe Research application pre-installed.[35,36] The Hyfe Research app continuously listens for explosive sounds like coughs without continuously recording every sound. If detected, the embedded AI algorithm records and classifies these explosive sounds as cough or not with a sensitivity of 91% and specificity of 98%.[36] Less than 0.5 seconds of each explosive sound is recorded, ensuring that acoustic environments and conversations are not recorded, and that the participant's privacy is protected.[36]

Smartphone models varied by country, details of which have been reported for R2D2 participating countries.[37] In Madagascar, the Motorola G9 Play model was used. The smartphones were configured such that only the Hyfe Research app was allowed to run, preventing participants from downloading other apps or making calls. For 14 consecutive days, participants were instructed to carry the smartphone in a pouch around their neck, directing the microphone towards their face (**Figure 5.1**). They were instructed to keep the device with them as much as possible to ensure continuous cough recording, including placing it near their head while sleeping. Participants were asked to charge the phone nightly and were shown how to pause and restart recording when needed.[37]

Figure 5.1. Image of participant wearing study smartphone around their neck. Smartphone microphone is pointed towards the participant's face. Participants were instructed to wear the smartphone in the pouch as often as possible throughout their day.

5.5.6. Cough data preparation and treatment of missing data

All cough sounds detected from the Hyfe app are associated with a timestamp. All cough timestamps were adjusted to reflect the local time zone in the country where the data was collected. A full day of follow-up was 12pm (noon) until 11:59am the following day in the local time zone. Data on recording activity was also available. This data reported timestamps corresponding to periods of active and inactive recording sessions by the Hyfe Research app during the 14-day follow-up. Cough recording completeness was assessed at the hourly level to address within-day missingness and prevent underestimation of daily cough counts. Cough counts were pro-rated if ≥30 minutes of recording were observed during the hour. Hours with <30 minutes of recording were set as missing. Missing hourly cough counts were imputed using multiple imputation by chained equation (MICE) (**Figure S5.1**).[38] Imputations were performed at the hourly level to account for the missingness that occurred during specific hours of the day. A multi-level multiple imputation with predictive mean matching was done using the mice (version 3.16.0) and miceadds (version 3.17.44) packages in R.[38,39] Within the MICE model, clustering at the individual level was accounted for, along with random effects for day, harmonic sine and cosine terms, and interactions between day and harmonic terms. Harmonic terms were included to account for the circadian trend that cough tends to follow (**Figure S5.2**).[15] Data was also imputed for several person-level predictors, including history of TB (0.5% missing), HIV-status (1.5% missing), and baseline CAD scores (3.9% missing). Overall, 28% of hourly cough data was missing and a total of 30 datasets were imputed and their analytical results combined. Density plots comparing observed and imputed distributions are provided in **Figure S5.3**

5.5.7. Statistical analysis

All statistical analyses were conducted using R software (version 4.3.3). Participants were stratified by Xpert Ultra semi-quantitative results and the median coughs per hour (cph) per day of follow-up was reported. The metric of median cph was previously used to reported longitudinal trends in cough rates.[17,40] Spearman's rank correlation test was used to evaluate the relationships between baseline CAD4TB and total cough counts on each day of follow-up.

Cough count data is known to be over-dispersed and zero-inflated.[22,41] This distribution was also observed in this dataset (**Figure S5.4**), therefore, a zero-inflated negative binomial distribution was used for models. Different random effects and correlation structures, including autoregressive 1 (AR1), exchangeable, and Toeplitz were evaluated using Maximum Likelihood estimation and compared.[42] The final model was selected based on the Akaike Information Criterion (AIC) and residual assessment using the DHARMa package (**Table S5.1, Figure S5.5, Figure S5.6**).[43]

: Xpert semi-quantitative result (categorical) and CAD4TB score (continuous). Both models employed a random-effects zero-inflated negative binomial generalized linear mixed model (GLMM) for daily cough count.[42] An AR1 correlation structure was included for the random slope of the day, which accounts for the temporal dependence in the data, where observations close in time are not independent (**Figure S5.7**). Participant-level random intercepts were fit to account for individual variability in the probability of excess zeros. Participant-level fixed effects adjusted for pre-determined confounders between markers of TB severity and cough frequency, including sex, age, country, smoking status, HIV status, COVID-19 status, and TB history. A quadratic term for day of follow-up was included in both the count and zero-inflation components to account for the non-linear progression of cough over time. Final models were fit using restricted maximum likelihood (REML) estimation to provide unbiased estimates of the variance components. The models output rate ratios (RRs), where lower RRs indicate that lower predictor values (such as lower CAD4TB scores or lower Xpert semi-quantitative categories) associated with decreased cough frequency. For the Xpert and the CAD4TB models, the glmmTMB (version 1.1.9) package was used.[44]

5.5.8. Sensitivity analysis

Overestimation of cough counts can occur when cough sounds from individuals other than the participant are recorded. Differentiating between the participant's cough’s acoustic signature from that of other individuals is currently not possible. To address this, a sensitivity analysis was conducted to assess the impact of censoring hourly cough frequencies above different percentile thresholds. Daily cough counts were capped at the 50th, 75th, and 90th percentile per day within

Xpert semi-quantitative groups. These percentiles were chosen to represent a range of censoring intensities. Hourly cough counts that were greater than the percentile were right censored at the percentile. For instance, the 90th percentile for hourly cough counts on day 1 for the 'High' Xpert group was 64 cph. Therefore, any hourly cough counts >64 cph were set as 64 cph. This process was repeated for each day and each Xpert group. These right-censored hourly datasets were summed to create daily totals. The previously described GLMM, was then applied to these censored datasets for evaluating the association between cough frequency and baseline TB burden.

5.5.9. Ethics statement

Ethical approval for this study was obtained from institutional review boards (IRB) in the US (University of California San Francisco IRB #20-32670), Canada (Centre de Recherche du Centre Hospitalier de l'Université de Montréal IRB #2021-9270, 20.226) and each study site: Vietnam (Ministry of Health Ethical Committee for National Biological Medical Research IRB #94/CN-HĐĐĐ, National Lung Hospital Ethical Committee for Biological Medical Research IRB #566/2020/NCKH and the Hanoi Department of Health, Hanoi Lung Hospital Science and Technology Initiative Committee IRB #22/BVPHN); India (Christian Medical College IRB #13256); South Africa (Stellenbosch University Health Research Ethics Committee #17047); Uganda (Makerere University College of Health Sciences School of Medicine Research Ethics Committee #2020-182); the Philippines (De La Salle Health Sciences Institute Independent Ethics Committee #2020-33-02-A); and Madagascar (Comité d'Éthique à la Recherche Biomédicale #IORG0000851 - N°051-MSANP/SG/AMM/CERBM).

5.6. Results

5.6.1. Participant information

A total of 209 individuals with microbiologically confirmed PTB were included in this analysis. Baseline demographic and clinical characteristics, categorized by sputum Xpert semi-quantitative results, are summarized in **Table 5.1**. A third of participants (71/209, 34.0%) had a 'High' Xpert result, indicative of a high baseline TB burden. Of these 'High' Xpert result individuals, the

majority (40/71, 56.3%) were from Madagascar. Individuals with higher sputum bacterial burden presented with more severe baseline clinical symptoms, including lower BMIs, higher temperatures, longer duration of cough, and elevated heart rates. The severity of these symptoms diminished with decreasing Xpert semi-quantitative result. CAD analyses for TB detection using CAD4TB displayed similar trends, with higher CAD scores in the ‘High’ Xpert group.

Table 5.1. Baseline demographic and clinical characteristics by GeneXpert semi-quantitative result.

	High (N=71)	Medium (N=48)	Low (N=39)	Very low (N=23)	Xpert Trace/Neg. & Culture pos. (N=24)	Overall (N=209)
Female	28 (39.4)	18 (37.5)	15 (38.5)	8 (34.8)	11 (45.8)	80 (39.0)
Age (years)	29.0 (23.0, 41.5)	40.0 (25.0, 51.0)	30.0 (22.0, 45.5)	41.0 (33.5, 59.0)	43.0 (29.5, 62.0)	33.0 (25.0, 50.0)
Country						
Madagascar	40 (56.3)	13 (27.1)	10 (25.6)	3 (13.0)	5 (20.8)	71 (34.6)
Philippines	2 (2.8)	5 (10.4)	7 (17.9)	7 (30.4)	7 (29.2)	28 (13.7)
Uganda	22 (31.0)	18 (37.5)	9 (23.1)	7 (30.4)	6 (25.0)	62 (30.2)
Vietnam	7 (9.9)	12 (25.0)	13 (33.3)	6 (26.1)	6 (25.0)	44 (21.5)
Smoked in past 7 days	10 (14.1)	6 (12.5)	8 (20.5)	1 (4.3)	4 (16.7)	29 (14.1)
HIV positive	4 (5.6)	1 (2.1)	2 (5.1)	3 (13.0)	4 (16.7)	14 (6.8)
Missing; n (%)	*1 (1.4)*	*1 (2.1)*	*1 (2.6)*	*0 (0.0)*	*0 (0.0)*	*3 (1.5)*
COVID-19 positive	8 (11.3)	1 (2.1)	5 (12.8)	1 (4.3)	2 (8.3)	17 (8.3)
Prior TB	4 (5.6)	6 (12.5)	7 (17.9)	2 (8.7)	3 (12.5)	22 (10.5)
Missing	*0 (0.0)*	*0 (0.0)*	*1 (2.6)*	*0 (0.0)*	*0 (0.0)*	*1 (0.5)*
BMI (kg/m^2)	18.3 (16.6, 20.2)	18.7 (17.5, 20.4)	18.7 (17.0, 21.6)	19.6 (18.0, 23.2)	19.8 (17.6, 21.6)	18.7 (17.3, 20.9)
MUAC (mm)	235 (220, 250)	239 (226, 254)	227 (210, 252)	240 (232, 266)	239 (219, 265)	235 (220, 254)
Temperature (°C)	37.2 (36.7, 37.5)	36.8 (36.6, 37.1)	36.7 (36.5, 37.0)	36.8 (36.5, 36.9)	36.8 (36.4, 37.1)	36.8 (36.5, 37.4)
Cough days	45 (30, 90)	45 (30, 90)	30 (30, 90)	30 (16, 64)	29 (15, 58)	35 (25, 90)
Heart rate (bpm)	98 (83, 120)	94 (75, 105)	90 (80, 100)	91 (84, 108)	89 (82, 95)	95 (80, 109)
CAD4TB score	83.5 (73.1, 93.8)	84.0 (71.9, 91.7)	80.6 (69.3, 90.2)	45.5 (15.1, 67.0)	47.1 (29.2, 72.7)	78.4 (61.9, 90.7)
Missing, n (%)	*3 (4.2)*	*2 (4.2)*	*3 (7.7)*	*0 (0.0)*	*0 (0.0)*	*8 (3.9)*

Categorical variables are reported as count (percentage) and continuous variables are presented as median (Q1, Q3).
BMI, body mass index; bpm, beats per minute; C, Celsius; MUAC, mid-upper arm circumference; n, number

5.6.2. Cough data quality control

The anticipated cumulative number of recording hours was 72,224 hours, representing the 24h of recording over 14 days of follow-up for all 209 eligible participants (24*14*209). Of these hours, 70.7% had complete recording while 27.9% were completely missing. Partial missingness (<1h recording) was observed in 1.4% of recording hours. Cough counts were pro-rated if ≥30 minutes of recording were observed during the hour (0.9% of recordings). The remaining 0.5% of partial hourly cough recordings were set as missing.

Participants recorded for a median of 21 hours a day. Overall, 38% of participants had at least one hour of daily recording throughout the entire 14-day follow-up period while 70% of individuals achieved one hour or more of recording for at least of 7 days during the study. A gradual increase in missing recording hours was observed as the study progressed. The proportion of missing hours of recording rose from 16% on the first day of monitoring to 46% by day 14 (**Figure S5.1**).

5.6.3. Overview of cough trends

Baseline (day 1) cough frequency positively correlated with TB bacterial burden. Individuals with a 'High' Xpert result had an elevated median baseline frequency of 17.5 cph (interquartile range [IQR]: 8.0, 32.0) compared to those with 'Medium' (8.0 cph; IQR: 4.0, 17.0), 'Low' (6.0 cph; IQR: 2.0, 20.0), 'Very low' (3.0 cph; IQR: 1.0, 11.5), and 'Xpert Trace/Neg. and 'Culture pos.' (2.0; IQR: 0.0, 4.5) results. This relationship between cough frequency and TB bacterial burden persisted throughout the initial 14 days of anti-TB treatment (**Figure 5.2**). A general decline in cough frequency was observed across all groups during this period and by day 14, all groups demonstrated a median cough frequency of fewer than 2 cph. Despite the overall reduction in cough frequency, significant heterogeneity in cough trajectories was observed among individuals (**Figure 5.2**).

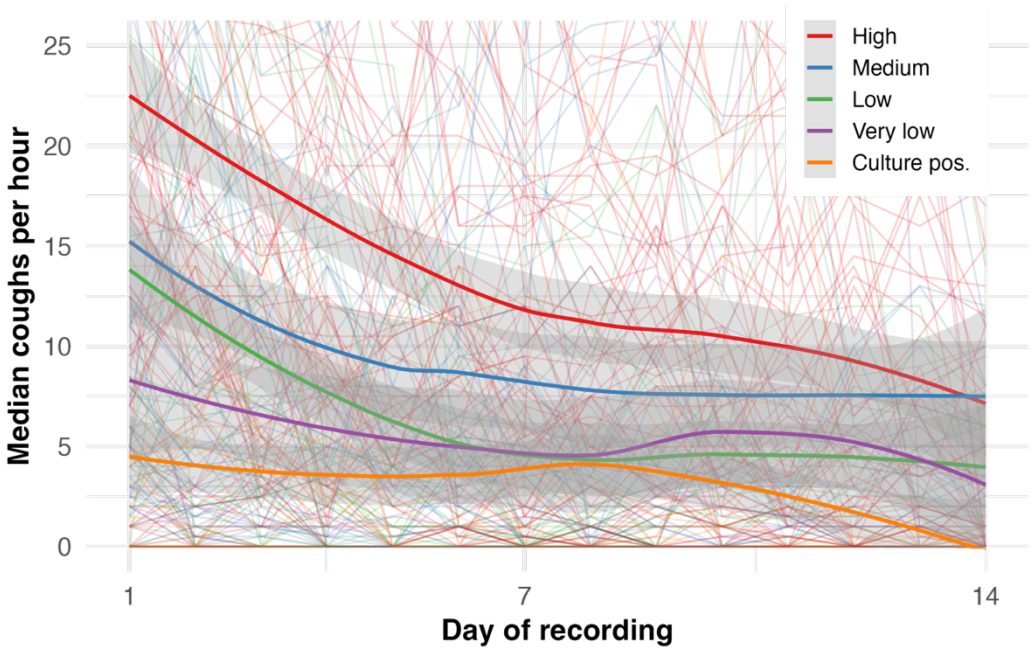


Figure 5.2. Trends in median coughs per hour during the first 14-days of anti-TB treatment among patients with microbiologically confirmed pulmonary TB. Participants were grouped according to their Xpert semi-quantitative results, which served as a marker of baseline TB bacterial burden. Individuals in the 'Culture pos.' group were culture positive but either Xpert negative or 'Trace'. Thick lines are locally estimated scatterplot smoothing (LOESS) curves for the median coughs per hour, with the shaded regions representing LOESS confidence intervals. The thin lines represent individual cough trajectories of different participants. The y-axis has been truncated at 25 coughs per hour for visualization.

When assessing the relationship between baseline CAD4TB score and cough frequency at each day of follow-up, a moderate positive correlation was observed that remained consistent throughout the 14-day period. Correlation coefficients (r) fluctuated between 0.33 (day 13) and 0.46 (day 6) (**Figure 5.3**). This finding suggests that despite the overall decrease in cough frequency (**Figure 5.2**), individuals with higher baseline CAD4TB scores maintained relatively higher cough rates throughout the study period.

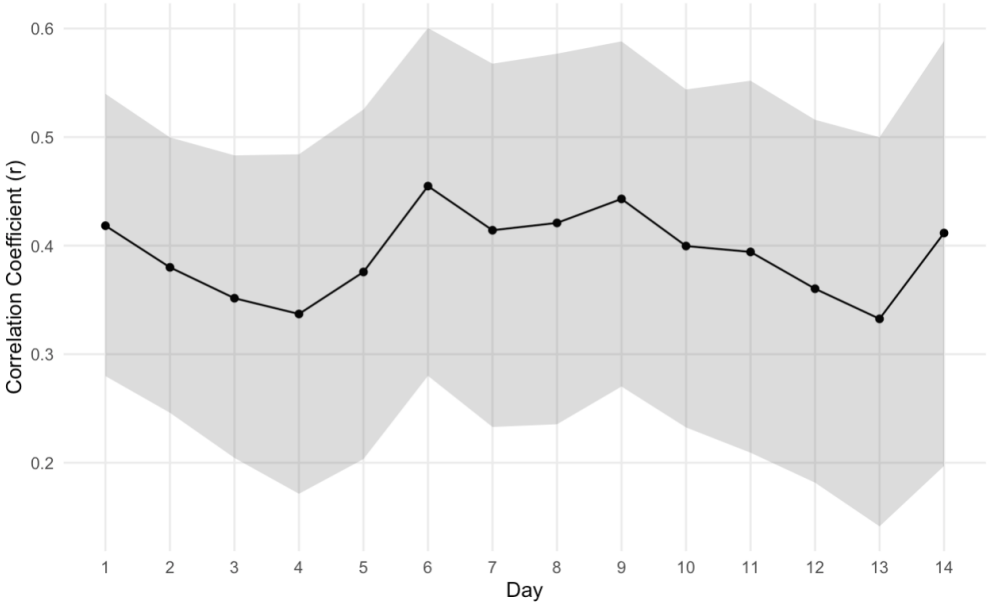


Figure 5.3. Correlation coefficient between median coughs per hour during the first 14-days of TB treatment and baseline CAD4TB scores. Shaded area represents the 95% confidence interval. The graph shows a consistent moderate correlation (around 0.4) between a patient's initial CAD4TB score and their coughing frequency throughout the 14-day period. This suggests that patients who started with higher CAD4TB scores generally maintained higher coughing rates throughout their first two weeks of treatment compared to patients who started with lower CAD4TB scores.

5.6.4. Daily cough counts – model results

The results of the regression analysis are presented in **Table 5.2**. Model 1 presents the RR for the multivariable Xpert semi-quantitative result analysis, showing that as the Xpert semi-quantitative result decreases, the cough rate also decreases. Model 2 presents the RRs for the CAD4TB analysis, illustrating that a decrease in log CAD4TB score results in a decrease in the cough rate (RR: 0.80; 95% CI: 0.67, 0.95). Overall, both models indicate that there is a positive association between TB bacterial burden (higher Xpert semi-quantitative result or higher CAD4TB score) and daily cough frequency.

The effect of time (day of follow-up) on cough frequency is illustrated in **Figure S5.8**. Since the coefficients for day and day^2 are the same for both Model 1 and Model 2 (**Table 5.2**), these results apply to both models. As participants progress on TB treatment, the cough rate decreases over the first 14 days of TB treatment. Results from the logit component of the zero-inflated model indicated that as time progressed, the odds of observing a zero cough count (i.e., no coughs recorded) increased. Specifically, for each unit increase in time, the odds of a zero count increased by 1.14 (95% CI: 1.05, 1.24) in both models.

The random effects in the models reveal individual heterogeneity in cough patterns (**Table 5.2**). In the negative binomial component, the standard deviation of the random slope for day (0.80 for Model 1 and 0.77 for Model 2) indicates variability in how cough frequency changes over time among participants. The logistic component of the model further illustrates this heterogeneity. The elevated standard deviation in random intercepts (7.40 for Model 1 and 7.80 for Model 2) suggests marked differences in individuals' propensity to record zero coughs. This large variability indicates that some participants are much more likely than others to have days with no recorded cough. Finally, the high AR1 correlation (0.94 for Model 1 and 0.95 for Model 2) in this component indicates strong day-to-day dependence in cough counts, underscoring the importance of accounting for temporal autocorrelation in the analysis.

Table 5.2. Results from multivariable zero-inflated negative binomial generalized linear mixed models evaluating the relationship between TB severity (Xpert semi-quantitative result and CAD4TB) and daily cough counts.

	Model 1: Xpert* RR (95% CI)	Model 2: CAD4TB† RR (95% CI)
	Negative binomial model	
Xpert semi-quantitative result		
High	Ref.	-
Medium	0.79 (0.59, 1.05)	
Low	0.64 (0.47 0.87)	
Very low	0.61 (0.41, 0.90)	
Xpert Trace/Neg. & Culture Pos.	0.43 (0.29, 0.62)	
CAD4TB score (log)	-	0.80 (0.67, 0.95)
Day	0.91 (0.88, 0.94)	0.91 (0.88, 0.94)
Day^2	1.00 (1.00, 1.01)	1.00 (1.00, 1.01)
Female	1.01 (0.81, 1.26)	1.05 (0.85, 1.33)
Age (years)	1.00 (0.97, 1.01)	1.00 (0.99, 1.01)
Country		
Madagascar	Ref.	Ref.
Philippines	0.50 (0.33, 0.75)	0.43 (0.28, 0.65)
Uganda	0.38 (0.28, 0.51)	0.39 (0.28, 0.53)
Vietnam	0.69 (0.47, 1.03)	0.56 (0.38, 0.82)
Smoked in past 7 days	1.10 (0.79, 1.54)	1.04 (0.74, 1.46)
HIV positive	1.06 (0.66, 1.70)	0.95 (0.59, 1.54)
COVID-19 positive	0.89 (0.59, 1.36)	0.80 (0.52, 1.22)
Prior TB	1.36 (0.96, 1.92)	1.32 (0.93, 1.87)
	Logit model	
Day	1.14 (1.05, 1.24)	1.14 (1.05, 1.24)

CAD, computer-aided detection; CI, confidence interval; EPTB, extrapulmonary tuberculosis; RR, rate ratio; TB, tuberculosis

* Negative binomial: standard deviation random in slope (day) = 0.76; AR1 correlation = 0.94. Logit: standard deviation in random intercept = 7.40.

† Negative binomial: standard deviation in random slope (day) = 0.77; AR1 correlation = 0.95. Logit: standard deviation in random intercept = 7.80.

5.6.7. Sensitivity analysis

Sensitivity analyses were conducted to assess the potential impact of cough count overestimation on the association between TB severity and cough frequency (**Table S5.2**; **Table S5.3**). These analyses revealed that censoring hourly cough counts at lower percentiles strengthened this association. For instance, when hourly cough counts were right censored at the 90th percentile, the

RR for ‘Medium’ semi-quantitative result versus ‘High’ semi-quantitative result decreased to 0.73 (95% CI: 0.57, 0.94) and to 0.79 (95% CI: 0.68, 0.92) in the CAD4TB model. In the extreme case where cough counts were right-censored at the 50th percentile, the association became even stronger with a RR of 0.50 (95% CI 0.44, 0.58) for Xpert ‘High’ versus ‘Medium’ and a RR of 0.72 (95% CI: 0.59, 0.88) for the CAD4TB model.

5.7. Discussion

The frequency of coughing in individuals with TB can provide valuable insights into the underlying disease state. Despite significant heterogeneity in longitudinal cough patterns among patients, distinct trends emerge over time. These trends correlate with TB bacterial burden at baseline, as indicated by markers such as Xpert semi-quantitative results and CAD4TB digital chest X-ray prediction scores. Patients with higher TB burden consistently demonstrate elevated cough frequencies at diagnosis, aligning with previous research findings.[15,23] Recent studies have emphasized the potential of monitoring changes in clinical symptoms as indicators of treatment response.[9,10] Understanding the variability in cough counts and the influence of baseline disease severity on cough count trajectories can therefore inform the development and evaluation of future cough monitoring tools for TB management.

Aside from TB bacterial burden, the setting where the participant was enrolled also influenced cough frequency. Participants in Madagascar exhibited particularly high rates of cough compared to those in other countries, even after controlling for other predictors of cough. The variation raises questions about the influence of context-specific factors on cough rates, including local epidemiology, sources of indoor air pollution, and environmental pollutants (e.g., particulate matter).[45] From a technical standpoint, there may also be cohort-specific behavioral differences in cough recording practice and adherence that may have contributed to observed differences. Further studies are needed to better characterize how all of these factors influence cough and modify the association with TB burden and eventually clinical response.

This study is the first to model the association between TB bacterial burden and cough counts recorded continuously over a sustained monitoring period. Prior studies exploring this association

used intermittently monitored 24h cough recordings, which do not capture the significant variability in cough over time.[35] Additionally, this study accounts for the underlying distribution of longitudinal cough counts, acknowledging mechanisms through which cough counts may be both underestimated (through the use of a zero-inflated distribution) and overestimated (through sensitivity analyses).

Cough counts are a zero-inflated marker, both at the daily and hourly level. The zero-inflation occurs through two mechanisms: 1) when the individual has truly not coughed in an hour period and 2) when coughs occurred but the monitor was either not recording or not in range. Prior analyses of cough data have shown that the expected frequency of 0 cph periods should be similar to the frequency 1 and 2 cph periods.[22] In this dataset, the proportion of 0 cph was significantly higher than the proportion of 1 and 2 cph (**Figure S5.4**). By extension, when hourly cough counts were summed at the daily level, the number of 0 coughs per day was high. While the exact contributions of the zero-inflation mechanisms cannot be ascertained without contextual data, feasibility challenges for continuous cough monitoring have previously outlined some of the factors that may contribute to excess zeros.[35,37] These can be related to the ergonomics of carrying a smartphone continuously around the neck as well as safety concerns of having a smartphone visible, putting individuals at risk of being robbed. Technological innovations (e.g., smartwatches, which are more ergonomic and subtle) may address some of these issues, and in turn, improve the quality of cough data.

Sensitivity analyses exploring the effect of cough overestimation, due to coughs being recorded from other individuals, found that overestimated cough counts may potentially attenuate the association between TB burden and cough frequency. However, this analysis was a crude overview of the potential impact of measurement error in cough detection. The sensitivity approach used, while addressing overestimation, may also inadvertently remove genuine high cough counts from severely ill patients, potentially introducing bias in the sensitivity analysis. In the long-term, advances in AI and monitoring devices may allow for speaker (cougher) identification algorithms that can be applied to cough audio sounds to help overcome this challenge.[46] Such technological improvements could significantly enhance the accuracy and reliability of cough monitoring in TB patients, potentially leading to more precise assessments of disease severity and treatment efficacy.

This study has several limitations. First, although it includes the largest sample of TB-positive individuals in a longitudinal cough study to date, the number of participants across different Xpert semi-quantitative groups remains relatively small. This is particularly evident in the lower semi-quantitative groups, which had fewer participants than the 'High' group. This imbalance may reflect selection bias, as recruitment from healthcare facilities likely favored individuals with more severe symptoms who were more likely to seek care and participate in the study. Second, there may be some unmeasured confounding between TB burden and cough frequency that was not accounted for, such as environmental exposures and cough-related co-morbidities. Third, contextual data on participant recording habits and environment was not available, preventing the concrete assessment of how under- and overestimation of cough influenced results. Accounting for zero-inflation in the model as well as performing sensitivity analyses attempted to control for this, however future studies may consider having patients self-report on their daily activities and using wearable devices which may increase the reliability and completeness of cough time series. Fourth, for ethical reasons, the study could not include a control group of untreated TB-positive individuals. This limitation makes it difficult to distinguish whether changes in cough frequency were due to treatment effects or natural disease progression. Finally, clinical and microbiological treatment outcomes were not available for these participants, so it was not possible to associate cough trends during the first two weeks of treatment with end of treatment outcomes. This study was not intended to determine whether cough could serve as a biomarker for evaluating treatment monitoring. Future studies should investigate this association to further assess the potential role of cough as a biomarker for treatment response, considering the findings presented here on baseline cough heterogeneity.

Overall, cough is an appealing biomarker for TB treatment response since it is low-cost, non-invasive, and capable of providing real-time data. This work provides insights into the variability of cough counts and their association with TB severity, aiding in the development of cough-based treatment response markers.

5.8. Acknowledgements

The datasets used for the analyses described were contributed by Dr. Adithya Cattamanchi at UCSF and Dr. Simon Grandjean Lapierre at University of Montreal and were generated in collaboration with researchers at Stellenbosch University (PI Grant Theron), Walimu (PIs William Worodria and Alfred Andama); De La Salle Medical and Health Sciences Institute (PI Charles Yu), Vietnam National Tuberculosis Program (PI Nguyen Viet Nhung), Christian Medical College (PI DJ Christopher), Centre Infectiologie Charles Mérieux Madagascar (PIs Mihaja Raberahona & Rivonirina Rakotoarivelo), and Ifakara Health Institute (PIs Issa Lyimo & Omar Lweno) with funding from the U.S. National Institutes of Health (U01 AI152087), The Patrick J. McGovern Foundation and Global Health Labs

5.9. Contributions

Conceptualization (A.J.Z., S.G.L.), R2D2 data collection and curation (D.J., D.J.C., N.V.N., G.T., W.W., C.Y., P.N., C.M.D., A.C., S.H.), Madagascar data collection and curation (M.R., R., R., S.G.L.), data preparation and cleaning (A.J.Z., S.H.), data visualization (A.J.Z.), data analysis (A.J.Z., A.B., S.H.), writing – original draft (A.J.Z.), writing – review and editing (M.R., D.J., R.R., D.J.C., N.V.N., G.T., W.W., C.Y., P.N., C.M.D., A.B., M.P., A.C., S.G.L., S.H.).

5.10. References

1. World Health Organization. Global Tuberculosis Report 2024.; 2024. Accessed November 18, 2024. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024
2. World Health Organization. WHO Operational Handbook on Tuberculosis - Module 2: Systematic Screening for Tuberculosis Disease. WHO; 2022. https://www.who.int/publications/i/item/9789240022614
3. Turner RD, Bothamley GH. Cough and the Transmission of Tuberculosis. The Journal of Infectious Diseases. 2015;211(9):1367-1372. doi:10.1093/infdis/jiu625
4. Suzuki T, Shirai T, Hirai K, et al. Improved cough- and sputum-related quality of life after initiation of treatment in pulmonary tuberculosis. Respiratory Investigation. 2019;57(3):252-259. doi:10.1016/j.resinv.2019.01.002

5. Stadler J a. M. Updated WHO definitions for tuberculosis outcomes: Simplified, unified and future-proofed. Afr J Thorac Crit Care Med. 2022;28(2). doi:10.7196/AJTCCM.2022.v28i2.224
6. Horne DJ, Royce SE, Gooze L, et al. Sputum monitoring during tuberculosis treatment for predicting outcome: systematic review and meta-analysis. The Lancet Infectious Diseases. 2010;10(6):387-394. doi:10.1016/S1473-3099(10)70071-2
7. Gupta-Wright A, Den Boon S, MacLean E, et al. Target product profiles: tests for tuberculosis treatment monitoring and optimization. Bull World Health Org. 2023;101(11):730-737. doi:10.2471/BLT.23.290901
8. Zimmer AJ, Lainati F, Aguilera Vasquez N, et al. Biomarkers That Correlate with Active Pulmonary Tuberculosis Treatment Response: a Systematic Review and Meta-analysis. McAdam AJ, ed. J Clin Microbiol. 2022;60(2):e01859-21. doi:10.1128/jcm.01859-21
9. Heyckendorf J, Georghiou SB, Frahm N, et al. Tuberculosis Treatment Monitoring and Outcome Measures: New Interest and New Strategies. Clin Microbiol Rev. 2022;35(3):e00227-21. doi:10.1128/cmr.00227-21
10. Wejse C, Gustafson P, Nielsen J, et al. TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course. Scandinavian Journal of Infectious Diseases. 2008;40(2):111-120. doi:10.1080/00365540701558698
11. Turner R, Bothamley G, Birring S. P240 Validation of the Leicester Cough Questionnaire in pulmonary tuberculosis. Thorax. 2015;70(Suppl 3):A197.2-A198. doi:10.1136/thoraxjnl-2015-207770.376
12. Turner RD, Birring SS, Bothamley GH. Measuring cough-related quality of life and cough frequency in pulmonary TB. int j tuberc lung dis. 2024;28(2):93-98. doi:10.5588/ijtld.23.0312
13. Turner R, Repossi A, Matos S, Birring S, Bothamley G. S79 Cough Prevalence And Frequency In Pulmonary Tuberculosis. Thorax. 2014;69(Suppl 2):A43-A44. doi:10.1136/thoraxjnl-2014-206260.85
14. Williams CM, Abdulwhhab M, Birring SS, et al. Exhaled Mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: prospective observational studies. The Lancet Infectious Diseases. 2020;20(5):607-617. doi:10.1016/S1473-3099(19)30707-8
15. Proaño A, Bravard MA, López JW, et al. Dynamics of Cough Frequency in Adults Undergoing Treatment for Pulmonary Tuberculosis. Clinical Infectious Diseases. 2017;64(9):1174-1181. doi:10.1093/cid/cix039
16. Lee GO, Comina G, Hernandez-Cordova G, et al. Cough dynamics in adults receiving tuberculosis treatment. Ehtesham HS, ed. PLoS ONE. 2020;15(6):e0231167. doi:10.1371/journal.pone.0231167
17. Huddart S, Asege L, Jaganath D, et al. Continuous cough monitoring: a novel digital biomarker for TB diagnosis and treatment response monitoring. int j tuberc lung dis. 2023;27(3):221-222. doi:10.5588/ijtld.22.0511
18. Sharma M, Nduba V, Njagi LN, et al. TBscreen: A passive cough classifier for tuberculosis screening with a controlled dataset. Sci Adv. 2024;10(1):eadi0282. doi:10.1126/sciadv.adi0282

19. Huddart S, Yadav V, Sieberts SK, et al. Solicited Cough Sound Analysis for Tuberculosis Triage Testing: The CODA TB DREAM Challenge Dataset. Published online March 28, 2024. doi:10.1101/2024.03.27.24304980
20. Botha GHR, Theron G, Warren RM, et al. Detection of tuberculosis by automatic cough sound analysis. Physiol Meas. 2018;39(4):045005. doi:10.1088/1361-6579/aab6d0
21. Zimmer AJ, Ugarte-Gil C, Pathri R, et al. Making cough count in tuberculosis care. Commun Med. 2022;2(1):83. doi:10.1038/s43856-022-00149-w
22. Rudd M, Song WJ, Small PM. The Statistics of Counting Coughs: Easy as 1, 2, 3? Lung. 2022;200(5):531-537. doi:10.1007/s00408-022-00559-6
23. Proaño A, Bui DP, López JW, et al. Cough Frequency During Treatment Associated With Baseline Cavitary Volume and Proximity to the Airway in Pulmonary TB. Chest. 2018;153(6):1358-1367. doi:10.1016/j.chest.2018.03.006
24. Magombedze G, Pasipanodya JG, Gumbo T. Bacterial load slopes represent biomarkers of tuberculosis therapy success, failure, and relapse. Commun Biol. 2021;4(1):664. doi:10.1038/s42003-021-02184-0
25. R2D2 TB Network. R2D2 TB Network. Accessed October 24, 2024. https://www.r2d2tbnetwork.org
26. Martin-Higuera MC, Rivas G, Rolo M, Muñoz-Gallego I, Lopez-Roa P. Xpert MTB/RIF Ultra CT value provides a rapid measure of sputum bacillary burden and predicts smear status in patients with pulmonary tuberculosis. Sci Rep. 2023;13(1):1591. doi:10.1038/s41598-023-28869-6
27. Opota O, Mazza-Stalder J, Greub G, Jaton K. The rapid molecular test Xpert MTB/RIF ultra: towards improved tuberculosis diagnosis and rifampicin resistance detection. Clinical Microbiology and Infection. 2019;25(11):1370-1376. doi:10.1016/j.cmi.2019.03.021
28. World Health Organization. WHO Meeting Report of a Technical Expert Consultation: Non-Inferiority Analysis of Xpert MTB/RIF Ultra Compared to Xpert MTB/RIF.; 2017.
29. Nielsen MC, Clarner P, Paroha R, Lee S, Thwe PM, Ren P. Comparison of Analytical Sensitivity (Limit of Detection) of Xpert MTB/RIF and Xpert MTB/RIF Ultra for Non-Sputum Specimens. Pathogens. 2023;12(2):157. doi:10.3390/pathogens12020157
30. Chilukutu L, Mwanza W, Kerkhoff AD, Somwe P, Kagujje M, Muyoyeta M. Prevalence and interpretation of Xpert® Ultra trace results among presumptive TB patients. public health action. 2022;12(1):28-33. doi:10.5588/pha.21.0066
31. Murphy K, Habib SS, Zaidi SMA, et al. Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. Sci Rep. 2020;10(1):5492. doi:10.1038/s41598-020-62148-y
32. Worodria W, Castro R, Kik SV, et al. An independent, multi-country head-to-head accuracy comparison of automated chest x-ray algorithms for the triage of pulmonary tuberculosis. Published online June 19, 2024. doi:10.1101/2024.06.19.24309061
33. Soares TR, Oliveira RD de, Liu YE, et al. Evaluation of chest X-ray with automated interpretation algorithms for mass tuberculosis screening in prisons: a cross-sectional study. Lancet Reg Health Am. 2023;17:100388. doi:10.1016/j.lana.2022.100388
34. Khan FA, Majidulla A, Tavaziva G, et al. Chest x-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: a prospective study of diagnostic

accuracy for culture-confirmed disease. The Lancet Digital Health. 2020;2(11):e573-e581. doi:10.1016/S2589-7500(20)30221-1

35. Lee SE, Rudd M, Kim TH, et al. Feasibility and Utility of a Smartphone Application-Based Longitudinal Cough Monitoring in Chronic Cough Patients in a Real-World Setting. Lung. 2023;201(6):555-564. doi:10.1007/s00408-023-00647-1
36. Galvosas M, Gabaldón-Figueira JC, Keen EM, et al. Performance evaluation of the smartphone-based AI cough monitoring app - Hyfe Cough Tracker against solicited respiratory sounds. F1000Res. 2023;11:730. doi:10.12688/f1000research.122597.2
37. Huddart S, Asege L, Dang H, et al. Feasibility and Analysis Considerations of Continuous Community-Based Cough Monitoring in Low- and Middle-Income Settings. Public and Global Health; 2022. doi:10.1101/2022.07.06.22277224
38. Buuren SV, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. J Stat Soft. 2011;45(3). doi:10.18637/jss.v045.i03
39. Robitzsch A, Grund S, Thorsten H. Package 'miceadds.' Published online January 9, 2024. https://cran.r-project.org/web/packages/miceadds/miceadds.pdf
40. Raberahona M, Zimmer A, Rakotoarivelo RA, et al. Continuous digital cough monitoring during 6-month pulmonary tuberculosis treatment. ERJ Open Research. Published online January 1, 2024. doi:10.1183/23120541.00655-2024
41. Gabaldón-Figueira JC, Keen E, Rudd M, et al. Longitudinal passive cough monitoring and its implications for detecting changes in clinical status. ERJ Open Res. 2022;8(2):00001-02022. doi:10.1183/23120541.00001-2022
42. Hedeker D, Gibbons RD. Longitudinal Data Analysis. John Wiley & Sons; 2006.
43. Hartig F, Lohse L. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. Published online September 8, 2022. Accessed July 4, 2024. https://CRAN.R-project.org/package=DHARMa
44. Brooks M, Bolker B, Kristensen K, Maechler M, Magnusson A, McGillycuddy M. glmmTMB: Generalized Linear Mixed Models using Template Model Builder. Published online March 20, 2024. https://cran.uni-muenster.de/web/packages/glmmTMB/glmmTMB.pdf
45. Zimmer AJ, Tsang LY, Jolicoeur G, et al. Incidence of cough from acute exposure to fine particulate matter (PM2.5) in Madagascar: A pilot study. Puttaswamy N, ed. PLOS Glob Public Health. 2024;4(7):e0003530. doi:10.1371/journal.pgph.0003530
46. Zhang M, Chen Y, Li L, Wang D. Speaker Recognition with Cough, Laugh and "Wei." Published online 2017. doi:10.48550/ARXIV.1706.07860

5.11. Supplementary information

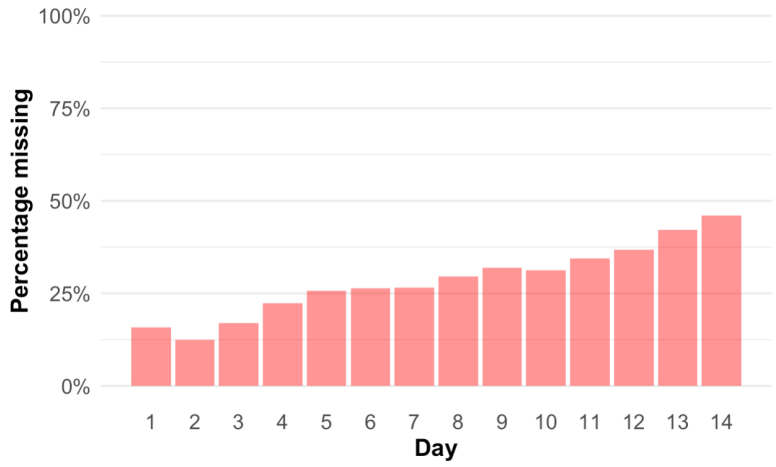


Figure S5.1. Percentage of hourly cough data that are missing over follow-up day.

The amount of missingness increased as follow-up continued. Overall, 28.4% of hourly cough data was imputed.

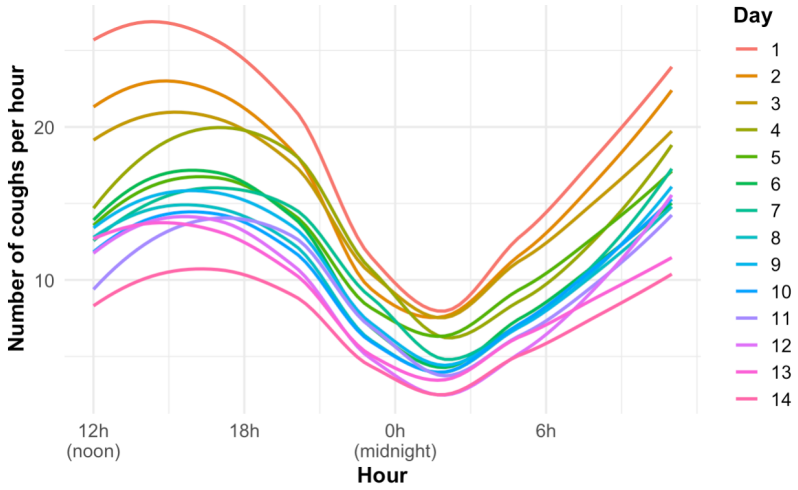


Figure S5.2. Circadian trend in cough over each day of follow-up.

Lines are locally estimated scatterplot smoothing (LOESS) curves for number of coughs per hour. Each line colour represents a different day of follow-up. Cough frequency appears to peak in the afternoon (14h-15h) and is lowest at night (1h-2h). This trend persists throughout the 14 days of follow-up, however the magnitude of coughing decreases with time of follow-up.

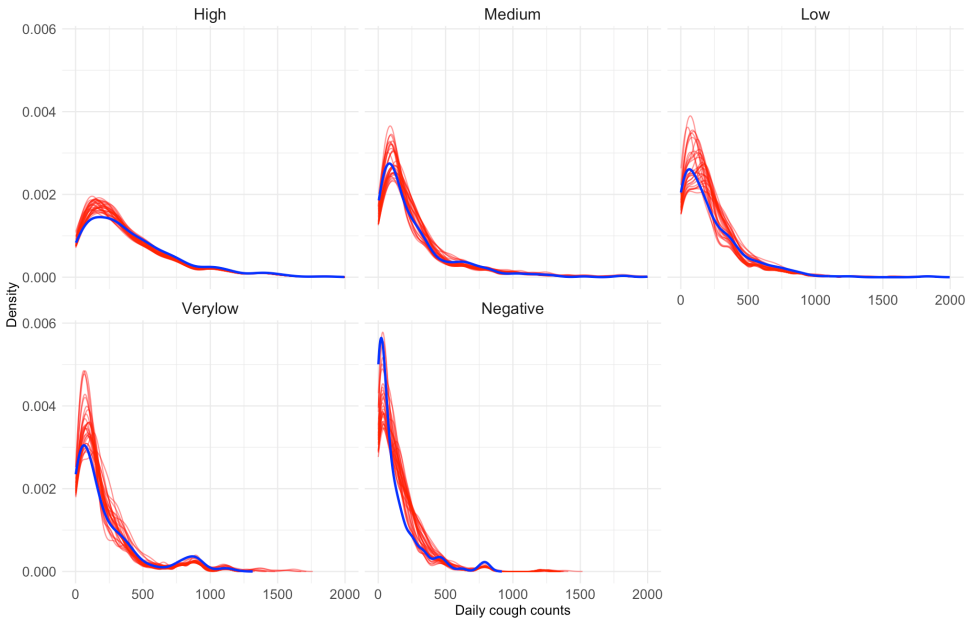


Figure S5.3. Density plot of observed and imputed daily cough count data by GeneXpert semi-quantitative result. Blue lines represent the distribution of the original, unimputed dataset. Red lines show the 30 multiple imputations.

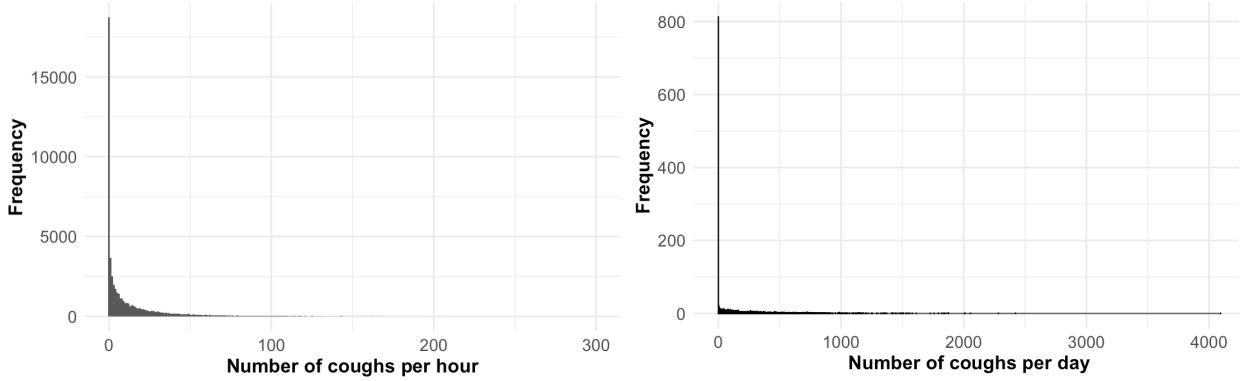


Figure S5.4. Distribution of cough data among all participants and all follow-up hours. The width of each band represents a single increment in cough frequency on the x-axis. Left: Zero-inflated distribution of hourly cough counts. Right: Zero-inflated distribution of daily cough counts. The expected proportion of 0 coughs per hour (cph) should be similar to the number of 1 cph and 2 cph. In this dataset, 26% of hourly cough counts were 0 cough (18,724/72,224 hours), which is greater than the 5% (3,670/72,224 hours) of 1 cph and the 3% of 2 cph (2,493/72,224 hours). This zero-inflation was also observed when summing hourly cough counts at the daily level, with 27% (813/3,010 days) having 0 coughs throughout the day.

Table S5.1. Comparative analysis of model structures for daily cough count data for the Xpert and CAD4TB multivariable models using the Akaike Information Criterion (AIC).

#	Negative binomial count model*	Zero-inflation model†	Xpert Mean AIC	CAD4TB Mean AIC
Negative binomial				
1	Random intercept	N/A	36,162	36,171
2	Random intercept Random slope (day)	N/A	35,770	35,777
3	AR1 covariance structure	N/A	35,739	35,754
Zero-inflated negative binomial				
4	Random intercept	Random intercept	35,265	35,272
5	Random intercept Random slope (day)	Random intercept	34,975	34,983
6	Random intercept Random slope (day)	Random intercept Random slope (day)	Failed§	34,923
7	**AR1 covariance structure**	**Random intercept**	**34,932**	**34,940**
8	Exchangeable covariance structure	Random intercept	Failed§	Failed§
9	Toeplitz covariance structure	Random intercept	Failed§	Failed§
10	AR1 covariance structure	AR1 covariance structure	Failed§	Failed§

Maximum Likelihood estimation was employed for all model comparisons.
Model in bold was used in the final analysis.
*Fixed effects: Xpert/log(CAD4TB), Day, Day^2, Sex, Age, Country, Smoking, HIV status, COVID, Prior TB.
†Fixed effects: Day
§Model did not converge.
AR1, autoregressive 1; CAD, computer-aided detection; Xpert, GeneXpert

Equation S5.1. Equation for final selected model.

Count model (negative binomial):

$$\log(\mu_{it}) = \beta_0 + \beta_1 X_{xpert/cad} + \beta_2 X_{day} + \beta_3 X_{day^2} + \beta_4 X_{sex} + \beta_5 X_{age} + \beta_6 X_{country} + \beta_7 X_{smoking} + \beta_8 X_{hiv} + \beta_9 X_{covid} + \beta_{10} X_{priorTB} + u_{it}$$

Where:

μ_{it} = the expected cough count for person *i* at time (day) *t*

u_{it} = the random effect for subject *i* at time *t* with an AR(1) correlation structure

Zero-inflation (logistic regression):

$$\text{logit}(\pi_{it}) = \gamma_0 + \gamma_1 X_{day} + v_i$$

Where:

π_{it} = the probability of excess zero for persion i at time t

v_i = the random intercept for person i

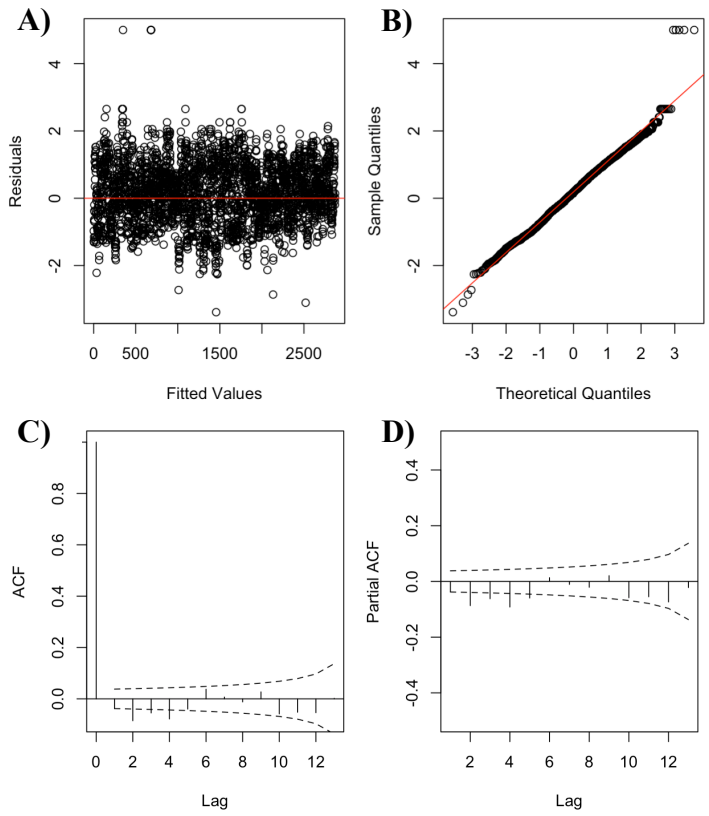


Figure S5.5. Residual and autocorrelation analysis for final zero-inflated negative binomial model using CAD4TB as a predictor. Using the final model (#7 in Table S5.1). A) Residuals vs. Fitted Values, B) Q-Q plot of residuals, C) Partial Autocorrelation Function (Partial ACF), and D) Autocorrelation Function (ACF).

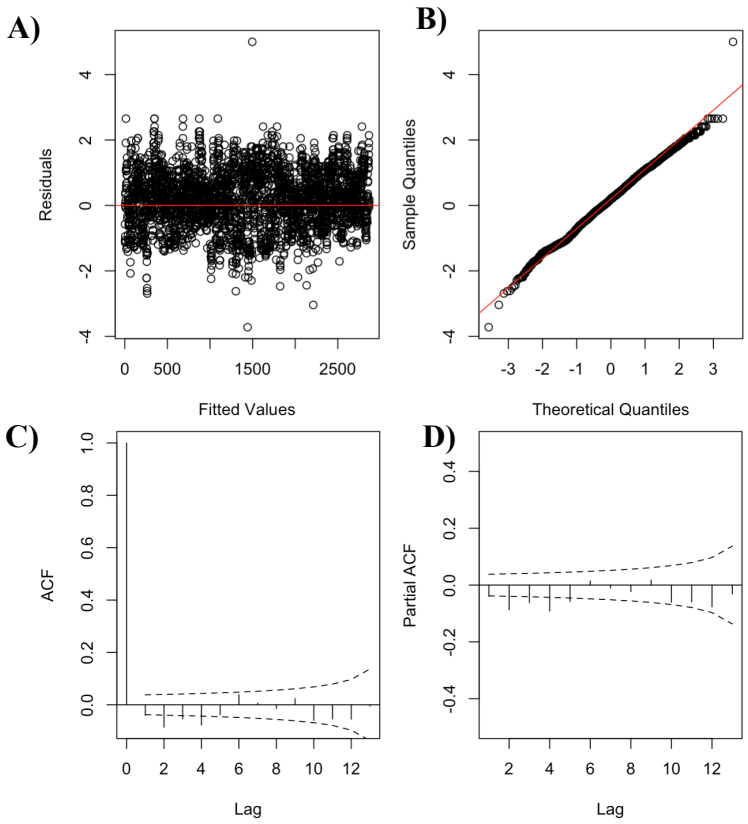


Figure S5.6. Residual and autocorrelation analysis for final zero-inflated negative binomial model using Xpert semi-quantitative score as a predictor.

Using the final model (#7 in Table S5.1). A) Residuals vs. Fitted Values, B) Q-Q plot of residuals, C) Partial Autocorrelation Function (Partial ACF), and D) Autocorrelation Function (ACF).

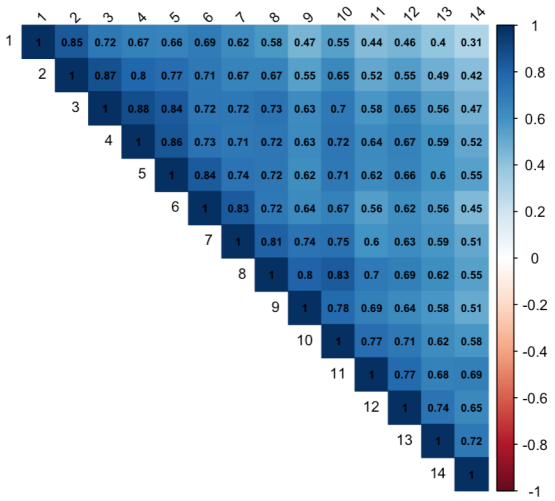


Figure S5.7. Correlation matrix of cough frequency between days of follow-up. Evidence of an auto-regressive trend can be observed, with decreasing correlation with increasing time between follow-up days.

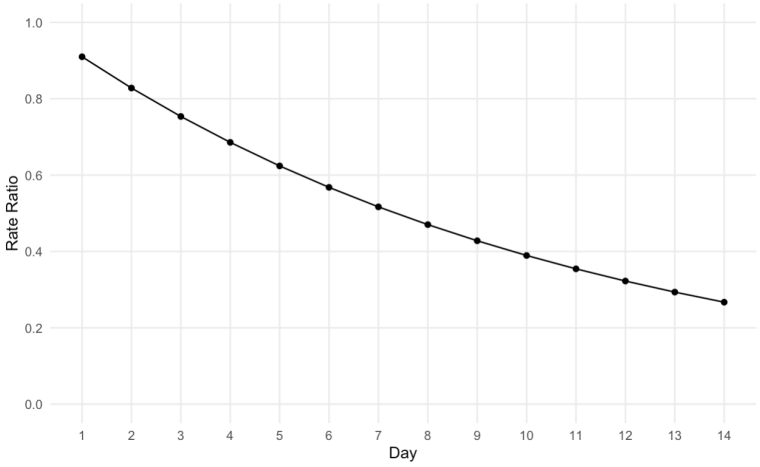


Figure S5.8. Change in rate ratio (RR) of cough counts over the first 14 days of TB treatment. The RR for each day of follow-up is displayed, controlling for additional covariates. The smooth, downward-sloping curve demonstrates a steady reduction in the rate of cough counts over time. Starting at approximately 0.9 on Day 1, the RR decreases consistently each day, reaching about 0.23 by Day 14.

Table S5.2. Rate ratios of cough after censoring hourly cough counts at various percentiles for Xpert semi-quantitative results.

	50th percentile RR (95% CI)	**75th percentile RR (95% CI)**	**90th percentile RR (95% CI)**
		Negative binomial model	
Xpert semi-quant			
High	Ref.	Ref.	Ref.
Medium	0.50 (0.44, 0.58)	0.63 (0.52, 0.77)	0.73 (0.57, 0.94)
Low	0.37 (0.32, 0.43)	0.52 (0.42, 0.65)	0.60 (0.45, 0.77)
Very low	0.28 (0.23, 0.33)	0.43 (0.33, 0.57)	0.54 (0.39, 0.76)
Xpert Trace/Neg. & Culture Pos.	0.14 (0.12, 0.17)	0.27 (0.21, 0.35)	0.36 (0.26, 0.50)
Day	0.85 (0.83, 0.87)	0.89 (0.87, 0.91)	0.90 (0.87, 0.92)
Day^2	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.01)
Female	0.99 (0.90, 1.10)	0.99 (0.85, 1.15)	1.00 (0.82, 1.20)
Age (years)	1.00 (1.00, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.01)
Country			
Madagascar	Ref.	Ref.	Ref.
Philippines	0.71 (0.59, 0.86)	0.60 (0.46, 0.79)	0.53 (0.38, 0.76)
Uganda	0.67 (0.59, 0.78)	0.53 (0.43, 0.65)	0.44 (0.34, 0.57)
Vietnam	0.89 (0.75, 1.06)	0.80 (0.62, 1.03)	0.74 (0.53, 1.03)
Smoked in past 7 days	1.04 (0.89, 1.21)	1.07 (0.85, 1.34)	1.09 (0.81, 1.45)
HIV positive	0.89 (0.72, 1.11)	0.95 (0.69, 1.30)	1.00 (0.67, 1.50)
COVID-19 positive	0.97 (0.80, 1.17)	0.95 (0.71, 1.26)	0.93 (0.65, 1.34)
Prior TB	1.13 (0.96, 1.33)	1.21 (0.95, 1.53)	1.29 (0.96, 1.74)
		Logit model	
Day	1.46 (1.36, 1.57)	1.14 (1.05, 1.23)	1.14 (1.05, 1.23)

CI, confidence interval; cph, coughs per hour; RR, rate ratio; TB, tuberculosis

Table S5.3. Rate ratios of cough after censoring hourly cough counts at various percentiles for CAD4TB scores.

	50th percentile RR (95% CI)	75th percentile RR (95% CI)	90th percentile RR (95% CI)
		Negative binomial model	
CAD4TB score (log)	0.72 (0.59, 0.88)	0.77 (0.68, 0.89)	0.79 (0.68, 0.92)
Day	0.85 (0.81, 0.89)	0.89 (0.87, 0.91)	0.90 (0.87, 0.92)
Day^2	1.00 (1.00, 1.00)	1.00 (1.00, 1.00)	1.00 (1.00, 1.01)
Female	1.03 (0.76, 1.39)	1.02 (0.86, 1.24)	1.04 (0.84, 1.28)
Age (years)	0.99 (0.99, 1.00)	1.00 (0.99, 1.00)	1.00 (0.99, 1.00)
Country			
Madagascar	Ref.	Ref.	Ref.
Philippines	0.44 (0.31, 0.62)	0.45 (0.33, 0.61)	0.44 (0.30, 0.63)
Uganda	0.62 (0.39, 0.97)	0.51 (0.40, 0.66)	0.44 (0.33, 0.58)
Vietnam	0.57 (0.43, 0.77)	0.58 (0.44, 0.77)	0.57 (0.41, 0.79)
Smoked in past 7 days	1.00 (0.76, 1.32)	1.04 (0.78 1.33)	1.03 (0.76, 1.39)
HIV positive	0.69 (0.44, 1.07)	0.80 (0.55, 1.16)	0.87 (0.57, 1.33)
COVID-19 positive	0.75 (0.53, 1.06)	0.80 (0.57, 1.12)	0.81 (0.56, 1.19)
Prior TB	1.08 (0.80, 1.45)	1.17 (0.88, 1.54)	1.25 (0.91, 1.71)
		Logit model	
Day	1.45 (1.30, 1.63)	1.14 (1.05, 1.23)	1.14 (1.05, 1.23)

CI, confidence interval; cph, coughs per hour; RR, rate ratio; TB, tuberculosis

Chapter 6. Manuscript IV: Cough acoustics for COVID-19 detection: a comparative study of patient cohorts from Lima, Peru and Montreal, Canada

6.1. Preface

While the previous chapters explored cough counts as a biomarker, cough can also be examined through lens of its acoustic properties. Since the COVID-19 pandemic, there has been substantial interest in developing ML and AI tools for cough sound analysis, particularly for COVID-19 screening. While early algorithms reported promising results, these were largely based on crowdsourced datasets with self-reported diagnoses, raising concerns about their real-world applicability. Moreover, these algorithms were typically validated using internal datasets only, without assessment of their performance in distinct populations.

In this study, we investigated the challenges of developing AI-based cough screening tools using prospectively collected data from two distinct populations in Montreal, Canada and Lima, Peru. Unlike previous studies, all participants underwent rigorous reference standard diagnostic testing to confirm their COVID-19 status and screen for other respiratory pathogens. By comparing acoustic features and model performance across these populations, we examined whether cough-based screening algorithms developed in one setting could be successfully transferred to another. This work addresses critical gaps in our understanding of population-specific differences in cough acoustics and their implications for developing globally applicable screening tools.

This project was funded by a CIHR grant for which I was one of the primary grant writers.

6.2. Title page

Cough acoustics for COVID-19 detection: a comparative study of patient cohorts from Lima, Peru and Montreal, Canada

Alexandra J. Zimmer[1,2], Vijay Ravi[3], Patricia Espinoza Lopez[4,5], George P. Kafentzis [6], Mirco Ravanelli[7,8], Samira Abbasgholizadeh Rahimi[7,9,10], Madhukar Pai[1,2], César Ugarte-Gil[4,5,11], Simon Grandjean Lapierre[2,12,13]

[1] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

[2] McGill International TB Centre, Montreal, Canada

[3] Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, USA

[4] School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

[5] Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, Peru

[6] Hyfe Inc., Wilmington, DE, USA

[7] Mila – Quebec AI Institute, Montreal, Quebec, Canada

[8] Department of Computer Science and Operations Research, Université de Montréal, Montreal, Quebec, Canada

[9] Family Medicine, Faculty of Medicine and Health Sciences and Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Quebec, Canada

[10] Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

[11] Department of Epidemiology, School of Public and Population Health, University of Texas Medical Branch, Galveston, USA

[12] Immunopathology Axis, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, 900 Rue Saint-Denis, Montréal, Canada

[13] Department of Microbiology, Infectious Diseases and Immunology, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Canada

6.3. Abstract

Introduction

While digital cough screening shows promise for COVID-19 detection, understanding how population differences affect both cough acoustics and screening accuracy is crucial. This study investigates cough characteristics and COVID-19 screening performance across two populations in Lima, Peru and Montreal, Canada.

Methods

Cough recordings and clinical data were collected prospectively from 605 adults with any cough across both sites. COVID-19 status was confirmed by nucleic acid amplification testing (NAAT). Additional NAAT was conducted to detect other respiratory pathogens common to each site. Acoustic features (including spectral and temporal characteristics) were extracted from cough recordings and compared between cohorts. COVID-19 classification was performed using eXtreme Gradient Boosting (XGBoost) and the Emphasized Channel Attention, Propagation and Aggregation Time Delay Neural Network (ECAPA-TDNN), a neural network architecture optimized for audio signal processing. Model performances were assessed through within-dataset validations (internal validity) and external validations (transferability) both for audio-only, clinical-only, and combined audio and clinical models. The areas under the curve (AUCs) were reported as averages with standard errors (SE) across 5 independent training iterations. A sub-analysis investigated differences in XGBoost COVID-19 prediction scores for the within-dataset models according to underlying disease status (COVID-19 positive, other disease positive, or negative on available tests performed).

Results

Descriptive analyses highlighted significant heterogeneity in cough acoustic features between Lima and Montreal cohorts. Audio-based models trained and tested in Lima demonstrated superior performance (AUC: 0.71; SE ± 0.08) compared to audio models trained and tested within Montreal

(AUC: 0.53; SE ± 0.04). Both models showed poor performance during external validation, with Lima-trained audio models dropping to an AUC of 0.51 (SE ± 0.01) when tested on Montreal data, and Montreal-trained audio models remaining at an AUC of 0.53 (SE ± 0.03) when tested on Lima data. ECAPA-TDNN models demonstrated similar trends with superior within-dataset performance in Lima and poor external validity across both datasets. In the sub-analysis, individuals positive for other respiratory diseases (e.g., influenza, tuberculosis) had higher COVID-19 prediction scores in Montreal compared to Peru, suggesting potential variations in model performance across different epidemiological contexts.

Conclusion

The findings demonstrate that cough acoustics are population-specific, with distinct cough feature distributions between sites. The utility of cough-based classification algorithms may differ depending on the setting and epidemiological profile of the population. There was limited transferability of COVID-19 cough screening models between different geographical, demographic, and epidemiological contexts. This study highlights the challenges in developing globally applicable cough-based COVID-19 screening tools when the training data is not representative of the target population

6.4. Introduction

The COVID-19 pandemic catalyzed the development of innovative screening and diagnostic solutions,[1] including algorithms for cough sound analysis. These efforts have resulted in the creation of various algorithms, employing both advanced artificial intelligence (AI) neural network approaches,[2–5] and more traditional machine learning (ML) methods.[6–10] While neural networks and deep learning show promise for improving model performance and accuracy with large datasets, ML methods provide greater transparency in their decision-making process,[11] allowing researchers to more easily interpret and explain which acoustic features are most influential in the model's predictions.

Early COVID-19 classification algorithms, both AI- and ML-based, reported impressive performance metrics, with many achieving an area under the curve (AUCs) or accuracy that exceeds 80%.[2–10] These algorithms were proposed as rapid, non-invasive screening methods to identify individuals who may need further diagnostic testing. However, they used crowdsourced cough datasets, such as Coswara,[12] Virufy,[13] and the Cambridge COVID-19 Sounds.[14] While these datasets were valuable resources for initial research, they are subject to several biases that may impact the generalizability and real-world applicability of the resulting models.[15] For instance, they rely on self-reported COVID-19 diagnoses rather than gold-standard laboratory confirmation, potentially introducing misclassification bias. Second, the variable recording quality inherent in crowdsourced data, where participants use their own devices and recording environments, may introduce noise and inconsistencies in the recordings. Third, selection biases may be present, as participants in these datasets may not represent the broader population of COVID-19 patients who would seek care for their cough. Finally, the limited availability and quality of clinical and diagnostic labels for other diseases that produce cough in these datasets restrict the ability to assess and control for potential confounding factors. Moreover, these datasets were largely generated during the early stages of the COVID-19 pandemic when other respiratory viruses were less prevalent due to public health measures, and when clinical attention was primarily focused on COVID-19, resulting in datasets enriched with or biased toward a single diagnosis.

Another limitation of the cough classification literature is its narrow focus on disease state differentiation. Studies have focused on distinguishing between various health conditions (e.g., healthy versus COVID-19, healthy versus tuberculosis [TB], etc.), without addressing the underlying heterogeneity of acoustic features among people with the same disease profile but who are from different patient populations. In other words, there is a lack of evidence regarding whether populations have different acoustic feature distributions, and how these differences may impact the transferability of models developed in one population to others. Additionally, the epidemiological profiles of respiratory diseases can vary significantly across regions, affecting disease prevalence patterns and, consequently, the population-level distribution of acoustic signatures that diagnostic models must distinguish. This variability in disease landscapes across settings further complicates the development of universally applicable cough classification algorithms. This research gap raises important questions about the feasibility of developing "global" cough classification tools that can be generalized to broader populations versus the need for localized algorithms that can identify population-specific acoustic trends.

The objective of this study is twofold: 1) characterize population-level differences in cough acoustic features between distinct geographical cohorts and 2) assess the external validity of AI and ML COVID-19 cough classification algorithms between those cohorts. The study uses prospectively collected cough sounds from coughing patients in Lima, Peru and Montreal, Canada with gold-standard nuclear acid amplification tests (NAATs) as a reference. By comparing these distinct populations from the Global North and the Global South, this study aims to shed light on potential variations in cough acoustics and their impact on the performance of classification algorithms across different settings.

6.5. Methods

6.5.1. Settings

This study used prospective data collected from Lima, Peru and Montreal, Canada. Participants in Lima were recruited from two healthcare settings: Hospital de Huaycán, a public secondary referral hospital in the Ate-Vitarte district, and a network of 33 primary health centers in the San

Juan de Lurigancho distrtict. Both districts are characterized by high population density, with San Juan de Lurigancho being Lima's most populous district. These districts also bear a significant TB burden, with San Juan de Lurigancho reporting the highest TB incidence rate in Lima.[16,17] Both health facilities have clinics dedicated to diagnosing and treating TB patients. During the study period, COVID-19 prevalence among patients attending these facilities was notably high at 35%.[16] Recruitment in Lima occurred in two phases. Phase 1 ran from March 2022 to January 2023, where participants were recruited from the cohort of an ongoing parent study investigating the diagnostic accuracy of integrated TB and COVID-19 testing.[16] Phase 2 took place from July 2023 to March 2024.

In Montreal, recruitment took place at the Centre hospitalier de l'Université de Montréal COVID-19 screening center between April 2022 and November 2023. During this period, Quebec experienced high COVID-19 positivity rates,[18] while Canada saw increased circulation of other respiratory viruses, particularly during the severe 2022-2023 influenza season when test positivity peaked at 24% in November 2022.[19] Unlike Lima, Montreal and Canada maintain a low TB burden, with a national TB incidence of 5.1 per 100,000 population in 2022.[20]

6.5.2. Participants

In Lima, consecutive adults presenting with a cough of any duration were enrolled. During Phase 1, the exclusion criteria of the parent study applied, excluding adults with confirmed COVID-19 in the last three months or who had taken anti-TB medication in the past six months.[16] For both phases, participants were excluded from the sub-study if they were currently taking cough-suppressive medication.

In Montreal, consecutive adults presenting with a cough of any duration were enrolled. Participants were excluded if they were taking cough-suppressive medication at the time of enrollment.

6.5.3. COVID-19 and other respiratory disease diagnosis

In Lima, nasopharyngeal samples were collected, stored at 2-8°C, and transported to the Humberto Guerra Alisson laboratory, a reference laboratory at the Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia. During Phase 1, samples were tested for COVID-19 using Xpert Xpress SARS-CoV-2 cartridges on the GeneXpert automated molecular platform (Cepheid, Sunnyvale, CA, USA). During Phase 2, samples were tested using the Xpert Xpress SARS-CoV-2/Flu/RSV cartridges (Cepheid, Sunnyvale, CA, USA). Both cartridges received emergency-use authorization from the U.S. Food and Drug Administration and have demonstrated high diagnostic accuracy for COVID-19 diagnosis.[21–24] They are NAATs that run on the GeneXpert platform. If the initial NAAT result produced an error or was invalid, the test was repeated using the same sample.

In both phases of the study in Peru, participants provided sputum samples for TB testing using the WHO-endorsed Xpert MTB/RIF Ultra (Xpert Ultra) (Cepheid, Sunnyvale, CA, USA) and bacterial culture (BD BACTEC MGIT, BD, Franklin Lakes, NJ, USA).[25] Xpert Ultra tests were repeated if initial testing produced an error, indeterminate, or a 'Trace' semi-quantitative result. MGIT culture was repeated if the initial test was contaminated. Participants were TB positive if either of the following conditions were met: a positive sputum Xpert Ultra result, a MGIT positive, or two Trace positive results on Xpert Ultra. Participants were considered PTB negative if none of the tests had a positive result and at least two sputum tests (Xpert Ultra, Xpert Ultra repeat, MGIT, or MGIT repeat) were negative.

Participants recruited in Montreal provided nasopharyngeal samples that were stored at 2-8°C and tested at the CHUM using Biofire Multiplex PCR Respiratory Panel 2.1 (Biomerieux, Marcy-L'Étoile, France).[26] This panel tests for common respiratory pathogens, including viruses (adenovirus, COVID-19, coronaviruses, influenza A, influenza B, metapneumovirus, parainfluenza viruses, rhinovirus, respiratory syncytial virus [RSV]) and bacteria (Bordetella parapertussis, Bordetella pertussis, Chlamydia pneumoniae, Mycoplasma pneumoniae). Given the low tuberculosis prevalence in Montreal, participants at this site were not tested for TB.

6.5.4. Cough recording procedures

Cough recording procedures were standardized across all study sites. Study Android smartphones were used to record solicited coughs upon participant enrollment, at the health facility. In Lima, the Xiaomi Redmi 9A was used while in Montreal the Motorola G6 smartphone was used.

Cough recordings were performed using the Hyfe Research smartphone app, which has been previously validated for cough detection.[27] The embedded cough detection algorithm assigns a predictive score from 0 to 1, with a higher score indicating that the sound recorded is more likely to be a cough. Only cough sounds with a probability score of ≥0.8 were included in this analysis. The cough sounds are recorded in 0.5s bursts and only capture the cough sound, preventing the recording of conversations. In both settings, a trained research nurse assisted the participant in recording their cough sounds. Participants were prompted to cough forcefully into the smartphone's microphone (held approximately 30cm away from their mouth), with instructions to produce 5-10 coughs or as many times as was physically comfortable. All recordings were performed in an isolated setting to minimize the amount of background noise. Healthcare personnel were trained to assist participants with the recording and wore appropriate personal protective equipment, including N95 face masks.

The sampling rate was 44.1 Hz and files had a 16-bit PCM format. Cough sound .WAV files were uploaded to a secured study server that was only accessible by the study staff. A unique identifier was used for recording coughs that could only be linked to the participant's demographic and clinical data by study personnel.

6.5.5. Acoustic feature extraction and descriptive analysis

Prior studies performing cough-based ML classification have used different audio features capable of effectively representing acoustic signals.[6,28,29] Following literature review and expert consultation with co-author GPK, we extracted 75 acoustic features. These characteristics can be grouped into two main categories: time-based and frequency-based features. Time-based features measured how the sound's intensity changed throughout the duration of the cough, similar to how

a volume meter moves up and down while someone is speaking. Frequency-based features captured the pitch and tonal qualities of the cough, much like how one can distinguish between a deep, rumbling cough and a high-pitched, wheezing one. Features included standard acoustic parameters (zero-crossing rate, energy, intensity, and various spectral measurements), 13 MFCCs, their 13 first-order temporal derivatives (ΔMFCC), and 40 filter bank (FBank) features. Detailed descriptions of these features are provided in **Table S6.1**. Given the non-stationary nature of cough signals and their rapid temporal variations, the features were analyzed using short segments known as frames. This processing techniques assumes that the audio signal is stationary over time within the analysis frame. To achieve an optimal balance between meaningful feature representation and achieving the stationary assumption, frames of 50 milliseconds (ms) in length were used, advancing across the cough signal in 25-ms increments (the frame rate), resulting in a 50% overlap between consecutive frames. Each frame of the audio signal was processed using a Hamming windowing to prevent spectral leakage,[30] which can occur when abrupt frame boundaries create artificial frequency components that distort the signal's true spectral content. This windowing technique assists by gradually attenuating the signal amplitude at frame boundaries while preserving the central signal information.

Extracted features from all frames are statistically summarized per cough sound to reduce dimensionality. Statistical summaries used were the mean, standard deviation, median, 25th percentile (q1), 75th percentile (q3), skewness, kurtosis, minimum, maximum, and range. This aggregation effectively reduces the information to a single vector that represents the overall statistical behavior of the features across the entire cough signal.

Box plots were used to visualize the distribution of acoustic features extracted from cough recordings at both sites (Lima and Montreal). To compare audio features between Lima and Montreal, independent Wilcoxon tests of 15 acoustic parameters were performed (only the first and last MFCC, ΔMFCC, and FBank were compared). The Bonferroni correction method was applied to control the family-wise error rate and reduce the probability of type I errors (false positives) that naturally increase when conducting multiple statistical tests. While this conservative approach increases the risk of type II errors (false negatives), we prioritized minimizing false discoveries given the exploratory nature of these acoustic analyses. Statistical significance was set

at α=0.05, resulting in an adjusted significance threshold of 0.003 (0.05/15) for individual comparisons. Additional comparisons were made within each dataset between COVID-19 positive and COVID-19 negative individuals.

6.5.6. Machine learning approach

An eXtreme Gradient Boosting (XGboost) classifier was employed as the machine learning approach for training on the extracted features.[31] XGBoost is a non-parametric decision tree ensemble algorithm used primarily for classification tasks for complex tabular datasets. The ensemble approach combines multiple simpler models (in this case, decision trees) that work together to make predictions.[32] Each new tree in the ensemble learns from the mistakes of previous trees, focusing on correcting misclassified samples by assigning higher weights, ultimately producing a robust classifier that is more accurate than any single decision tree could be alone (**Figure 6.1**). This choice of classifier was informed by the complexity of the dataset, expert consultation (GPK), and previous cough classification studies, which demonstrated XGboost and other ensemble boosting model's ability to achieve high performance while maintaining interpretability.[33–35] The model was developed in Python version 3.12 using the package xgboost.[31] Since the study's objective is not aimed at generating the best performing classifier, but instead to investigate differences in acoustic features and model performance between, no other ML classifier was tested.

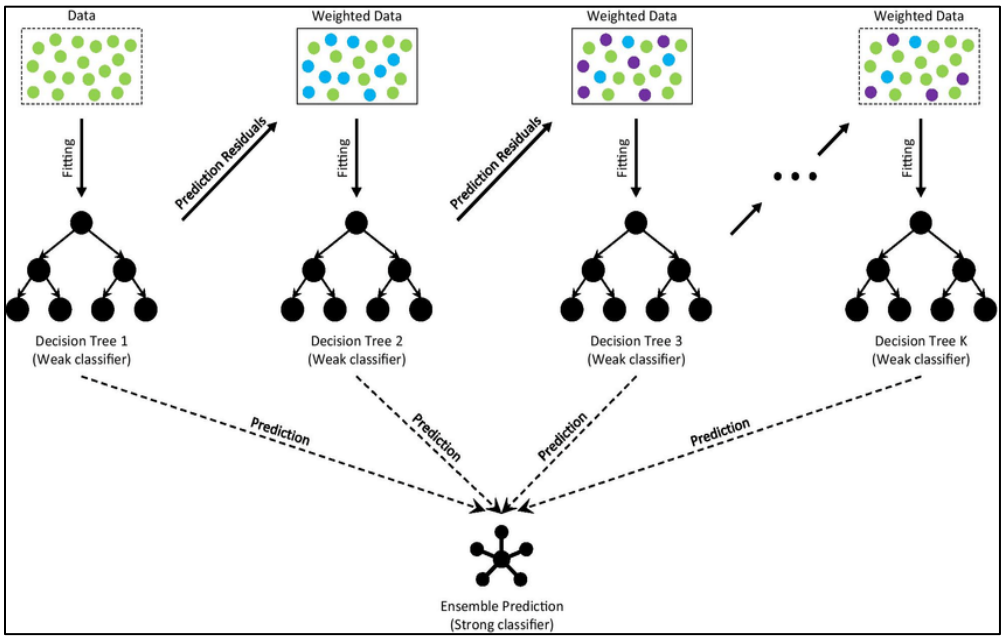


Figure 6.1. Schematic representation of the Gradient Boosting (GB) process that occurs as part of Extreme Gradient Boosting (XGBoost) architecture. The algorithm sequentially builds decision trees where each subsequent tree learns from the errors of previous trees. Initially, data points (shown in green) are processed through the first decision tree. For subsequent trees, the data is reweighted (shown in different colors) based on prediction residuals. Each tree is a weak classifier, but the ensemble combination of all trees' predictions produces a strong classifier, demonstrating XGBoost's ensemble boosting principle. In addition to the GB depicted, XGBoost has additional regularization terms that reduce overfitting. (Source: Deng et al.[36])

XGBoost uses a loss function (**Equation 6.1**) for optimizing model performance by simultaneously minimizing prediction errors and controlling model complexity through regularization. The first term, $\sum_i l(\hat{y}_i, y_i)$,, is the loss function that measures the sum of the difference between predicted values ($\hat{y}_i$) and actual values (y_i), quantifying how well the model makes predictions. The second term, $\sum_k \Omega(f_k)$, is the regularization component where Ω measures the complexity of each individual decision tree (f_k) in the ensemble, and the sum (Σ_k) accumulates these complexity penalties across all trees. The complexity measure Ω considers factors such as the number of leaves (prediction outputs), effectively penalizing trees that become too elaborate. By adding these complexity penalties to the objective function, XGBoost automatically favors simpler tree structures that are less likely to overfit the training data.

Equation 6.1. $$L(\Phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

Hyperparameters are model configuration settings that control the learning process and must be set prior to training. **Table 6.1** presents the optimized hyperparameters, their functional definitions, and the ranges over which they were tuned. These include parameters controlling tree construction (n_estimators, max_depth), regularization terms (reg_lambda, reg_alpha), and various optimization settings that help prevent overfitting and manage model complexity.

Table 6.1. Hyperparameters optimized in the XGBoost model.

Hyperparameter	Definition	Range
n_estimators	The total number of sequential prediction models (trees) combined to make the final prediction.	200, 300, 400, 500
max_depth	The maximum number of sequential decisions allowed in each tree model.	3, 4, 5
gamma	A threshold value that determines if further splitting of data is worthwhile. Higher values result in more conservative (simpler) models.	0.3, 0.5, 0.7
reg_lambda	A penalty term that helps prevent extreme prediction values, leading to more stable and generalizable models.	3, 5, 7
reg_alpha	A penalty term that helps identify and focus on the most important predictive features.	0.5, 1, 2
learning_rate	Controls how much each new tree contributes to the final prediction. Smaller values make the model more robust but require more trees.	0.05, 0.01, 0.1
colsample_bytree	For each tree, randomly uses only a portion of the available predictive factors. This helps prevent the model from becoming overly dependent on specific factors.	0.5, 0.7, 0.9
colsample_bylevel	At each decision point in a tree, randomly uses only a portion of the available predictive factors, encouraging the model to consider diverse relationships in the data.	0.5, 0.7, 0.9

The cough sound and clinical datasets were divided into a 4:1 ratio for model training and validation, respectively. This split was stratified by participant to ensure that all cough sounds from the same individual remained in the same subset. For experiments evaluating models on combined Lima and Montreal datasets, the split was additionally stratified by setting. To address class

imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was applied to the feature space of the training set.[37] Hyperparameter tuning (**Table 6.1**) was then performed using grid search with a nested, stratified 5-fold cross-validation on the training set. A final model was evaluated on the test set. To enhance robustness, this training and validation process was repeated across five different random seeds.

To quantify and interpret the contribution of individual features to the model's predictions, a SHapley Additive exPlanations (SHAP) analysis was performed.[38,39] SHAP calculates each feature's contribution to the model's predictions by systematically evaluating how each feature, alone or in combination with others, influences the model's output. This approach allows for ranking features by their relative importance and understanding their individual impact on model predictions. For this analysis, the shap package was used in Python.[40]

6.5.7. Neural network approach

For each participant, the raw .WAV cough sound files were concatenated into single audio files in temporal order of when the person coughed (**Figure 6.2**). After testing different audio chunks and overlap lengths, the audio files were segmented into 3-second chunks with a 1-second overlap. If an audio file was shorter than the chunk length (3 seconds), it was repeated until it reached the required length. SMOTE was applied to the feature space of the training set. This approach ensured a consistent data format and provided the best balance between longer recording periods and sufficient data points for analysis.

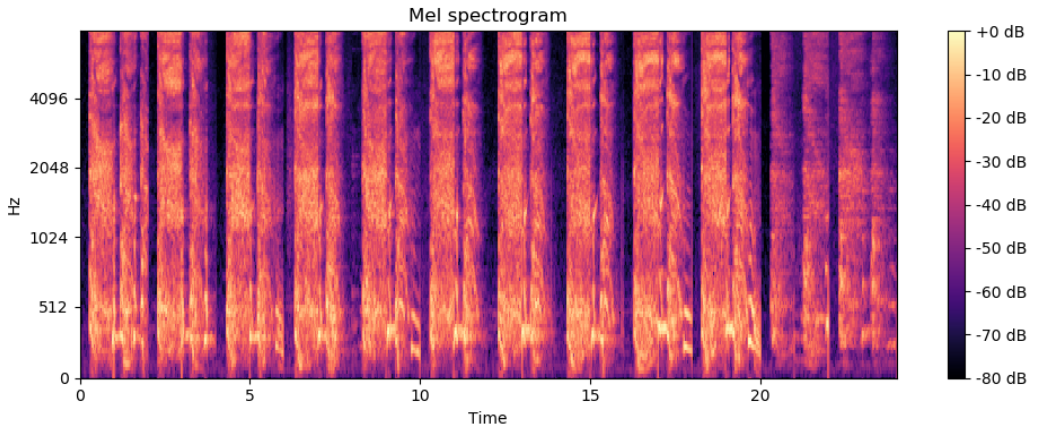


Figure 6.2. Mel spectrogram visualization of concatenated cough audio recordings. The spectrogram shows multiple 3-second chunks of cough audio data with 1-second overlap between segments. Each vertical segment represents a 3-second chunk, with darker bands indicating the transitions between chunks. The frequency components (y-axis, 0-4096 Hz) and their intensities (color scale, -80 to 0 dB) reveal the acoustic characteristics of the cough sounds across time. The repetitive patterns visible in some segments likely indicate where shorter audio files were repeated to reach the standard 3-second chunk length

The raw audio signals were converted into Mel-spectrograms using the Fbank module from SpeechBrain—a PyTorch-powered toolkit specializing in audio AI analyses.[33] This process includes framing the audio into overlapping windows, applying a short-time Fourier transform (STFT) to each window,[34] and then mapping the resulting frequency domain representation onto the Mel scale using a filter bank. The STFT is configured with a window length of 25 ms, a hop length of 10 ms, and 400 Fast Fourier Transform points.

The ECAPA-TDNN (Emphasized Channel Attention, Propagation, and Aggregation Time Delay Neural Network) model was employed for COVID-19 cough classification. This architecture, known for its state-of-the-art performance in audio recognition tasks,[35] is designed to capture intricate temporal and spectral patterns in audio data. The ECAPA-TDNN model consists of multiple convolutional layers with varying kernel sizes and dilation rates, allowing it to effectively model both short-term and long-term dependencies in the audio signal.

SpeechBrain was used to train, validate, and test the ECAPA-TDNN algorithms. The segmented audio files were split into a ratio of 4:1 for model training and testing. The input features were extracted as Fbank representations with 80 Mel-frequency bands, a value determined empirically through testing different numbers of frequency bands. This choice represents a balance between capturing sufficient spectral information and avoiding model overfitting through excessive parameterization, as lower numbers of bands resulted in reduced model performance.[36] These features are normalized and processed through the ECAPA-TDNN layers to produce a fixed length embedding of size 512. This embedding is passed through a fully connected layer followed by softmax output layer, which generates probability scores for the COVID-positive and COVID-negative classes. The model training employed an additive angular margin loss function to enhance discriminative power by maximizing the margin between different classes in the feature space.

6.5.8. Classification experiments

A series of experiments were conducted to evaluate the models' performance and generalizability across different datasets (**Table 6.2**). Various training and testing configurations were explored, including within-dataset validation (e.g., training and testing on Lima dataset) and external validation (e.g., training on Lima dataset and testing on Montreal dataset). An additional within-dataset experiment explored the performance of models trained and tested on the combined dataset (Lima and Montreal).

Within-dataset validation (e.g., training and testing on Lima dataset) were performed using both the LR and ECAPA-TDNN model configurations. For comparison, an additional within-dataset model was trained using a combination of data from Montreal and Peru. Features used for this analysis derived from the same dataset.

Table 6.2. Summary of classification experiments.

Experiment	Training set	Test set
Within-dataset	Lima Montreal Combined	Lima Montreal Combined
External validation	Lima Montreal	Montreal Lima

For XGBoost, experiments listed in **Table 6.2** were performed three times with different features inputs: using only audio features ("Audio"), using only clinical and demographic features ("Clinical"), and once using audio and clinical/demographic features ("Audio+Clinical"). The clinical and demographic information used includes participant age, sex, smoking status, body mass index, cough duration, additional symptoms (e.g., fever, headache), and country (for the combined evaluation) that were collected at time of enrollment. For the ECAPA-TDNN, only experiments using audio features were performed.

6.5.9. Model evaluation

The primary performance metric utilized for model evaluation was the AUC, computed using the pROC package in R.[45] The AUC measures the model's ability to discriminate between positive and negative classes (e.g., disease present vs. absent) across all possible classification thresholds. This is particularly valuable as it assesses model performance independently of any single decision threshold, making it robust for comparing models across different populations where optimal classification cutoffs may vary. Supplementary performance metrics included the specificity at 70% and 80% sensitivity, accuracy, and F1 score. Specificity at fixed sensitivity levels (70% and 80%) indicates the model's ability to correctly identify negative cases while maintaining a predetermined rate of correct positive case identification. Accuracy is the proportion of correct predictions (both true positives and true negatives) among all predictions made (**Equation 6.2**), providing a general measure of model performance. The F1-score is the harmonic mean of precision and recall (**Equation 6.3**), providing a balanced measure of model performance that considers both false positives and false negatives, which is particularly useful when dealing with imbalanced datasets. Both accuracy and F1-score were calculated based on a threshold of 0.5, meaning that prediction scores >0.5 were COVID-19 positive. To account for variability introduced by different random seeds in data partitioning, all metrics were computed for each seed iteration. The central tendency of each metric was represented by the mean across all seeds, with the accompanying standard error (SE) serving as a measure of dispersion.

Equation 2. $$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}}$$

Equation 3. $$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{\text{TP}}{\text{TP} + {}^{1}/_{2}(\text{FP}+\text{FN})}$$

6.5.10. Sub-analysis

To further evaluate the discriminative ability of the XGBoost classifier, a sub-group analysis was conducted using prediction scores from the above-motioned within-dataset XGBoost models. Based on microbiological test results, participants were categorized into three groups: 1) COVID-19 positive, 2) positive for other tested respiratory diseases (e.g., TB, influenza, etc.), and 3) negative for all tested diseases. The Wilcoxon test was employed to assess statistical differences in XGBoost prediction scores among these groups. In alignment with the cough classifier's training methodology, individuals co-infected with COVID-19 and other diseases were classified as COVID-19 positive.

6.5.11. Ethics

All participants provided written informed consent. In Lima, Peru, the studies involved in the recruitment phases received ethical approvals from the Comité Institucional de Ética en Investigación at Universidad Peruana Cayetano Heredia (Phase 1: SIDISI 202931/SIDISI 206951, and Phase 2: SIDISI 210857) and the McGill University Health Centre Research Ethics Board (2021-6866). The studies were registered in the PRISA repository at Instituto Nacional de Salud in Peru (Phase 1: EI00000001484/EI00000002576, and Phase 2: EI00000003221).

In Montreal, Canada, ethical approval was obtained from the Comité d'éthique à la recherche du Centre de Recherche du Centre Hospitalier de l'Université de Montréal under IRB # MP-02-2022-10470, 21.351.

6.6. Results

6.6.1. Participants

Overall, 610 participants were enrolled in the study, of whom 310 were recruited in Lima and 300 in Montreal. In Lima, 150 individuals were recruited in Phase 1 and an additional 160 were recruited during Phase 2. Four individuals were excluded from Lima due to a loss of cough recording data (n=1; Phase 1) or because both COVID-19 tests produced invalid results (n=3; Phase 2). In Montreal, one person was excluded due to a loss of cough recording data. The final sample size for this study was 605 participants. From these participants, 10,721 individual cough recordings were used for model training and evaluation: 5,889 from Lima and 4,832 from Montreal.

Baseline demographic and clinical features are summarized in **Table 6.3** by setting. In general, participants recruited in Lima presented with more severe symptoms, including longer cough duration days and a higher prevalence of symptoms. COVID-19 was more prevalent in Montreal, with 34.1% of participants receiving a positive diagnosis compared to 11.8% of participants in Lima. In Lima, all individuals enrolled in Phase 1 and Phase 2 were additionally tested for TB. A total of 94 (94/306; 31.3%) had a positive TB diagnosis. During Phase 2 of recruitment (n=157), additional testing was performed for influenza A, influenza B, and RSV. Nine (9/157; 5.7%) individuals were influenza A positive and two (2/157; 1.3%) were RSV positive. In Montreal, the multiplex panel was performed on all participants. The most common other infections were due to rhinovirus (n=50; 16.7%) followed by influenza A (n=10; 3.3%). Of the 605 participants, 11 (1.8%) were co-infected with COVID-19 and another pathogen: 5/306 (1.6%) in Lima and 6/299 (2.0%) in Montreal.

Table 6.3. Baseline demographic and clinical characteristics of participants recruited in Lima and Montreal.

	Lima n=306	**Montreal** n=299	**Total** n=605
Age (years), median (IQR)	36 (24, 52)	40 (30, 53)	38 (27, 52)
Sex, n (%)			
Female	135 (44.1)	218 (72.9)	353 (58.3)
Male	171 (55.9)	81 (27.1)	252 (41.7)
Smoking status, n (%)			
Current	10 (3.3)	38 (12.7)	48 (7.9)
Former	23 (7.5)	42 (14.0)	65 (10.7)
Never	273 (89.2)	219 (73.2)	492 (81.3)
BMI, median (IQR)	23.9 (21.7, 26.8)	25.7 (22.1, 29.3)	24.8 (21.9, 27.8)
Cough duration (days), median (IQR)	21 (10, 35)	3 (2, 7)	8 (3, 25)
Symptoms, n (%)			
Headache	241 (78.8)	143 (47.8)	384 (63.5)
Fever	161 (52.6)	66 (22.1)	227 (37.5)
Sore throat	237 (77.5)	154 (51.5)	391 (64.6)
Loss of smell/taste	25 (8.2)	15 (5.0)	40 (6.6)
Muscle pain	218 (71.2)	80 (26.8)	298 (49.3)
Fatigue	145 (47.4)	130 (43.5)	275 (45.5)
Nausea	87 (28.4)	9 (3.0)	96 (15.9)
Diarrhea	68 (22.2)	15 (5.0)	83 (13.7)
COVID-19, n (%)	36 (11.8)	102 (34.1)	138 (22.8)
Other diagnoses			
Adenovirus	NA	1 (0.3)	1 (0.2)
Bordetella parapertussis	NA	0 (0.0)	0 (0.0)
Bordetella pertussis	NA	0 (0.0)	0 (0.0)
Chlamydia pneumoniae	NA	0 (0.0)	0 (0.0)
Influenza A	9 (5.7)*	10 (3.3)	19 (3.1)
Influenza B	0 (0.0)*	0 (0.0)	0 (0.0)
Metapneumovirus	NA	6 (2.0)	6 (1.0)
Mycoplasma pneumoniae	NA	0 (0.0)	0 (0.0)
Other coronavirus**	NA	11 (3.7)	11 (1.8)
Parainfluenza virus	NA	7 (2.3)	7 (1.2)
Rhinovirus	NA	50 (16.7)	50 (8.3)
RSV	2 (0.4)*	2 (0.7)	2 (0.3)
TB	94 (31.3)	NA	94 (15.5)

*Influenza A, Influenza B, and RSV testing was only performed for participants recruited during Phase 2 in Lima (n=157 participants).

**Coronavirus HKU1, coronavirus NL63, Coronavirus 229E, Coronavirus OC43.

BMI, body mass index; IQR, interquartile range; NA, not available; n, number; RSV, respiratory syncytial virus; TB, tuberculosis

6.6.2. Feature distribution

Differences in mean acoustic features among individuals recruited in Lima and Montreal are illustrated in **Figure 6.3**. Detailed definitions for each acoustic feature are presented in **Table S6.1**. While a statistically significant difference can be seen between most features, some features show more pronounced differences. For instance, the box plots for spectral features (centroid, spread, entropy, flux, rolloff) all show minimal overlap. These differences between populations are more pronounced than within-dataset differences between COVID-19 positive and COVID-19 negative individuals, as illustrated for Lima in **Figure S6.1** and Montreal in **Figure S6.2**.

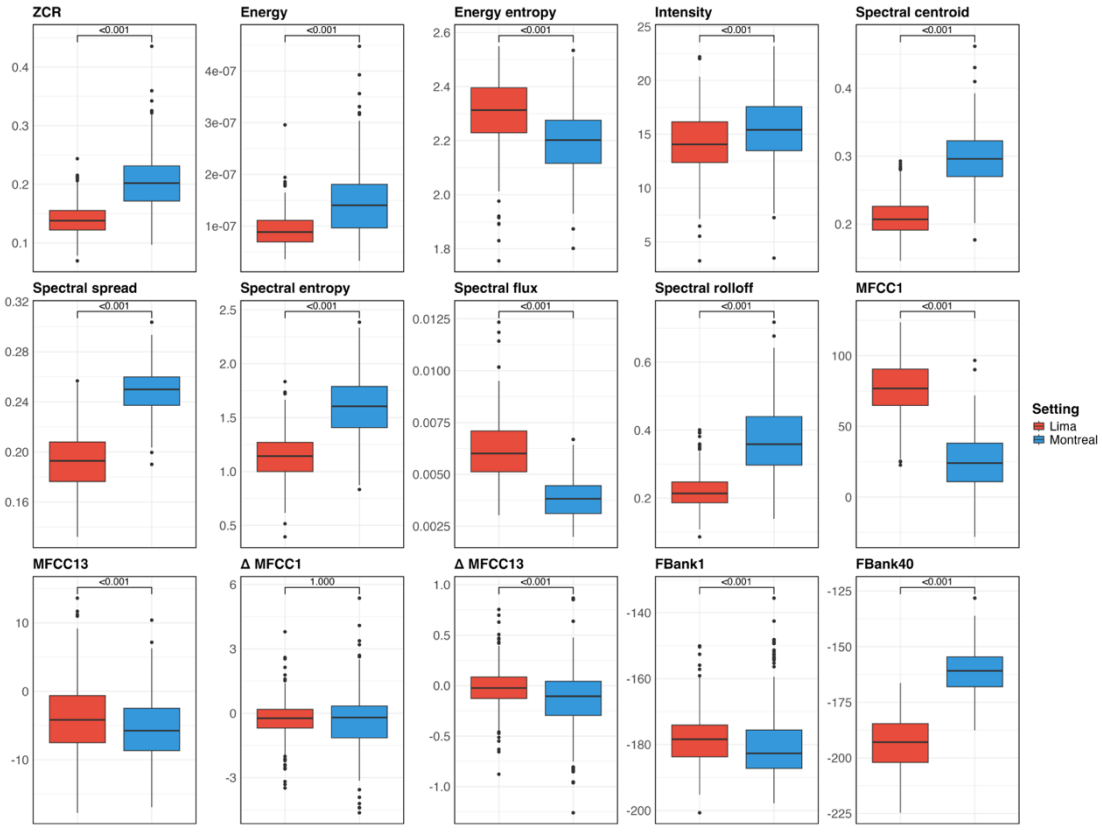


Figure 6.3. Comparison of mean acoustic cough feature distributions among participants in Lima, Peru (red; N=306 participants) and Montreal, Canada (blue; N=299 participants). Only the first and last MFCC, ΔMFCC, and FBank are illustrated. Statistically significant differences (P<0.001) based on t-tests with Bonferroni correction are observed for all features except the ΔMFCC1. ZCR, zero-crossing rate; MFCC, Mel-frequency cepstral coefficients; FBank, filter-bank. Definitions for all features are presented in **Table S6.1**.

6.6.3. Classifier results

Model performance varied across datasets (**Figure 6.4**). When using the same dataset for training and testing (**Figure 6.4A**), distinct patterns emerged between Lima and Montreal cohorts. In the Lima dataset, audio-based models showed superior performance, with the 'Audio' and 'Audio+Clinical' models achieving mean AUCs of 0.71 (standard error [SE] ± 0.08) and 0.73 (SE ± 0.06), respectively, both outperforming the 'Clinical' model (0.63; SE ± 0.07). In contrast, the Montreal dataset showed different trends, where models incorporating clinical features performed better: both the 'Audio+Clinical' (0.66; SE ± 0.05) and 'Clinical' (0.64; SE ± 0.06) models achieved higher AUCs compared to the 'Audio' model alone (0.53; SE ± 0.04). SHAP analyses revealed distinct combinations of audio features that enhanced predictive performance across the "Audio" models (**Figure S6.3**). The "Audio+Clinical" models demonstrated similar variations in audio feature importance (**Figure S6.4**), while clinical parameters including cough duration, age, and BMI remained consistently important predictors across Lima, Montreal, and the combined dataset analyses.

All three within-dataset models in Lima demonstrated poor F1-scores <0.25 (**Table S6.2**), indicating poor performance in detecting the minority case when the threshold for COVID-19 positive was set at >0.5. Unlike the Peru dataset, the Montreal cohort demonstrated more balanced performance metrics, with F1-scores approximating 0.50 across all three models, reflecting the more balanced distribution of COVID-19 cases in the Montreal dataset.

When model training combined the Montreal and Lima datasets, the mean AUC for the "Audio" model was 0.67 (SE ± 0.03). Combined "Clinical" and "Audio+Clinical" improved with a mean AUC of 0.71 (SE ± 0.03) and 0.72 (SE ± 0.03) respectively. However, when disaggregating the combined results by setting (**Figure S6.5**), a difference in AUCs was observed between sites. For the "Audio" model, the AUC in Lima remained the same as the overall AUC (0.67; SE ± 0.04) however the AUC in the Montreal dataset decrease to 0.59 (SE ± 0.09). This reduction occurred for both Lima and Montreal in the "Audio+Clinical" model with AUCs of 0.62 (SE ± 0.09) and 0.53 (SE ± 0.05) respectively.

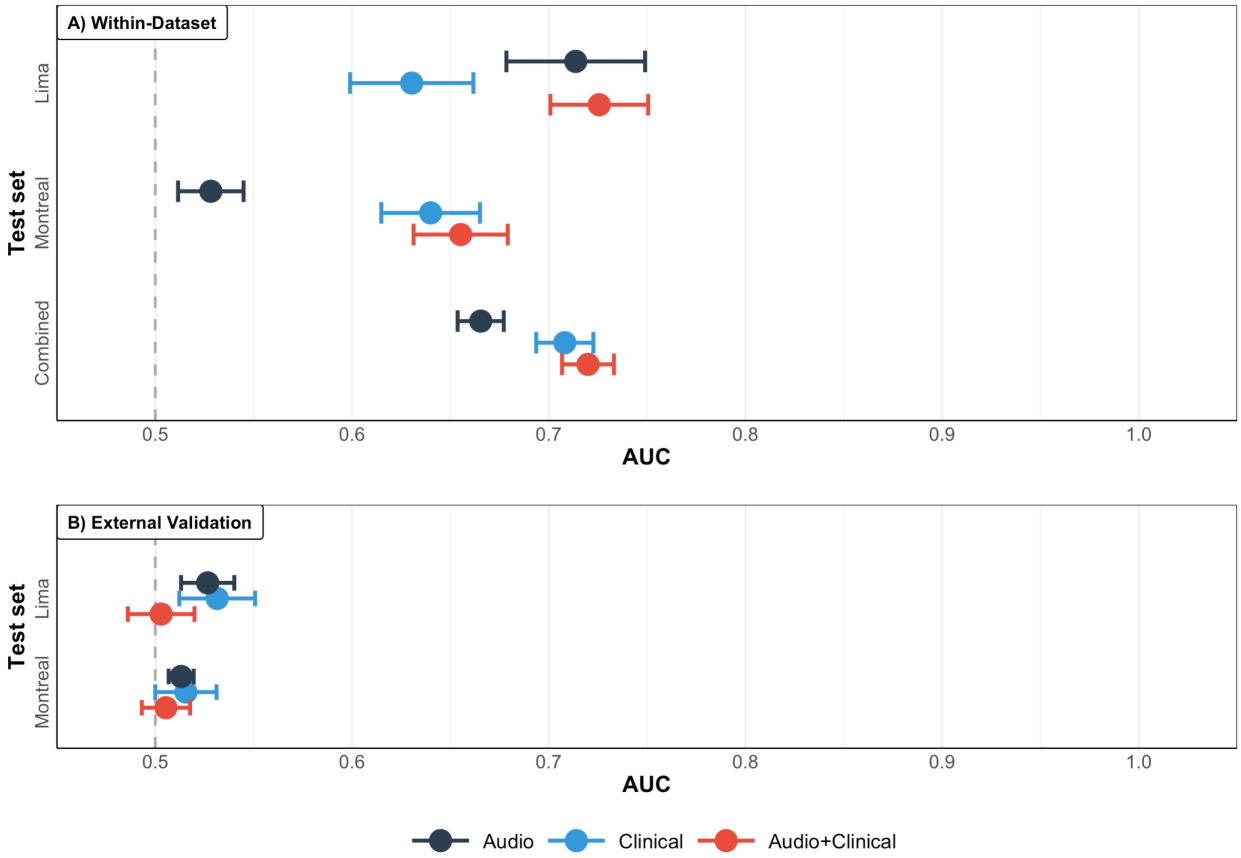


Figure 6.4. Comparison of model performance using three feature sets in XGBoost classifiers using datasets from Lima, Peru and Montreal, Canada. A) Within-dataset validation demonstrates AUC scores across Lima, Montreal, and Combined datasets. B) External validation results show generalizability of models across different testing populations. Models were evaluated using Area Under the Curve (AUC) metrics averaged across five independent seeds, with error bars indicating the standard error across seeds.

External validations (**Figure 6.4B**) revealed substantial performance decreases. Models optimized and trained on one dataset performed poorly when applied to the other. "Audio" models trained in Montreal and validated in Lima achieved a mean AUC of 0.53 (SE ± 0.03) while the "Audio" models trained in Lima and validated in Montreal achieved a mean AUC of 0.51 (0.01). Models that used "Clinical" and "Audio+Clinical" variables reported similar poor performance.

Results from the neural network ECAPA-TDNN assessment of "Audio" models are presented in the **Table S6.3**. As observed in the XGBoost "Audio" model, the mean AUC was higher in the

Lima dataset (0.60; SE ± 0.07) compared to the Montreal dataset (0.54; SE ± 0.02). External validation in the counterpart dataset resulted in a decrease in performance in both models as was observed in the XGBoost models.

6.6.4. Sub-analysis: Logistic regression COVID-19 prediction score based on underlying disease

The distribution of COVID-19 prediction scores from the XGBoost within-dataset models (models presented in **Figure 6.4A**) reveals distinct patterns across settings, underlying disease status, and model configurations ("Audio", "Clinical", and "Audio+Clinical") (**Figure 6.5**). The prediction scores for the XGBoost models ranged from 0 to 1.0, with a higher score indicating that the observation is more likely to be labeled as COVID-19 positive.

Prediction scores in the Lima dataset were predominantly below 0.50 across all three models, reflecting the previously-mentioned low F1-score. For the "Audio" model, COVID-19 positive cases showed the highest median prediction score (0.49; IQR: 0.46, 0.55), which was significantly higher than both other diseases (0.44; IQR: 0.35, 0.49; p=0.001) and disease-negative cases (0.46; IQR: 0.39, 0.51; p=0.04) (**Figure 6.5A**). While statistically significant, these differences were modest in absolute terms. The Montreal dataset showed even smaller distinctions between disease groups, with identical median scores for COVID-19 positive cases and other diseases (0.50; IQR: 0.48, 0.49 and 0.50; IQR: 0.49, 0.52 respectively; p=0.101).

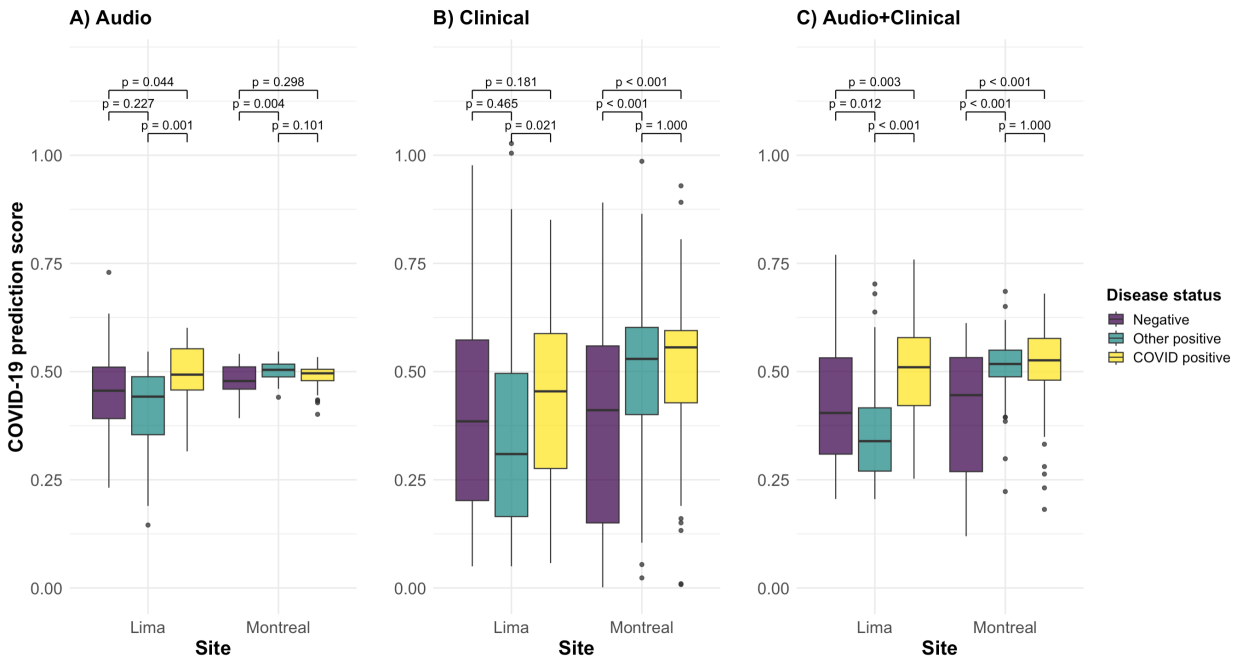


Figure 6.5. Distribution of COVID-19 prediction scores from the XGBoost classifier's within-dataset validation, stratified by disease status (COVID-19 positive, other disease positive, and negative for all diseases tested) and study site. The XGBoost models were trained and tested on the same dataset. A) "Audio" models only used audio features as inputs. B) "Clinical" models only used clinical and demographic features as inputs. C) "Audio+Clinical" models combined audio features with clinical and demographic features. P-values from Wilcoxon tests with Bonferroni correction for multiple comparisons indicate statistical significance of between-group differences. Horizontal bars show pairwise comparisons, with p-values displayed above each comparison. Statistical significance was set at $p<0.05$.

The "Clinical" models demonstrated greater spread in prediction score distributions (**Figure 6.5B**). In Lima, the increased overlap between distributions reduced the statistical significance compared to the "Audio" models. Montreal's results showed higher prediction scores for COVID-19 positive cases compared to negative cases (0.56; IQR: 0.45, 0.59 versus 0.41; IQR: 0.15, 0.53; $p<0.001$), though no significant difference emerged between COVID-19 positive and other disease cases (0.53; IQR: 0.41, 0.60; $p=1.000$).

The combined "Audio+Clinical" model (**Figure 6.5C**) enhanced performance in Lima, most notably between COVID-19 positive and other disease groups (0.53; IQR: 0.44, 0.61 versus 0.35;

IQR: 0.28, 0.43 respectively; p<0.001), with a less pronounced difference between COVID-19 positive and disease-negative groups (0.40; IQR: 0.30, 0.51; p=0.012). Montreal's results paralleled the "Clinical" model pattern, showing better discrimination between COVID-19 and disease-negative groups (0.52; IQR: 0.47, 0.56 vs 0.44; IQR: 0.27, 0.53; p<0.001) but no significant difference between COVID-19 positive and other disease groups (0.52; IQR: 0.48, 0.57; p=1.000).

6.7. Discussion

This study demonstrates that cough acoustic features display significant heterogeneity between individuals recruited in Lima, Peru and Montreal, Canada. These differences underscore the potential influence of population-specific demographic and physiological characteristics on cough acoustics; an aspect of cough classification research that to date has been largely unexplored. Consideration must also be given to the impact of technological variability between smartphones used for recording in different settings, as differences in audio capture capabilities could significantly affect the measurement of acoustic features. The heterogeneity between cohorts is further reflected in the SHAP analyses of XGBoost models. Although both datasets predominantly relied on features derived from Mel-frequency analysis (specifically FBanks and MFCCs), consistent with previous research,[9,46,47] the relative importance of specific acoustic features for optimal COVID-19 classification differed between the two populations.

A critical finding from this study is the poor transferability of cough-based COVID-19 detection models between settings, regardless of the analytical approach used (machine learning or neural networks). This limitation aligns with previous studies examining model transferability across crowdsourced datasets.[48] Notably, the addition of clinical features to audio data did not improve external validation performance, suggesting that factors such as demographics and clinical presentations influence model transferability.

Models that combined the Lima and Montreal datasets resulted in relatively high AUC scores. However, when the results from the combined model were disaggregated by study site and analyzed separately, differences in performance metrics were observed between sites. COVID-19

prediction scores assigned to the observations from the combined dataset also differed: Montreal participants received higher COVID-19 prediction scores compared to Lima participants, even when using only audio features without any location information. This bias suggests the model may be detecting subtle recording characteristics specific to each site and incorrectly associating them with COVID-19 status, possibly influenced by Montreal's higher COVID-19 prevalence. Rather than learning true acoustic signatures of COVID-19, the model might be learning to recognize site-specific recording artifacts and erroneously using these to make predictions. These findings highlight critical technological challenges for developing globally applicable cough-based diagnostic algorithms, as variations in smartphone hardware specifications, microphone quality, and ambient recording conditions across different settings can introduce systematic biases to the classifier.

Analysis of the COVID-19 classifier performance across different disease profiles revealed distinct patterns between the two cohorts. In Lima, participants with other diseases (predominantly TB) received notably lower COVID-19 prediction scores compared to both COVID-19 positive and disease-negative individuals. In contrast, the Montreal cohort showed similar prediction scores between COVID-19 positive cases and those with other diseases, which were primarily viral respiratory infections. These findings align with previous research suggesting that algorithms can more effectively distinguish COVID-19 coughs from bacterial infections than from other viral infections, likely due to distinct pathological mechanisms.[49–53] The observed differences highlight how local disease prevalence patterns may influence cough classifier performance, particularly when deploying models across settings with varying epidemiological landscapes.

This study has several limitations. It is important to note that the primary objective of this study was to investigate audio features across different populations rather than to develop the best-performing COVID-19 classifier. As such, only one AI and one ML approach were explored. While this focused approach served our research goals, it also presents a limitation. Additional work would be necessary to explore alternative models and potentially achieve higher classification performance both within dataset and in external validation. There was also a risk of overfitting, which is particularly concerning when working with limited datasets and complex models. However, the study's use of nested cross-validation for hyperparameter tuning helped mitigate this

risk by providing unbiased performance estimates and ensuring that the validation data remained truly independent from the hyperparameter optimization process. Class imbalance emerged as a significant issue, particularly pronounced in the Lima dataset. While obtaining larger, more balanced datasets would undoubtedly lead to improved models, this solution may not always be feasible due to resource constraints. An alternative could involve applying advanced generative AI techniques to create synthetic cough sounds for the minority class.[54] The use of different smartphones between the two sites with varying microphone qualities may have influenced the recorded cough sounds and, more critically, affected model performance in external validation scenarios. Although this presents a challenge for developing globally applicable algorithms, it's important to note that the phones selected were appropriate for local deployment, considering pricing and availability in each context. A notable limitation of this study lies in the sub-analysis evaluating the impact of co-circulating diseases on model performance across different sub-groups. Potential misclassification of participants may have occurred, particularly when labeling individuals as disease negative. This is because participants could not be comprehensively tested for all possible infectious diseases. This limitation was especially pronounced in the Peru cohort, where only four other diseases were considered in Phase 2, and only TB was additionally tested in Phase 1. Consequently, some participants classified as 'negative' may have had undetected respiratory infections, potentially influencing the interpretation of the model's performance and the true impact of co-circulating diseases on cough classification accuracy.

In conclusion, this study provides evidence that cough acoustics are population-specific, with models generated in one setting demonstrating poor external validity in another. These findings underscore the significant challenges in developing globally applicable algorithms for cough-based COVID-19 detection. Instead, our results suggest that efforts should be focused on developing and implementing local algorithms in settings where the epidemiological landscape presents clearer distinction between disease profiles. This local approach is already recommended for AI algorithms for reading chest X-rays,[55] and would likely yield more accurate and reliable screening tools for COVID-19 and potentially other respiratory diseases. Moving forward, there is a pressing need to transition from reliance on crowdsourced datasets with patient-reported outcomes to more robust reference standard testing. By addressing these aspects, researchers can enhance the validity and clinical applicability of cough-based diagnostics.

6.8. Acknowledgements

We extend our sincere gratitude to the study personnel at both the Lima, Peru and Montreal, Canada sites for their dedication in participant recruitment, data collection, and laboratory diagnostics processing. We are particularly grateful to our study participants, whose willingness to participate and valuable contributions made this research possible.

6.9. Contributions

Conceptualization (A.J.Z., S.A.R., M.P., C.U.G., S.G.L.), Peru data collection (A.J.Z., P.E.L., C.U.G.), Montreal data collection (A.J.Z., S.G.L.), data preparation and cleaning (A.J.Z.), data visualization (A.J.Z.), data analysis (A.J.Z., V.R., G.P.K., M.R., S.A.R.) writing – original draft (A.J.Z.), writing – review and editing (V.R., P.E.L., G.P.K., M.R., S.A.R., M.P., C.U.G., S.G.L.).

6.10. References

1. Sheridan C. COVID-19 spurs wave of innovative diagnostics. Nat Biotechnol. 2020;38(7):769-772. doi:10.1038/s41587-020-0597-x
2. Aly M, Alotaibi NS. A novel deep learning model to detect COVID-19 based on wavelet features extracted from Mel-scale spectrogram of patients' cough and breathing sounds. Informatics in Medicine Unlocked. 2022;32:101049. doi:10.1016/j.imu.2022.101049
3. Kapoor T, Pandhi T, Gupta B. Cough Audio Analysis for COVID-19 Diagnosis. SN COMPUT SCI. 2022;4(2):125. doi:10.1007/s42979-022-01522-1
4. Schuller BW, Coppock H, Gaskell A. Detecting COVID-19 from Breathing and Coughing Sounds using Deep Neural Networks. Published online 2020. doi:10.48550/ARXIV.2012.14553
5. Banerjee A, Nilhani A. A Residual Network based Deep Learning Model for Detection of COVID-19 from Cough Sounds. Published online 2021. doi:10.48550/ARXIV.2106.02348
6. Brown C, Chauhan J, Grammenos A, et al. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. Published online 2020. doi:10.48550/ARXIV.2006.05919
7. Chowdhury NK, Kabir MA, Rahman MdM, Islam SMS. Machine learning for detecting COVID-19 from cough sounds: An ensemble-based MCDM method. Computers in Biology and Medicine. 2022;145:105405. doi:10.1016/j.compbiomed.2022.105405

8. Anupam A, Mohan NJ, Sahoo S, Chakraborty S. Preliminary Diagnosis of COVID-19 Based on Cough Sounds Using Machine Learning Algorithms. In: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE; 2021:1391-1397. doi:10.1109/ICICCS51141.2021.9432324
9. Pahar M, Klopper M, Warren R, Niesler T. COVID-19 cough classification using machine learning and global smartphone recordings. Computers in Biology and Medicine. 2021;135:104572. doi:10.1016/j.compbiomed.2021.104572
10. Vrindavanam J, Srinath R, Shankar HH, Nagesh G. Machine Learning based COVID-19 Cough Classification Models - A Comparative Analysis. In: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE; 2021:420-426. doi:10.1109/ICCMC51019.2021.9418358
11. Adadi A, Berrada M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access. 2018;6:52138-52160. doi:10.1109/ACCESS.2018.2870052
12. Bhattacharya D, Sharma NK, Dutta D, et al. Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection. Sci Data. 2023;10(1):397. doi:10.1038/s41597-023-02266-0
13. Chaudhari G, Jiang X, Fakhry A, et al. Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough. Published online 2020. doi:10.48550/ARXIV.2011.13320
14. Xia T, Spathis D, Brown C, et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. In: ; 2021. Accessed September 3, 2024. https://openreview.net/forum?id=9KArJb4r5ZQ
15. Xiong H, Berkovsky S, Kâafar MA, Jaffe A, Coiera E, Sharan RV. Reliability of crowdsourced data and patient-reported outcome measures in cough-based COVID-19 screening. Sci Rep. 2022;12(1):21990. doi:10.1038/s41598-022-26492-5
16. MacLean ELH, Villa-Castillo L, Espinoza-Lopez P, et al. Integrating tuberculosis and COVID-19 molecular testing in Lima, Peru: a cross-sectional, diagnostic accuracy study. The Lancet Microbe. 2023;4(6):e452-e460. doi:10.1016/S2666-5247(23)00042-3
17. Barletta F, Otero L, de Jong BC, et al. Predominant Mycobacterium tuberculosis Families and High Rates of Recent Transmission among New Cases Are Not Associated with Primary Multidrug Resistance in Lima, Peru. J Clin Microbiol. 2015;53(6):1854-1863. doi:10.1128/JCM.03585-14
18. Institut national de santé publique du Québec. Données COVID-19 au Québec. Institut national de santé publique du Québec. Accessed November 4, 2024. https://www.inspq.qc.ca/covid-19/donnees
19. Schmidt K, Ben Moussa M, Buckrell S, et al. National Influenza Annual Report, Canada, 2022-2023: Canada's first fall epidemic since the 2019-2020 season. Can Commun Dis Rep. 2023;49(10):413-424. doi:10.14745/ccdr.v49i10a02
20. Public Health Agency of Canada. Tuberculosis in Canada: Infographic (2022). March 4, 2024. Accessed November 4, 2024. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/tuberculosis-canada-infographic-2022.html

21. Cao XJ, Fang KY, Li YP, Zhou J, Guo XG. The Diagnostic Accuracy of Xpert Xpress to SARS-CoV-2: A systematic review. J Virol Methods. 2022;301:114460. doi:10.1016/j.jviromet.2022.114460
22. Mostafa HH, Carroll KC, Hicken R, et al. Multicenter Evaluation of the Cepheid Xpert Xpress SARS-CoV-2/Flu/RSV Test. J Clin Microbiol. 2021;59(3):e02955-20. doi:10.1128/JCM.02955-20
23. Cepheid. Cepheid Receives Emergency Use Authorization from FDA for Rapid SARS-CoV-2 Test. News Release Archive. Accessed November 7, 2024. http://cepheid.mediaroom.com/2020-03-21-Cepheid-Receives-Emergency-Use-Authorization-from-FDA-for-Rapid-SARS-CoV-2-Test
24. Cepheid. Cepheid Receives Emergency Use Authorization for Xpert Xpress CoV-2/Flu/RSV plus. News Release Archive. Accessed November 7, 2024. http://cepheid.mediaroom.com/2021-09-14-Cepheid-Receives-Emergency-Use-Authorization-for-Xpert-Xpress-CoV-2-Flu-RSV-plus
25. World Health Organization. WHO Meeting Report of a Technical Expert Consultation: Non-Inferiority Analysis of Xpert MTB/RIF Ultra Compared to Xpert MTB/RIF.; 2017.
26. Creager HM, Cabrera B, Schnaubelt A, et al. Clinical evaluation of the BioFire® Respiratory Panel 2.1 and detection of SARS-CoV-2. Journal of Clinical Virology. 2020;129:104538. doi:10.1016/j.jcv.2020.104538
27. Galvosas M, Gabaldón-Figueira JC, Keen EM, et al. Performance evaluation of the smartphone-based AI cough monitoring app - Hyfe Cough Tracker against solicited respiratory sounds. F1000Res. 2023;11:730. doi:10.12688/f1000research.122597.2
28. Peeters, G. A large set of audio features for sound description (similarity and classification) in the cuidado project. CUIDADO Ist Project Report. 2004;54(0):1-25.
29. Sreeram ASK, Ravishankar U, Sripada NR, Mamidgi B. Investigating the potential of MFCC features in classifying respiratory diseases. In: 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS). IEEE; 2020:1-7. doi:10.1109/IOTSMS52051.2020.9340166
30. Podder P, Zaman Khan T, Haque Khan M, Muktadir Rahman M. Comparative Performance Analysis of Hamming, Hanning and Blackman Window. IJCA. 2014;96(18):1-7. doi:10.5120/16891-6927
31. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Published online 2016. doi:10.48550/ARXIV.1603.02754
32. Sagi O, Rokach L. Ensemble learning: A survey. WIREs Data Min & Knowl. 2018;8(4):e1249. doi:10.1002/widm.1249
33. Mouawad P, Dubnov T, Dubnov S. Robust Detection of COVID-19 in Cough Sounds: Using Recurrence Dynamics and Variable Markov Model. SN COMPUT SCI. 2021;2(1):34. doi:10.1007/s42979-020-00422-6
34. Bagad P, Dalmia A, Doshi J, et al. Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds. Published online 2020. doi:10.48550/ARXIV.2009.08790
35. Irawati ME, Zakaria H. Classification Model for Covid-19 Detection Through Recording of Cough Using XGboost Classifier Algorithm. In: 2021 International Symposium on

Electronics and Smart Devices (ISESD). IEEE; 2021:1-5. doi:10.1109/ISESD53023.2021.9501695

36. Deng H, Zhou Y, Wang L, Zhang C. Ensemble learning for the early prediction of neonatal jaundice with genetic features. BMC Med Inform Decis Mak. 2021;21(1):338. doi:10.1186/s12911-021-01701-9
37. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair. 2002;16:321-357. doi:10.1613/jair.953
38. Shapley LS. 17. A Value for n-Person Games. In: Kuhn HW, Tucker AW, eds. Contributions to the Theory of Games (AM-28), Volume II. Princeton University Press; 1953:307-318. doi:10.1515/9781400881970-018
39. Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst. 2014;41(3):647-665. doi:10.1007/s10115-013-0679-x
40. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. Published online November 25, 2017. doi:10.48550/arXiv.1705.07874
41. Ravanelli M, Parcollet T, Plantinga P, et al. SpeechBrain: A General-Purpose Speech Toolkit. Published online 2021. doi:10.48550/ARXIV.2106.04624
42. Owens FJ, Murphy MS. A short-time Fourier transform. Signal Processing. 1988;14(1):3-10. doi:10.1016/0165-1684(88)90040-0
43. Guo J, Zhu J, Lin S, Shi F. ECAPA-TDNN Embeddings for Speaker Recognition. In: 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). IEEE; 2024:1488-1491. doi:10.1109/AINIT61980.2024.10581514
44. Ghrabli S, Elgendi M, Menon C. Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments. Sci Rep. 2024;14(1):593. doi:10.1038/s41598-023-50371-2
45. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77. doi:10.1186/1471-2105-12-77
46. Melek Manshouri N. Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. Cogn Neurodyn. 2022;16(1):239-253. doi:10.1007/s11571-021-09695-w
47. Nayak SS, Darji AD, Shah PK. Machine learning approach for detecting Covid-19 from speech signal using Mel frequency magnitude coefficient. SIViP. 2023;17(6):3155-3162. doi:10.1007/s11760-023-02537-8
48. Akman A, Coppock H, Gaskell A, Tzirakis P, Jones L, Schuller BW. Evaluating the COVID-19 Identification ResNet (CIdeR) on the INTERSPEECH COVID-19 From Audio Challenges. Front Digit Health. 2022;4:789980. doi:10.3389/fdgth.2022.789980
49. Amin SU, Taj S, Hussain A, Seo S. An automated chest X-ray analysis for COVID-19, tuberculosis, and pneumonia employing ensemble learning approach. Biomedical Signal Processing and Control. 2024;87:105408. doi:10.1016/j.bspc.2023.105408
50. Kumar Acharya A, Satapathy R. A Deep Learning Based Approach towards the Automatic Diagnosis of Pneumonia from Chest Radio-Graphs. Biomed Pharmacol J. 2020;13(1):449-455. doi:10.13005/bpj/1905

51. Naji O, Darwish I, Bessame K, et al. A Comparison of the Epidemiological Characteristics Between Influenza and COVID-19 Patients: A Retrospective, Observational Cohort Study. Cureus. 2023;15(11):e49280. doi:10.7759/cureus.49280
52. Venkataramana L, Prasad DVV, Saraswathi S, Mithumary CM, Karthikeyan R, Monika N. Classification of COVID-19 from tuberculosis and pneumonia using deep learning techniques. Med Biol Eng Comput. 2022;60(9):2681-2691. doi:10.1007/s11517-022-02632-x
53. Goel S, Kipp A, Goel N, Kipp J. COVID-19 vs. Influenza: A Chest X-ray Comparison. Cureus. 2022;14(11):e31794. doi:10.7759/cureus.31794
54. Saleh Y. Synthesizing Cough Audio with GAN for COVID-19 Detection. Published online 2023. doi:10.48550/ARXIV.2305.04810
55. Tavaziva G, Harris M, Abidi SK, et al. Chest X-ray Analysis With Deep Learning-Based Software as a Triage Test for Pulmonary Tuberculosis: An Individual Patient Data Meta-Analysis of Diagnostic Accuracy. Clinical Infectious Diseases. 2022;74(8):1390-1400. doi:10.1093/cid/ciab639

6.11. Supplementary information

Table S6.1. Definition of extracted temporal and spectral acoustic features.

Feature	Description
Zero-crossing rate (ZCR)	The rate of sign-changes (positive to negative or negative to positive) within a frame. Higher frequencies have a high number of zero crossings while a low frequency should have a low number of zero crossings. ZCR is a temporal feature.
Energy	The energy of a frame is a measure of the audio signal's strength or magnitude. It captures the overall power distribution within the temporal structure of the audio signal. Energy is a temporal feature.
Energy entropy	A measure of the changes and distribution of energy in an audio signal over time. It helps distinguish between steady, consistent sounds and those with more variation or abrupt changes. Energy entropy is a temporal feature.
Intensity	Acoustic intensity is a physical quantity that represents the amount of sound energy flowing through a unit area per unit time. In the context of audio signal processing, intensity is often expressed in decibels (dB) relative to a reference level, typically the threshold of human hearing. Intensity is a temporal feature.
Spectral centroid	A measure that indicates where the "center of mass" of the spectrum is located. Higher values correspond to brighter sounds with more high-frequency content, while lower values indicate more bass-heavy sounds. Spectral centroid is a spectral feature.
Spectral spread	Measures how the frequency components of a sound are distributed around the spectral centroid. A larger spread indicates a more diverse range of frequencies, while a smaller spread suggests more focused frequency content. Spectral spread is a spectral feature.
Spectral entropy	Quantifies the complexity or unpredictability in the frequency content of a sound. Higher values indicate more random, noise-like sounds, while lower values suggest more organized, tonal sounds. Spectral entropy is a spectral feature.
Spectral flux	Measures the frame-to-frame change in the frequency spectrum of a sound. Higher values indicate rapid or significant changes in the spectral content, while lower values suggest more stable, consistent sounds. Spectral flux is a spectral feature.
Spectral rolloff	The frequency below which a specified percentage (usually 85-95%) of the total spectral energy is contained. This helps differentiate between harmonic content and noisy components in the signal. Spectral rolloff is a spectral feature.
Mel-Frequency Cepstral Coefficients (MFCCs)	A set of 13 coefficients that represent the short-term power spectrum of sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. These

	coefficients capture the general shape of the spectral envelope in a way that approximates human auditory perception.
MFCCΔ	The first-order temporal derivatives of the MFCC features, representing how MFCC values change over time. These features capture the dynamic aspects of speech and help in distinguishing between similar sounds with different temporal patterns.
Filter banks (FBanks)	A set of 40 frequency bands that divide the audio spectrum into regions that approximate how human hearing processes different frequencies. Each filter bank coefficient represents the energy present in its corresponding frequency band.

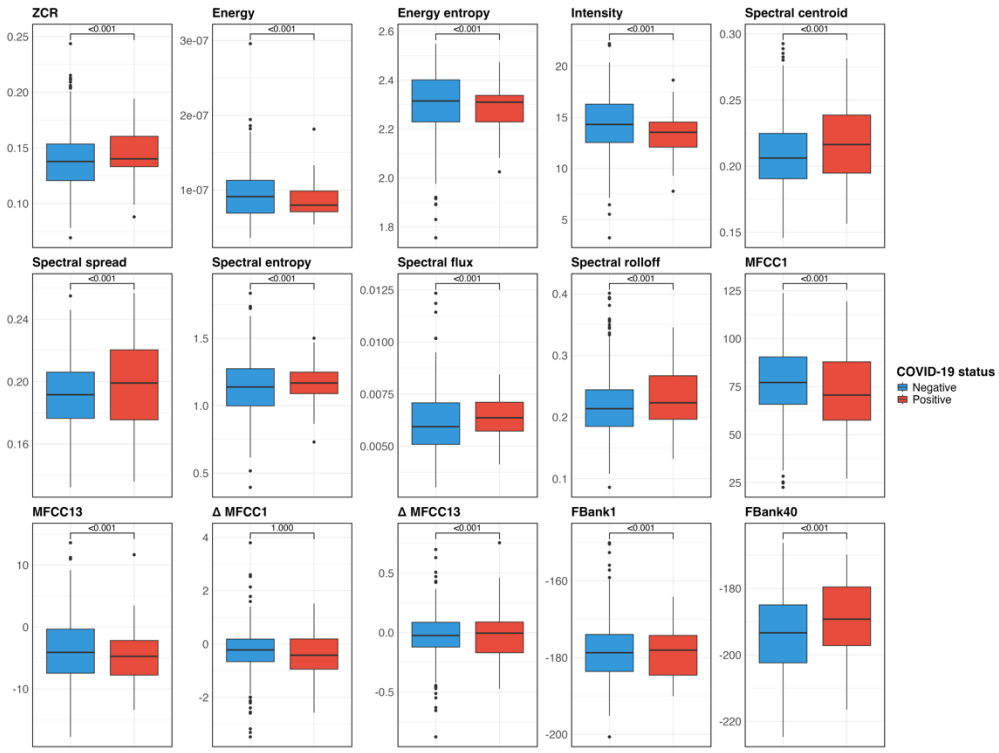


Figure S6.1. Comparison of mean acoustic cough feature distributions among participants in Lima, Peru who were COVID-19 negative (blue; N=270 participants) and COVID-19 positive (red; N=36 participants). Statistically significant differences (P<0.001) based on t-tests with Bonferroni correction are observed for all features except the ΔMFCC1. ZCR, zero-crossing rate; MFCC, Mel-frequency cepstral coefficients; FBank, filter-bank.

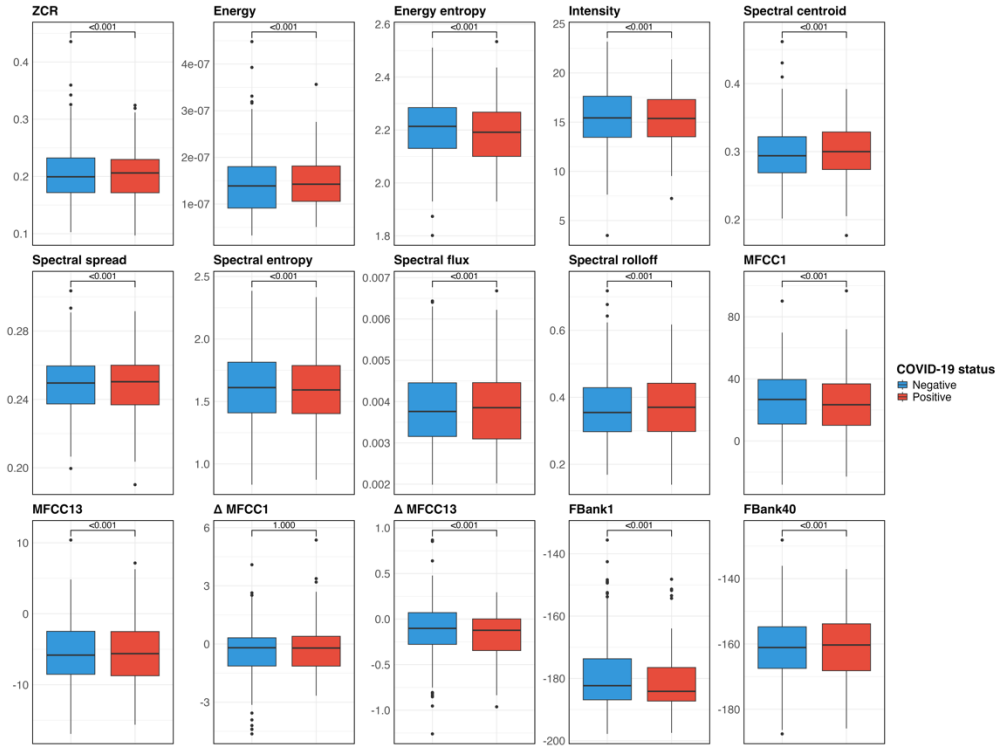


Figure S6.2. Comparison of mean acoustic cough feature distributions among participants in Montreal, Canada who were COVID-19 negative (blue; N=197 participants) and COVID-19 positive (red; N=102 participants). Statistically significant differences (P<0.001) based on t-tests with Bonferroni correction are observed for all features except the ΔMFCC1. ZCR, zero-crossing rate; MFCC, Mel-frequency cepstral coefficients; FBank, filter-bank.

Table S6.2. Table of model performance metrics for the XGBoost models.

			5-seed average (standard error)				
Configuration	**Test set**	**Features**	**AUC**	**Spec. at Sens.=0.70**	**Spec. at Sens.=0.90**	**Accuracy***	**F1***
Within-dataset	Lima	Audio	0.71 (0.08)	0.67 (0.13)	0.42 (0.11)	0.85 (0.08)	0.21 (0.20)
		Clinical	0.63 (0.07)	0.53 (0.17)	0.20 (0.18)	0.71 (0.07)	0.17 (0.06)
		Audio + Clinical	0.73 (0.06)	0.69 (0.12)	0.33 (0.17)	0.78 (0.02)	0.24 (0.07)
	Montreal	Audio	0.53 (0.04)	0.38 (0.12)	0.21 (0.11)	0.52 (0.01)	0.33 (0.03)
		Clinical	0.64 (0.06)	0.51 (0.09)	0.33 (0.11)	0.53 (0.09)	0.49 (0.04)
		Audio + Clinical	0.66 (0.05)	0.52 (0.14)	0.38 (0.11)	0.56 (0.06)	0.49 (0.09)

	Combined	Audio	0.67 (0.03)	0.59 (0.02)	0.25 (0.12)	0.64 (0.06)	0.40 (0.11)
		Clinical	0.71 (0.03)	0.60 (0.07)	0.32 (0.10)	0.64 (0.05)	0.46 (0.03)
		Audio + Clinical	0.72 (0.03)	0.65 (0.05)	0.37 (0.11)	0.65 (0.04)	0.49 (0.03)
External val.	Lima	Audio	0.53 (0.03)	0.31 (0.078)	0.15 (0.07)	0.86 (0.01)	0.04 (0.04)
		Clinical	0.53 (0.04)	0.35 (0.06)	0.12 (0.02)	0.88 (0.00)	0.00 (0.00)
		Audio + Clinical	0.50 (0.04)	0.47 (0.89)	0.25 (0.08)	0.88 (0.00)	0.00 (0.00)
	Montreal	Audio	0.51 (0.01)	0.32 (0.04)	0.10 (0.03)	0.65 (0.04)	0.11 (0.08)
		Clinical	0.52 (0.04)	0.32 (0.04)	0.11 (0.01)	0.66 (0.00)	0.00 (0.00)
		Audio + Clinical	0.51 (0.03)	0.27 (0.05)	0.10 (0.02)	0.46 (0.13)	0.35 (0.21)

*Used a threshold of 0.5, where observations >0.5 are classified as COVID-19 positive.
AUC, area under the curve; Sens., sensitivity; Spec., specificity; val., validation

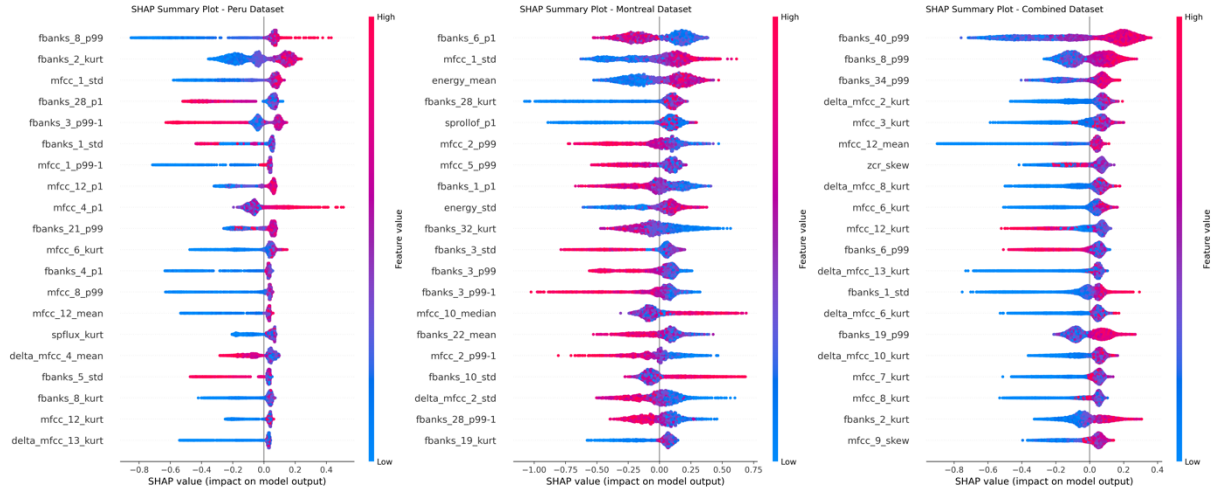


Figure S6.3. analysis of features used in the "Audio" analysis for Peru (left), Montreal (middle), and Combined (right). The features are ranked vertically by their impact on model predictions. The most common features were MFCCs and FBanks, though their relative ranking between analyses differed. Positive SHAP values (x-axis) indicate stronger prediction towards COVID-19 positive status. Color intensity represents feature values, with red indicating higher values (e.g., higher mean cough energy) and blue indicating lower values. Features showing mixed colors in their SHAP values indicate their effects on predictions are dependent on other feature values, revealing complex feature interactions in the model. MFCC, Mel-Frequency Cepstral Coefficients ; SHAP, SHapley Additive exPlanations

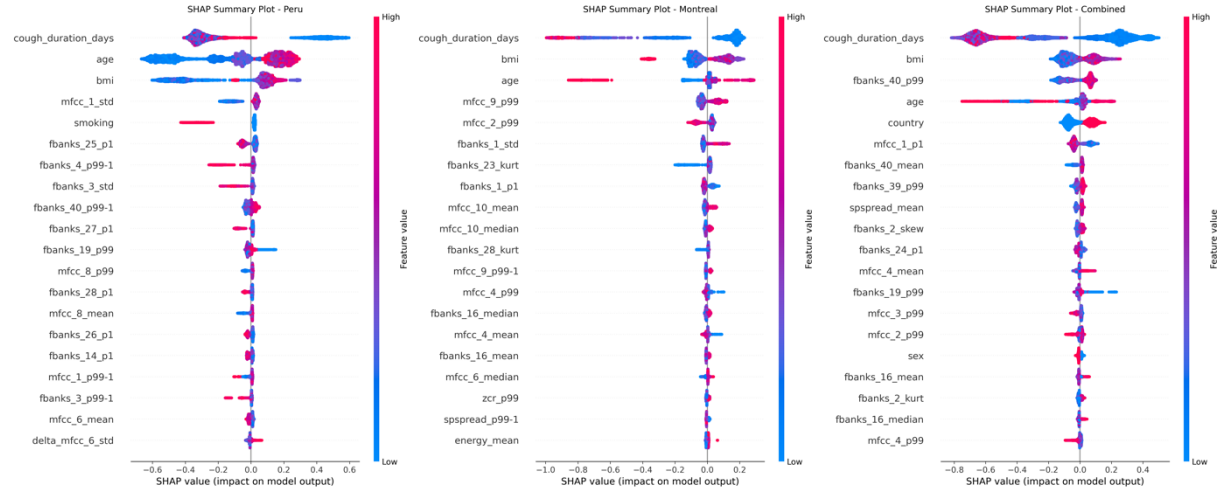


Figure S6.4. SHAP analysis of features used in the "Audio+Clinical" analysis for Peru (left), Montreal (middle), and Combined (right). The features are ranked vertically by their impact on model predictions, with cough duration, age, and BMI consistently emerging as the top contributors. Positive SHAP values (x-axis) indicate stronger prediction towards COVID-19 positive status. Color intensity represents feature values, with red indicating higher values (e.g., older age) and blue indicating lower values. Features showing mixed colors in their SHAP values indicate their effects on predictions are dependent on other feature values, revealing complex feature interactions in the model. Notably, shorter cough durations (blue) are associated with COVID-19 positive predictions. The wider spread of SHAP values for clinical variables (cough duration, age, BMI) indicates these features can more strongly influence predictions in either direction, while the narrower spread of audio features suggests their individual effects are more modest and might work better in combination. BMI, body mass index; SHAP, SHapley Additive exPlanations

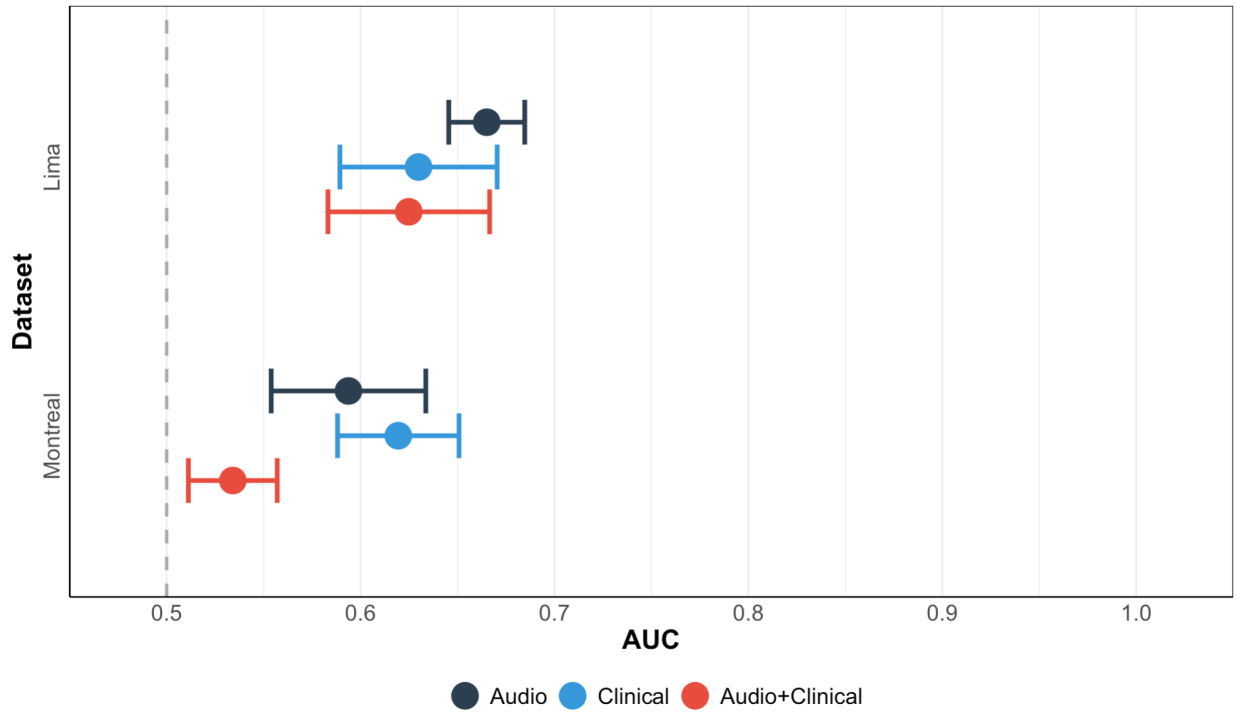


Figure S6.5. Conditional areas under the curve (AUCs) from the combined model stratified by setting (Lima, Peru and Montreal, Canada). The model shows varying performance between sites, with generally higher AUCs observed in Lima compared to Montreal for the "Audio" and "Audio+Clinical" feature set. Dots represent seed averages with error bars representing the standard error.

Table S6.3. Table of model performance metrics for the ECAPA-TDNN models.

			5-seed average (standard error)				
Configuration	**Test set**	**Features**	**AUC**	**Spec. at Sens.=0.70**	**Spec. at Sens.=0.90**	**Accuracy***	**F1***
Within-dataset	Lima	Audio	0.60 (0.07)	0.49 (0.30)	0.07 (0.05)	0.34 (0.13)	0.20 (0.10)
	Montreal	Audio	0.54 (0.02)	0.39 (0.12)	0.15 (0.07)	0.39 (0.10)	0.49 (0.04)
	Combined	Audio	0.64 (0.07)	0.52 (0.26)	0.19 (0.25)	0.32 (0.23)	0.35 (0.17)
External val.	Lima	Audio	0.52 (0.01)	0.35 (0.04)	0.14 (0.03)	0.28 (0.05)	0.32 (0.01)
	Montreal	Audio	0.52 (0.01)	0.33 (0.03)	0.13 (0.02)	0.44 (0.11)	0.45 (0.09)

*Used a threshold of 0.5, where observations >0.5 are classified as COVID-19 positive.
AUC, area under the curve; Sens., sensitivity; Spec., specificity; val., validation

Chapter 7. Manuscript V: External validation of cough-based algorithms for pulmonary tuberculosis triage from the CODA TB DREAM Challenge using cough data from Peru

7.1. Preface

The previous manuscript demonstrated that COVID-19 cough screening models trained on one population transferred poorly to other settings, highlighting the challenges of developing globally applicable AI tools for cough sound analysis. Building on these insights about the external validity of cough classification models, this manuscript examines similar questions in the context of TB triage models. We utilized models that were developed as part of the COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM Challenge using cough data from seven countries in Asia and Africa and externally validated them on a cough dataset collected in Peru. This work adds to the growing evidence that cough-based AI algorithms may be inherently population-specific, suggesting that future development efforts should focus on creating and validating models for local contexts rather than pursuing a one-size-fits-all approach.

7.2. Title page

External validation of cough-based algorithms for pulmonary tuberculosis triage from the CODA TB DREAM Challenge using cough data from Peru

Alexandra J. Zimmer[1,2], Patricia Espinoza-Lopez[3,4], Vijay Ravi[5], Solly K. Sieberts[6], Samira Abbasgholizadeh Rahimi[7,8,9], Madhukar Pai[1,2,10], César Ugarte-Gil[3,4,11], Simon Grandjean Lapierre[2,12,13]

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada

2 McGill International TB Centre, Montreal, Canada

3 School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

4 Instituto de Medicina Tropical Alexander von Humboldt, Universidad Peruana Cayetano Heredia, Lima, Peru

5 Department of Electrical and Computer Engineering, University of California Los Angeles, Los Angeles, CA, USA

6 Sage Bionetworks, Seattle, WA, USA

7 Mila – Quebec AI Institute, Montreal, Quebec, Canada

8 Family Medicine, Faculty of Medicine and Health Sciences and Faculty of Dental Medicine and Oral Health Sciences, McGill University, Montreal, Quebec, Canada

9 Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada

10 School of Population and Global Health, McGill University, Montreal, Canada

11 Department of Epidemiology, School of Public and Population Health, University of Texas Medical Branch, Galveston, USA

12 Immunopathology Axis, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, 900 Rue Saint-Denis, Montréal, Canada

13 Department of Microbiology, Infectious Diseases and Immunology, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Canada

7.3. Abstract

Introduction

The COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM Challenge recently evaluated the performance of artificial intelligence (AI) algorithms for tuberculosis (TB) triage using cough sounds. The CODA Challenge developed and internally validated 11 AI models using a dataset of 733,756 cough sounds collected from 2,143 adults from seven countries in Africa and Asia. This study externally evaluates the 11 AI models developed using the CODA Challenge with an external cough dataset from Peru.

Methods

Cough recordings from 303 coughing adults were collected from health facilities in Lima, Peru, using the Hyfe Research smartphone application. Eleven AI models from the CODA Challenge were evaluated using this independent dataset. Model performance was assessed using area under the curve (AUC) and compared to the original CODA Challenge results.

Results

The AUCs of the models ranged from 0.480 to 0.615, showing a decrease in performance compared to the model's performance when internally validated using the CODA Challenge, which ranged from 0.689 to 0.743. The best performing model in the CODA Challenge was also the best performing model in this external validation. Sub-group analyses revealed that models performed better in older (≥35 years) populations and among people with prior TB. No difference in model performance was observed between sex and whether the person presented with a fever.

Conclusion

The external validation revealed limitations in the generalizability of the CODA Challenge models to other settings. While some models showed promise, the overall performance decline highlights

the need for continued model validation on completely external datasets. It also underscores the importance of developing context-specific models as a viable alternative strategy. Such an approach could potentially yield higher accuracy in local settings, accounting for population-specific factors that influence cough characteristics and TB prevalence.

7.4. Introduction

Advances in artificial intelligence (AI) and digital health have sparked interest in the use of cough sounds as a biomarker for screening and monitoring respiratory diseases.[1,2] The emerging field of acoustic epidemiology applies methods in AI to detect signals in sound recordings, including cough, that may indicate underlying respiratory condition.

Pulmonary tuberculosis (PTB), a bacterial infectious disease, continues to pose significant global health challenges, with over 2.7 million people with TB remaining undetected or not notified in 2023.[3] To address this critical gap in detection, innovative diagnostic and screening tools are needed. Among these, AI-based cough screening tools have emerged as a promising solution for point-of-care triage.[4,5] These advanced technologies, accessible via smartphone applications, offer a novel approach that could improve TB screening by making it more readily available to both healthcare providers and patients.

Several AI models have been developed for PTB triage from cough sounds.[6–9] The most recent and largest initiative is the COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM Challenge (hereafter referred to as the CODA Challenge).[10,11] This public challenge, hosted on Synapse (www.synapse.org/tbcough), invited academics and companies with an interest in developing AI models for TB triage using cough sounds to participate. The CODA Challenge ran from October 2022 until February 2023. The top-performing model for cough-only (i.e., without using any other clinical data) challenge was the 'Blue Team' from Flywheel.io, achieving an area under the curve (AUC) of 0.743 (95% confidence interval [CI]: 0.703, 0.780).[11]

While these findings are promising, it is not clear how these models would perform on independent datasets from distinct populations, particularly from countries not involved in the models' original development. Factors such as the variability in human phenotypes, differences in recording devices and environments, and the local epidemiology of PTB and other respiratory diseases are expected to impact cough-based triage models' performance. To address this, the present study is an external evaluation of models submitted to the CODA Challenge that only used cough sounds for PTB classification (referred to as sub-challenge 1 in the CODA Challenge results manuscript).[11] In this

context, classification refers to the process of categorizing individuals as either likely to have PTB or not, based on their cough sounds. This evaluation employs cough data from a cohort of people with presumed PTB who were recruited in Lima, Peru, using the Hyfe Research application. By assessing these algorithms' performance on an independent and geographically distinct dataset, this study aims to provide insights into the transferability and real-world applicability of cough-based PTB screening tools.

7.5. Methods

7.5.1. Enrollment setting and timeline

Participants for the external validation dataset were prospectively enrolled from Hospital de Huaycán, a public secondary referral hospital in the Ate-Vitarte district, and a network of 33 primary health centers of the San Juan de Lurigancho district. Enrollment took place in two phases. Phase 1 ran from March 2022 to January 2023, where participants were recruited from the cohort of a parent study investigating the diagnostic accuracy of integrated TB and COVID-19 testing.[12] Phase 2, conducted from July 2023 to March 2024, was an independent cohort study that enrolled new participants to record their coughs at the same health facilities.

7.5.2. Inclusion and exclusion criteria

The study consecutively enrolled adults aged 18 years and older who presented with a new-onset cough. For Phase 1, the exclusion criteria from the parent diagnostic accuracy study were applied. This meant that individuals with a confirmed COVID-19 diagnosis within the previous three months or those who had taken anti-tuberculosis medication in the past six months were not eligible for enrollment. Additionally, for both phases of the study, participants were excluded from the sub-study if they were currently using cough-suppressive medications.

7.5.3. Microbiological reference standard

Sputum samples were tested for TB using Xpert MTB/RIF Ultra (Xpert Ultra) (Cepheid, Sunnyvale, CA, USA) and bacterial culture (BD BACTEC MGIT, BD, Franklin Lakes, NJ, USA) at the Humberto Guerra Alisson Laboratory, a reference laboratory at the Instituto de Medicina Tropical Alexander von Humboldt at the Universidad Peruana Cayetano Heredia. Xpert Ultra tests were repeated if initial testing produced an error, indeterminate, or a 'Trace' semi-quantitative result. MGIT culture was repeated if the initial test was contaminated. Participants were TB positive if either of the following conditions were met: a positive sputum Xpert Ultra result, a MGIT positive, or two Trace positive results on Xpert Ultra. Participants were considered PTB negative if none of the tests had a positive result and at least two sputum tests (Xpert Ultra, Xpert Ultra repeat, MGIT, or MGIT repeat) were negative. Participants had an indeterminate TB status if both sputum Xpert Ultra results were 'Trace' and culture was negative. Indeterminate TB status participants were not included in the analysis.

COVID-19 testing was also performed among all participants. During Phase 1, this was done using the Xpert Xpress SARS-CoV-2 cartridges (Cepheid, Sunnyvale, CA, USA). During Phase 2, this was done using Xpert Xpress SARS-CoV-2/Flu/RSV cartridges (Cepheid, Sunnyvale, CA, USA). Both cartridges run on the GeneXpert PCR platform.

7.5.4. Cough recording procedures

The Hyfe Research application (Hyfe) was used for all cough recording activities. Hyfe has an embedded AI algorithm that can accurately differentiate cough sounds from non-cough sounds. This model was previously reported to be 96% sensitive and 96% specific for cough detection (i.e., differentiating a cough sound due to any reason from ambient noise or non-cough sounds) using human-labeled sounds as a reference standard.[13] The algorithm assigns a score from 0 to 1, reflecting the probability that the sound recorded was a cough. The cough sounds and their probability score were stored on a secured server only accessible by the study team. Recordings with a probability ≥0.8 of being a cough sound were included for downstream analysis.

Hyfe only records and stores up to 0.5 seconds of sounds, preventing the recording of conversations or background acoustic environments. Cough recording was performed by a trained study nurse at the time of enrollment at the health facility. All nurses conducting the recordings wore appropriate PPE, including N95 masks, to ensure safety during the data collection process. The unique ID of each participant was entered into the app by the study nurse. Participants were instructed to hold the study smartphone (Xiaomi Redmi 9A) with the microphone approximately 30cm away from their mouth. The study nurse instructed the participant on how to use the application to record each cough sound. If the participant was not familiar with using a smartphone, the study nurse supported the patient by activating the recording of each cough sound. Then they purposively coughed 5-10 times (or as many times as was physically comfortable) holding the smartphone while the Hyfe application recorded. Some participants coughed more than 10 times due to coughing fits, and these coughs were included in the analysis (as was done for the original CODA Challenge). All recordings were performed in a private, ventilated tent or at the sample collection area to minimize background noise.

7.5.5. CODA TB DREAM Challenge

The data used as part of the CODA Challenge was collected from two multi-country TB diagnostic studies. The Rapid Research in Diagnostic Development TB Network (R2D2 TB Network) enrolled participants from outpatient facilities in India, the Philippines, South Africa, Uganda, and Vietnam. The Digital Cough Monitoring Project enrolled participants from outpatient facilities in Madagascar and Tanzania. In both studies, participants were eligible if they were 18 years or older and had a new or worsening cough in the past two weeks.

The cough data for the CODA Challenge was collected using the Hyfe application. Overall, 733,756 cough sounds were recorded from 2,143 participants across all countries. The cough data included 18,834 “solicited” coughs from all participants, which were purposefully collected at the time of enrollment by a healthcare worker (the same process as described in the Peru data collection). The remaining 714,922 coughs were “longitudinal” coughs, meaning that a subset of participants carried the smartphone with the Hyfe app continuously for two weeks to passively record any cough sounds. Approximately 25% (553/2,143) of the participants recruited were TB

positive based on a microbiological reference standard. Data was split at the participant level into training (n=1,105) and internal validation (n=1,308), meaning that all cough sounds belonging to the same individual were either in the training and internal validation set. Longitudinal cough sounds were only used for model training, not validation. Additional information on the CODA Challenge dataset has been published.[10]

The CODA Challenge put forth two sub-challenges for TB classification: the first used cough sounds alone and the second used cough sounds and basic demographic and clinical variables.[11] In this study, we exclusively validated the first sub-study as not all clinical and demographic variables used in the second CODA Challenge were available in the Peru dataset. Docker images from the 11 teams that publicly participated in the challenge were used to run all pre-processing steps and models on the Peru cough dataset. Docker is a platform that allows developers to package applications with all their dependencies into standardized units called containers, ensuring consistent performance across different computing environments. These files were saved in an Open Neural Network Exchange format. The model architectures and sound features used from the 11 teams have been summarized in **Table S7.1** Docker images were run by the data collection team and not the model developers.

7.5.6. Model evaluation

The same model ranking approach used by the CODA Challenge was applied. All models produced an output prediction score reflecting the probability of an individual participant having TB based on their cough sounds. The AUC was calculated for each model using the pROC package in R Software version 4.3.3.[14] The partial AUC (pAUC) that was presented in the CODA Challenge results was not calculated as the performance of the models were too low to achieve the set targets of sensitivity of 80% and specificity of 60%, which are both 10% less than the minimum World Health Organization (WHO) target product profile (TPP) for a TB triage tool which aims for 90% sensitivity and 70% specificity.[15] Additional analyses evaluated the models' specificity when the sensitivity was set to 80% and to 90% and the results were compared with the CODA Challenge evaluation set. Confidence intervals for the AUC, sensitivity, and specificity were calculated using bootstrap resampling (n=1000). Sub-analyses examined the distribution of AUCs across levels of

demographic and clinical predictors: age (<35 years vs. ≥35 years), sex, prior TB, and body mass index (BMI) (<25 kg/m^2 years vs. ≥25 kg/m^2), and self-reported fever (present vs. not). Fever was included as a non-specific symptom of disease.

7.5.7. Ethics

All participants provided written informed consent. The studies involved in the recruitment phases received ethical approval from the Comité Institucional de Ética en Investigación at Universidad Peruana Cayetano Heredia (Phase 1: SIDISI 202931/SIDISI 206951, and Phase 2: SIDISI 210857) and the McGill University Health Centre Research Ethics Board (2021-6866). The studies were registered in the PRISA repository at Instituto Nacional de Salud in Peru (Phase 1: EI00000001484/EI00000002576, and Phase 2: EI00000003221).

7.6. Results

7.6.1. Participant information

A total of 310 adults were enrolled in Lima. Seven individuals (2.3%) had an indeterminate TB status and were excluded from the analysis. Of the 303 included participants, 97 (32.0%) were TB positive and 206 (68.0%) were TB negative according to the microbiological reference standard (**Table 7.1**). The overall median age was 36 years (interquartile range [IQR]: 24, 52) and 135 (44.6%) were female. Approximately half of all individuals (53.1%) reported a fever at time of enrollment. Among all participants, 36 (12.0%) tested positive for COVID-19, with 5 of these cases being TB co-infection. Forty-two (13.9%) individuals reported a history of prior TB, of whom the majority (40/42) had prior pulmonary TB and 2/42 had extrapulmonary TB.

Table 7.1. Demographic and clinical characteristics stratified by TB diagnosis.

	TB positive n=97	TB negative n=206	Total n=303
Age			
<35 years	57 (58.8)	84 (40.8)	141 (46.5)
≥35 years	40 (41.2)	122 (59.2)	162 (53.5)
Sex			
Female	31 (32.0)	104 (50.5)	135 (44.6)
Male	66 (68.0)	102 (49.5)	168 (55.4)
Smoking status			
Current	3 (3.1)	7 (3.4)	10 (3.3)
Former	12 (12.4)	11 (5.3)	23 (7.6)
Never	82 (84.5)	188 (91.3)	270 (89.1)
BMI			
<25	78 (80.4)	103 (50.0)	181 (59.7)
≥25	19 (19.6)	103 (50.0)	122 (40.3)
Cough duration			
<20 days	15 (15.5)	120 (58.3)	135 (44.6)
≥25 days	82 (84.5)	86 (41.8)	168 (55.4)
Symptoms			
Headache	70 (72.2)	167 (81.1)	237 (78.2)
Fever	57 (58.8)	104 (50.5)	161 (53.1)
Sore throat	68 (70.1)	166 (80.6)	234 (77.2)
Loss of smell/taste	4 (4.1)	20 (9.7)	24 (7.9)
Muscle pain	64 (66.0)	152 (73.8)	216 (71.3)
Fatigue	41 (42.3)	103 (50.0)	144 (47.5)
Nausea	34 (35.1)	52 (25.2)	86 (28.4)
Diarrhea	21 (21.7)	46 (22.3)	67 (22.1)
COVID-19	5 (5.3)	31 (15.1)	36 (12.0)
Prior TB*	8 (8.3)	34 (16.5)	42 (13.7)

*Includes pulmonary and extrapulmonary TB.
All datra are presented as N (%).
BMI, body mass index; TB, tuberculosis

7.6.2. Model performance

The AUCs from the CODA Challenge internal validation ranged from 0.645 to 0.743 (**Figure 7.1A**). For the external validation on the Peru dataset, the AUCs ranged from 0.480 to 0.615 (**Figure 7.1B, Table 7.2**). The performance of all models decreased compared to the reported performance in the CODA Challenge. The models with the highest and lowest performance in the Peru validation were ranked similarly in the CODA Challenge. However, the order of the other models was not the same. The highest performing model in CODA Challenge was by the 'Blue

Team', with an AUC of 0.743 (0.703, 0.780). In the external Peru validation, the 'Blue Team's' model performed best, however the AUC declined to 0.615 (0.550, 0.680).

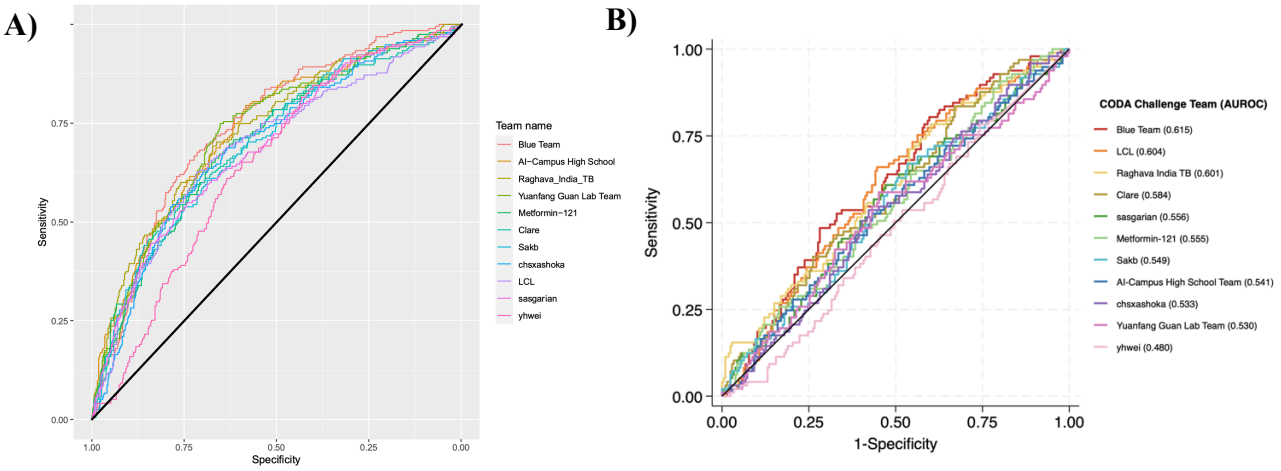


Figure 7.1. Comparison of 11 CODA Challenge algorithm performance in A) the internal validation of CODA Challenge models, using cough data from the CODA Challenge and B) the external validation, using cough data from Peru. Team names are ranked in order of their performance in the respective evaluations. The Blue Team model performed best in both the internal and external evaluation. Overall, performance was decreased for all models in the external evaluation. (Figure A is adapted from Jaganath et al.)[11]

For all models, there was a decrease in the specificity when setting the sensitivity to 80% and 90% (**Table 7.2**). Importantly, none of the models achieved the target WHO TPP for a triage test of 70% specificity at 90% sensitivity in the original CODA Challenge dataset. Thus, it was not expected to do the same in the external validation. The highest observed specificity when the sensitivity was set at 90% was 0.243 (95% CI: 0.102, 0.364) for the Blue Team, which decreased from a specificity of 0.350 (95% CI: 0.254, 0.487) in the original CODA Challenge dataset.

Table 7.2. Model performance for cough-based TB classification using the external Peru test set compared to the CODA challenge test set.

	Peru test dataset			CODA Challenge test dataset[11]		
Team	**AUC (95% CI)**	**Sensitivity = 0.800: Specificity (95% CI)**	**Sensitivity = 0.900: Specificity (95% CI)**	**AUC (95% CI)**	**Sensitivity = 0.800: Specificity (95% CI)**	**Sensitivity = 0.900: Specificity (95% CI)**
Blue Team	0.615 (0.550, 0.680)	0.393 (0.267, 0.476)	0.243 (0.102, 0.364)	0.743 (0.703, 0.780)	0.555 (0.466, 0.640)	0.350 (0.254, 0.487)
LCL	0.604 (0.540, 0.671)	0.364 (0.248, 0.485)	0.185 (0.087, 0.350)	0.689 (0.644, 0.733)	0.412 (0.287, 0.548)	0.193 (0.126, 0.299)
Raghava India TB	0.601 (0.533, 0.665)	0.350 (0.228, 0.466)	0.175 (0.087, 0.340)	0.730 (0.690, 0.773)	0.504 (0.407, 0.612)	0.343 (0.208, 0.422)
Clare	0.584 (0.517, 0.648)	0.335 (0.228, 0.413)	0.214 (0.141, 0.320)	0.699 (0.655, 0.746)	0.469 (0.376, 0.561)	0.235 (0.141, 0.398)
sasgarian	0.556 (0.487, 0.624)	0.228 (0.136, 0.389)	0.141 (0.034, 0.214)	0.689 (0.647, 0.732)	0.429 (0.356, 0.509)	0.313 (0.198, 0.370)
Metformin-121	0.555 (0.483, 0.622)	0.272 (0.194, 0.354)	0.189 (0.083, 0.272)	0.704 (0.660, 0.746)	0.476 (0.387, 0.558)	0.281 (0.180, 0.390)
Sakb	0.549 (0.477, 0.619)	0.233 (0.146, 0.398)	0.141 (0.029, 0.228)	0.695 (0.654, 0.739)	0.474 (0.363, 0.526)	0.323 (0.222, 0.383)
AI-Campus High School Team	0.541 (0.471, 0.610)	0.238 (0.141, 0.354)	0.126 (0.049, 0.218)	0.731 (0.691, 0.771)	0.571 (0.472, 0.629)	0.299 (0.207, 0.481)
chsxashoka	0.533 (0.467, 0.600)	0.243 (0.165, 0.364)	0.136 (0.073, 0.238)	0.693 (0.651, 0.736)	0.437 (0.336, 0.533)	0.277 (0.192, 0.345)
Yuanfang Guan Lab Team	0.530 (0.463, 0.606)	0.185 (0.087, 0.354)	0.073 (0.029, 0.160)	0.727 (0.685, 0.768)	0.560 (0.427, 0.661)	0.289 (0.224, 0.415)
yhwei	0.480 (0.412, 0.551)	0.217 (0.130, 0.320)	0.105 (0.065, 0.184)	0.645 (0.601, 0.687)	0.435 (0.368, 0.496)	0.299 (0.210, 0.382)

AUC, area under the curve; CI, confidence interval; CODA, COugh Diagnostic Algorithm for Tuberculosis

7.6.3. Sub-group analyses

The median model performance did not differ by sex or the presence of fever (**Figure 7.2**). However, there does appear to be greater variation in how models performed when patients present with fever. The models performed better among older participants (AUC: 0.579; IQR: 0.564, 0.614 for age ≥35) compared to their younger counterparts (AUC: 0.529; IQR: 0.494, 0.577 for age <35). The most pronounced difference was observed when comparing individuals with and without a history of TB disease, where models achieved substantially higher accuracy in those with prior TB (AUC: 0.702; IQR: 0.640, 0.713) versus those without (AUC: 0.554; IQR: 0.534, 0.601). A smaller but notable difference was also seen with BMI, where lower BMI scores were associated with slightly better model performance (AUC: 0.547; IQR: 0.534, 0.580 vs AUC: 0.518; IQR: 0.498, 0.586 for higher BMI).

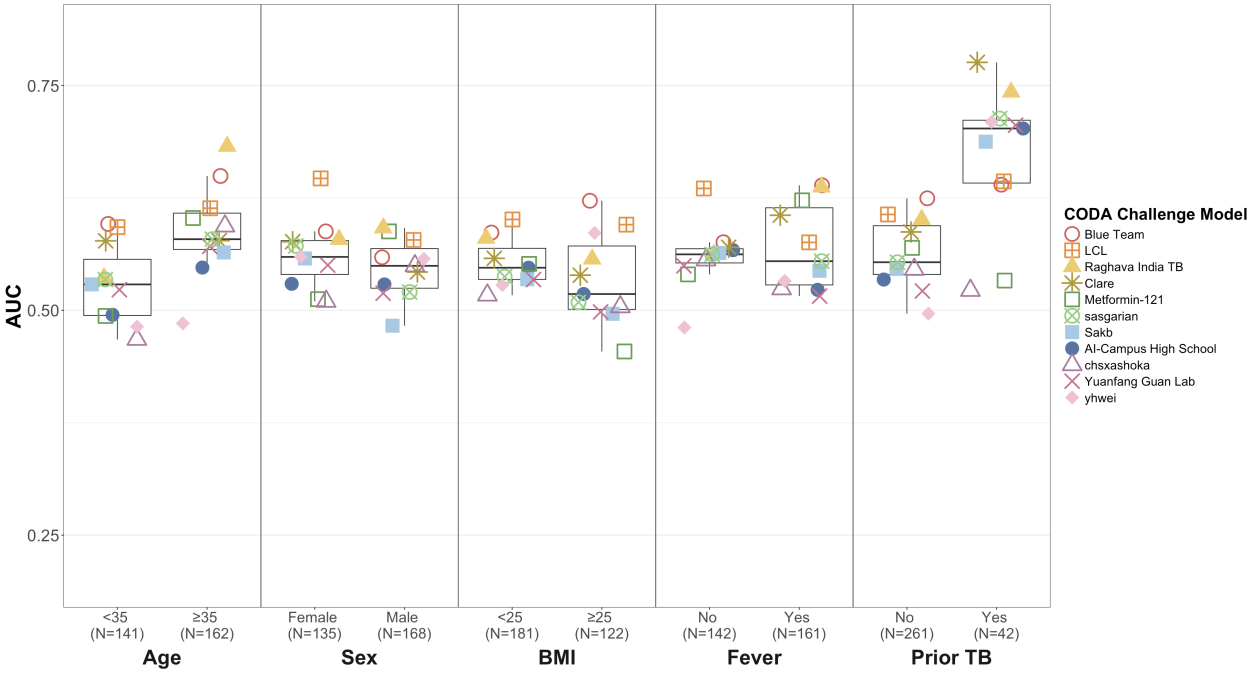


Figure 7.2. Area under the curve distribution of 11 CODA Challenge cough-only models in an external validation using data from Peru, stratified by demographic and clinical factors. Boxplots indicate the median and interquartile distribution of all models. Model-specific performance is indicated by markers. AUC, area under the curve; BMI, body mass index; TB, tuberculosis

7.7. Discussion

This external validation of the CODA Challenge models, which exclusively used cough sounds for TB classification, revealed a consistent decrease in performance compared to their original evaluations. This decline was anticipated, given that the original models were not trained using data from Peru. While there were some variations in model rankings between the CODA Challenge and this external evaluation, the best and worst performing models maintained their relative positions. Additionally, despite the overall performance drop, some models still achieved AUCs exceeding 0.600 in the external validation, with lower bound confidence intervals surpassing the 0.500 threshold.

It is important to note that this validation study utilized the CODA Challenge models as they were originally configured, without any further optimization. The hyperparameters remained unchanged from those tuned during the initial model training for the CODA Challenge. This approach was necessitated by the lack of access to the underlying model structures within the provided Docker files. Consequently, the potential for improving model performance through additional tuning remains unexplored. It is plausible that all 11 models could benefit from further optimization tailored to the specific characteristics of the Peru dataset. That said, this limitation mirrors real-world deployment scenarios, where pre-trained models are often applied to new data without modification. In practice, end users rarely have the expertise, resources, or access to retrain or fine-tune models for their specific use cases.

Sub-analyses revealed that the models' performance is not uniform across demographic and clinical subgroups. Notably, cough classification demonstrated superior performance among older individuals (≥35 years) and those with a history of prior TB compared to their respective counterparts. These differences may be attributed to physiological variations that modulate cough sounds, such as post-TB lung disease potentially altering cough characteristics.[16] However, it may also be attributed to characteristics of the dataset used to train the models. The training data from the CODA Challenge had a higher median age (40 years; IQR: 28, 53).[10] This composition could explain the better performance among older individuals and those with a TB history. Additionally, class imbalance within subgroups may have influenced model performance, which is particularly

evident among people with a history of prior TB, where only 8 out of 42 individuals had a positive TB diagnosis. Understanding the cause and patterns of heterogeneous performance across different subgroups is crucial for identifying potential biases and ensuring equitable model performance among different demographics. In the field of AI, this is often referred to as "aggregation bias" where assumptions about sub-groups are made based on observations made at the population level.[17] Further analyses on model performance in specific patient populations is needed to systematically map performance variations, understand their underlying causes, and develop strategies to mitigate potential disparities in model accuracy across different demographic and clinical subgroups.

A strength of this study is the utilization of the same Hyfe cough recording application employed by the CODA Challenge, ensuring comparable input data format and length, and thus enhancing the validity of comparisons with the CODA Challenge results. Additionally, the microbiological reference standard used for determining TB status resembled that of the CODA Challenge, further strengthening the comparative analysis.

There were several limitations with this external validation. The analysis was restricted to the sub-challenge involving only cough sounds due to the absence of certain clinical and demographic variables in the Peru dataset that were utilized in the second CODA sub-challenge, precluding the evaluation of those models. The CODA Challenge reported significant performance improvements when these variables were included, with some models approaching the WHO TPP sensitivity and specificity targets for a triage test. Future analyses should evaluate potential performance enhancements when combining the Peru cough data with available clinical and demographic variables. Despite using the same cough recording application, differences in smartphone hardware may have affected audio quality, potentially impacting model performance. However, the CODA Challenge stipulated that algorithms should be developed to ensure compatibility across different smartphones.[11] Further, cough data used in the CODA Challenge itself were recorded on different phones in the different countries.[10] Finally, while the microbiological reference standards were similar, some differences existed, notably the inclusion of urine samples for GeneXpert testing and solid Löwenstein–Jensen culture in the CODA Challenge dataset, which were not used in Peru.

Nevertheless, the risk of misclassification is minimal given the comprehensive array of other microbiological tests performed in Peru.

In conclusion, this external validation using cough data from a geographically distinct population reveals limitations in the generalizability of the CODA Challenge cough classification models. These findings underscore the critical importance of collecting cough data from diverse populations to develop robust, globally applicable models. Alternatively, efforts could be redirected towards developing and refining country-specific models, which may yield higher accuracy in local contexts.

7.8. Acknowledgements

We extend our gratitude to all patients who contributed data to this research, both for the original CODA Challenge dataset from the R2D2 TB Network and the Digital Cough Monitoring Project, as well as for the external validation study in Peru. Our appreciation also goes to the dedicated study personnel in all locations who made this work possible. We would like to recognize the efforts of all participants who developed models as part of the CODA TB DREAM Challenge, whose innovative work has advanced the field of cough-based AI TB triage.

7.9. Contributions

Conceptualization (A.J.Z., S.G.L.), Peru data collection (A.J.Z., P.E.L., C.U.G.), CODA Challenge Docker image generation (S.K.S.), Peru data preparation and cleaning (A.J.Z.), external validation data analysis (A.J.Z.), data visualization (A.J.Z.), writing – original draft (A.J.Z.), writing – review and editing (P.E.L., V.R., S.K.S., S.A.R., M.P., C.U.G., S.G.L.).

7.10. References

1. Alqudaihi KS, Aslam N, Khan IU, et al. Cough Sound Detection and Diagnosis Using Artificial Intelligence Techniques: Challenges and Opportunities. IEEE Access. 2021;9:102327-102344. doi:10.1109/ACCESS.2021.3097559

2. Kapetanidis P, Kalioras F, Tsakonas C, et al. Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review. Sensors. 2024;24(4):1173. doi:10.3390/s24041173
3. World Health Organization. Global Tuberculosis Report 2024.; 2024. Accessed November 18, 2024. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024
4. Zimmer AJ, Ugarte-Gil C, Pathri R, et al. Making cough count in tuberculosis care. Commun Med. 2022;2(1):83. doi:10.1038/s43856-022-00149-w
5. Nathavitharana RR, Garcia-Basteiro AL, Ruhwald M, Cobelens F, Theron G. Reimagining the status quo: How close are we to rapid sputum-free tuberculosis diagnostics for all? eBioMedicine. 2022;78:103939. doi:10.1016/j.ebiom.2022.103939
6. Botha GHR, Theron G, Warren RM, et al. Detection of tuberculosis by automatic cough sound analysis. Physiol Meas. 2018;39(4):045005. doi:10.1088/1361-6579/aab6d0
7. Pahar M, Klopper M, Reeve B, Warren R, Theron G, Niesler T. Automatic cough classification for tuberculosis screening in a real-world environment. Physiol Meas. 2021;42(10):105014. doi:10.1088/1361-6579/ac2fb8
8. Yellapu GD, Rudraraju G, Sripada NR, et al. Development and clinical validation of Swaasa AI platform for screening and prioritization of pulmonary TB. Sci Rep. 2023;13(1):4740. doi:10.1038/s41598-023-31772-9
9. Sharma M, Nduba V, Njagi LN, et al. TBscreen: A passive cough classifier for tuberculosis screening with a controlled dataset. Sci Adv. 2024;10(1):eadi0282. doi:10.1126/sciadv.adi0282
10. Huddart S, Yadav V, Sieberts SK, et al. A dataset of Solicited Cough Sound for Tuberculosis Triage Testing. Sci Data. 2024;11(1):1149. doi:10.1038/s41597-024-03972-z
11. Jaganath D, Sieberts SK, Raberahona M, et al. Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM Challenge. Published online May 14, 2024. doi:10.1101/2024.05.13.24306584
12. MacLean ELH, Villa-Castillo L, Espinoza-Lopez P, et al. Integrating tuberculosis and COVID-19 molecular testing in Lima, Peru: a cross-sectional, diagnostic accuracy study. The Lancet Microbe. 2023;4(6):e452-e460. doi:10.1016/S2666-5247(23)00042-3
13. Gabaldon-Figueira JC, Brew J, Doré DH, et al. Digital acoustic surveillance for early detection of respiratory disease outbreaks in Spain: a protocol for an observational study. BMJ Open. 2021;11(7):e051278. doi:10.1136/bmjopen-2021-051278
14. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77. doi:10.1186/1471-2105-12-77
15. World Health Organization. High-Priority Target Product Profiles for New Tuberculosis Diagnostics.; 2014. https://www.who.int/publications/i/item/WHO-HTM-TB-2014.18
16. Nihues SDSE, Mancuzo EV, Sulmonetti N, et al. Chronic symptoms and pulmonary dysfunction in post-tuberculosis Brazilian patients. The Brazilian Journal of Infectious Diseases. 2015;19(5):492-497. doi:10.1016/j.bjid.2015.06.005
17. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning. ACM Comput Surv. 2022;54(6):1-35. doi:10.1145/3457607

7.11. Supplementary information

Table S7.1. Overview of model architecture and audio features used for the 11 models from the cough-only CODA Challenge.

Team	Model type	Audio features used
Blue Team	Convolutional neural network	Spectrogram
LCL	Convolutional neural network, Light gradient-boosting machine	Zero-crossing rate, Mel frequency cepstral coefficients, chromagram, mel spectrogram, root mean square
Raghava India TB	Convolutional neural network	Mel spectrogram
Clare	Artificial Neural Network	Top 300 features extracted via OpenSMILE identified via principal component analysis
sasgarian	Artificial Neural Network	Top 1,024 features extracted via OpenSMILE identified via principal component analysis
Metformin-121	MetforNet	Z-score normalization of cough recordings
Sakb	Artificial Neural Network	Top 1,024 features extracted via OpenSMILE identified via principal component analysis
AI-Campus High School Team	Gradient Boosting Decision Tree	Mel-Frequency Cepstral Coefficients (MFCC), chromagram
chsxashoka	Artificial Neural Network	Mel Frequency Cepstral Coefficients, Mel spectrogram
Yuanfang Guan Lab Team	Light gradientboosting machine	Mel Frequency Cepstral Coefficients, first and second order time derivatives of MFCC, magnitude of pitch tracking, total number of coughs recorded
yhwei	Convolutional Neural Network	Spectrogram

Adapted from Jaganath et al.[1]

7.12. Supplementary references

1. Jaganath D, Sieberts SK, Raberahona M, et al. Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM Challenge. Published online May 14, 2024. doi:10.1101/2024.05.13.24306584

Chapter 8. Summary and conclusions

8.1. Summary of results

The five manuscripts presented in this thesis provide a comprehensive overview of advances made in acoustic epidemiology and independently contribute novel research findings to the field. Across these manuscripts, I examined two distinct applications of digital cough monitoring—cough counting for longitudinal disease monitoring and cough sound analysis for disease screening—with a particular focus on TB and COVID-19.

Manuscript I was a commentary on the use of cough as a biomarker for TB across the care cascade. Through review of the literature, we found that cough is a poorly utilized clinical symptom in TB care that could be transformed into an objective biomarker through emerging digital tools and technologies. While early studies demonstrate promise, there remains a critical need to develop well-annotated cough datasets and rigorously validate the clinical utility of these tools across.

In our scoping review (**Manuscript II**), we identified four clinical use cases for digital cough counting tools: disease diagnosis and severity assessment, treatment monitoring, health outcome prediction, and syndromic surveillance. Several factors impact the clinical application of tools, including privacy concerns regarding recordings and ergonomics of the cough monitors. Objective cough counts must be considered as part of a patient's broader clinical profile that involves other symptoms, biomarkers, and PROs.

In **Manuscript III**, we investigated the relationship between initial TB bacterial burden and cough frequency during the first two weeks of TB treatment. We assessed bacterial burden using two distinct markers: Xpert semi-quantitative results and CAD4TB scores. Our analysis revealed a clear dose-response relationship between bacterial burden and cough frequency. Compared to patients with 'High' Xpert results, those with lower semi-quantitative results demonstrated progressively decreasing cough frequencies: 'Medium' (RR: 0.79; 95% CI: 0.59, 1.05), 'Low' (RR: 0.64; 95% CI: 0.47, 0.87), and 'Very Low' (RR: 0.61; 95% CI: 0.41, 0.91). This association was

further corroborated by the CAD4TB model, where lower scores were significantly associated with reduced cough frequency (RR: 0.80; 95% CI: 0.67, 0.95). The robustness of these findings was confirmed through sensitivity analyses that accounted for potential cough count overestimation from neighboring participants, which strengthened the observed associations.

Manuscript IV and **Manuscript V** investigated the external validity of AI-based cough classification models across different populations, focusing on COVID-19 and TB detection, respectively. In **Manuscript IV**, our COVID-19 classifiers demonstrated strong internal validation performance in Lima. However, when these models were cross-validated on each other's datasets, their performance deteriorated significantly, with AUCs falling below 0.5 in both cases. **Manuscript V** revealed comparable challenges in model generalizability for TB detection. When we externally validated 11 audio-only TB classifiers from the CODA TB DREAM Challenge using our Peruvian dataset, we observed a substantial decline in performance, with AUCs dropping from 0.48 to 0.62 in the original challenge to 0.69 to 0.74 in our external validation.

8.2. Significance of the work

The digital health and AI landscape is experiencing unprecedented growth, with technological innovations transforming various sectors of society. However, healthcare has notably lagged in adopting these digital solutions,[141] largely due to justified concerns about reliability and the critical nature of medical decision-making. The successful implementation of digital health tools ultimately depends on their ability to meet end-user needs while maintaining clinical validity across diverse populations. The situation is especially concerning in lower-resource settings and the Global South, where a growing digital divide has resulted in limited validation and application of these technologies, despite their potential benefit in these regions.[142]

Despite the proliferation of AI algorithms for cough classification, there remains a significant gap between algorithm development and practical clinical implementation. This challenge is particularly evident in the field of digital cough counting technologies, where even long-standing tools like cough counters lack comprehensive validation across different populations and disease groups.

Our work began by examining digital cough tools through the lens of TB care, highlighting the untapped potential of cough-based technologies to improve TB outcomes. This perspective served to spotlight for the broader research community how digital innovations in cough monitoring and analysis could impact an underserved disease such as TB, particularly in resource-limited settings where such technologies are most needed yet least studied. Following this, our scoping review of cough counting tools revealed a crucial gap between technology development and clinical validation. Despite the long-standing availability of these tools, we found limited evidence of their validation across different settings and populations, highlighting the pressing need to translate promising technologies into validated tools that can meaningfully impact patient care.

To advance the field empirically, we investigated cough counts as a biomarker for TB treatment monitoring. Through analysis of prospectively collected data from high-burden TB countries, we identified established a relationship between cough counts and baseline TB bacterial burden. This work not only advanced our understanding of cough as a clinical indicator but also highlighted important methodological considerations for analyzing cough data, contributing to more robust research practices in this field.

Finally, our work challenged a fundamental assumption that underscored the field of cough-based AI classification: the universality of cough sounds across populations. Through external validation studies of both COVID-19 and TB classification models, we demonstrated that these algorithms may not generalize well across populations with different acoustic, demographic, and epidemiological profiles. This finding has important implications for future research and development in this field, suggesting a need for more localized approaches to model development and validation.

8.3. Strength and limitations

This thesis demonstrates notable strengths in both its scope and methodological diversity. The research spans multiple dimensions of cough as a biomarker, investigating both quantitative aspects (cough counts) and acoustic characteristics (cough sounds). The methodological approach

was comprehensive, encompassing literature reviews, primary data collection across geographically and demographically distinct locations, advanced statistical analyses of complex longitudinal data, and the application of machine learning and AI techniques to novel data types. This methodological breadth was supported by a truly interdisciplinary approach, bringing together expertise from epidemiology, biostatistics, machine learning, clinical medicine, and global health, resulting in a holistic examination of cough as a digital biomarker. Overall, this work underscores the importance of more cross-sectional research between public health and AI/ML sectors to ensure that technological developments are aligned with public health needs and can be effectively implemented in real-world settings.

Despite its broad methodological scope, this thesis has several important limitations. First, while the research explored multiple aspects of cough analysis, it focused primarily on two diseases: TB and COVID-19. This narrow disease focus, while allowing for depth of investigation, means that findings may not generalize to other respiratory conditions where cough is a prominent symptom.

A significant methodological challenge emerged in the longitudinal analysis of TB-related cough counts. The issue of measurement error in individual health data collection, particularly in the context of cough recordings, presented unique challenges. While missing data is not a unique issue, cough recording studies face the additional challenge of environmental contamination. Specifically, the inadvertent recording of coughs from individuals other than the target participant. While our study made important methodological advances by explicitly addressing this challenge through carefully designed sensitivity analyses, this remains an ongoing challenge in the field that will require continued technological innovation and methodological development.

The primary data collection efforts in Lima and Montreal for the fourth and fifth manuscripts also faced important limitations. The timing of data collection, which began in spring 2022, missed the major COVID-19 waves of 2020 and 2021, resulting in lower-than-anticipated COVID-19 prevalence rates among study participants. This timing issue, which diverged from the original expectations outlined in the 2021 grant proposal, resulted in a lower number of COVID-19 positive cases than anticipated, leading to more imbalanced comparison groups in our analyses. Additionally, the data collection in Peru was complicated by the need to conduct research over two

separate time periods. The absence of genetic screening tools further limited our ability to account for potential variations in COVID-19 strains, which may have influenced our findings given the evolving nature of the virus.

A notable limitation of this thesis is the absence of qualitative research exploring end-user experiences with digital cough tools. The successful implementation of any digital health solution fundamentally depends on its acceptability to patients and feasibility within local healthcare systems, aspects that are often overlooked in both digital health and diagnostic research more broadly.[143] While we did conduct interviews with patients and healthcare providers about their experiences recording cough sounds at health facilities in Lima (forthcoming publication to supplement my thesis), the analytical focus of this thesis did not capture these crucial implementation considerations.

8.4. Direction for future work

The transition from technological innovation to clinical implementation remains a critical challenge in the field of digital cough tools. Despite technological advances, there is a pressing need for prospective validation studies in clinical settings and standardized analytical frameworks to bridge this implementation gap.

The evolution of AI cough-based disease classification has been particularly instructive. When our initial manuscript was published in 2022, the field was characterized by considerable optimism, driven by promising early results from cough classification models. However, subsequent research, including our own studies, has revealed important complexities and limitations in using AI for disease detection through cough sounds. This has fostered a more nuanced and cautious approach to the field, highlighting the need for advances across multiple domains: technological innovation, model development, analytical methodologies, and real-world implementation.

Our external validation studies of COVID-19 and TB cough classification models revealed important implications for future research in this field. AI-based cough classification algorithms should prioritize local development and validation within target populations. While

methodological insights from local models can inform development in other settings, our evidence suggests that direct model transfer across populations is unlikely to be successful. This understanding should guide future research approaches in cough-based disease classification. Other important considerations for future cough-based classification include investigating how the presence of other respiratory diseases may confound binary COVID-19 positive vs. negative classification performance. Consideration should also be given towards expanding beyond binary classification to develop disease differentiation algorithms capable of distinguishing between multiple conditions such as COVID-19, TB, and influenza. As highlighted in **Manuscript I**, the development of such sophisticated classification systems will require large cough datasets with rigorous reference standard testing, a logistical challenge that remains a significant barrier to advancing this field.

In the specific context of TB treatment monitoring, cough presents a promising non-invasive biomarker for disease progression. Our research has established important foundational knowledge about the relationship between cough dynamics and key clinical factors such as TB bacterial burden. Building on this foundation, several key areas require further investigation. Research examining cough patterns throughout the entire treatment period is needed to better understand long-term dynamics, while studies investigating the association between cough patterns and treatment outcomes are essential to validate cough as a treatment response marker. Advances in cough diarization technology are also necessary to address measurement error from environmental cough sounds, ultimately improving the accuracy and reliability of cough monitoring for TB treatment response.

While this thesis explored two distinct perspectives of cough as a biomarker (longitudinal cough counts and cough acoustics) there may be value in integrating these complementary features. For example, in the context of TB treatment monitoring, treatment efficacy might be reflected not only in changes in cough frequency over time but also in the acoustic properties of the coughs themselves. This remains an unexplored area of research that could enhance our understanding of disease progression and treatment response for TB and other respiratory disease.

More qualitative investigations need to supplement the validation studies. For instance, investigation of the feasibility and acceptability of longitudinal cough monitoring among TB patients and healthcare providers in real-world settings will be crucial for successful implementation. From a technological perspective, feedback from end-users is necessary to aid in the development of more user-friendly cough monitoring tools which may in turn help improve issues related to data quality.

Box 8.1. Research questions for future directions.

1. How does the presence of concurrent respiratory infections and conditions affect the performance of cough-based disease classification algorithms?
2. What is the relationship between longitudinal cough patterns during TB treatment and end-of-treatment clinical outcomes, including treatment success and failure?
3. How can the integration of cough frequency metrics and acoustic properties enhance our understanding of disease progression and treatment response?
4. What are the key barriers and facilitators to implementing digital cough monitoring tools in routine clinical care from both patient and provider perspective?

8.5. Conclusion

We are entering an era of unprecedented digital innovation and big data analytics, transforming various sectors of society including healthcare. As medical practice increasingly embraces digital solutions, there is growing interest in leveraging technology to enhance clinical decision-making and patient care. Acoustic epidemiology emerges within this context as a novel field that seeks to transform subjective symptoms into objective, measurable biomarkers through digital tools and AI analytics. While this approach holds theoretical promise, particularly in the context of using cough as a biomarker for respiratory diseases, there remains limited evidence supporting its practical clinical impact. Our research has demonstrated that despite technological advances in both longitudinal cough monitoring and AI-based disease screening, significant challenges persist in translating these innovations into clinically meaningful tools. These challenges are particularly evident when attempting to validate these technologies across different populations and disease contexts, highlighting the need for continued research that rigorously evaluates the real-world applicability of these emerging technologies.

Overall, while this thesis does not claim to have fully validated cough as a clinical biomarker for any disease and all relevant use case, it has made substantial contributions toward this goal. Our findings provide crucial insights for researchers, clinicians, and technology developers working to bridge the gap between innovative digital health solutions and practical clinical implementation, particularly in resource-limited setting.

Back matter

References

1. Bouayed J. Sorry, I am sneezing and coughing but I do not have COVID-19. *Brain, Behavior, and Immunity*. 2022;101:57-58. doi:10.1016/j.bbi.2021.12.018
2. Ghrabli S, Elgendi M, Menon C. Challenges and Opportunities of Deep Learning for Cough-Based COVID-19 Diagnosis: A Scoping Review. *Diagnostics*. 2022;12(9):2142. doi:10.3390/diagnostics12092142
3. Pitts T. Airway Protective Mechanisms. *Lung*. 2014;192(1):27-31. doi:10.1007/s00408-013-9540-y
4. Morice AH. Epidemiology of Cough. *Pulmonary Pharmacology & Therapeutics*. 2002;15(3):253-259. doi:10.1006/pupt.2002.0352
5. Sharma S, Hashmi MF, Alhajjaj MS. Cough. In: *StatPearls*. StatPearls Publishing; 2023. Accessed November 13, 2024. http://www.ncbi.nlm.nih.gov/books/NBK493221/
6. Melek Manshouri N. Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. *Cogn Neurodyn*. 2022;16(1):239-253. doi:10.1007/s11571-021-09695-w
7. Haritaoglu ED, Rasmussen N, Tan DCH, et al. Using Deep Learning with Large Aggregated Datasets for COVID-19 Classification from Cough. Published online 2022. doi:10.48550/ARXIV.2201.01669
8. Chandy ST, Thangakunam B, Rudraraju G, et al. Identification of Respiratory Diseases using Peak Energy Analysis of Acoustic Cough. Published online May 31, 2024. doi:10.1101/2024.05.29.24308077
9. Fukuhara A, Saito J, Birring SS, et al. Clinical Characteristics of Cough Frequency Patterns in Patients with and without Asthma. *The Journal of Allergy and Clinical Immunology: In Practice*. 2020;8(2):654-661. doi:10.1016/j.jaip.2019.08.053
10. Huddart S, Asege L, Jaganath D, et al. Continuous cough monitoring: a novel digital biomarker for TB diagnosis and treatment response monitoring. *int j tuberc lung dis*. 2023;27(3):221-222. doi:10.5588/ijtld.22.0511
11. Biga R, Nottebaum S, Kozlakidis Z, Psomiadis S. Digitalization of Healthcare in LMICs: Digital Health and the Digital Divide Based on Technological Availability and Development. In: Kozlakidis Z, Muradyan A, Sargsyan K, eds. *Digitalization of Medicine in Low- and Middle-Income Countries*. Sustainable Development Goals Series. Springer International Publishing; 2024:185-193. doi:10.1007/978-3-031-62332-5_18
12. Zimmer AJ, Ugarte-Gil C, Pathri R, et al. Making cough count in tuberculosis care. *Commun Med*. 2022;2(1):83. doi:10.1038/s43856-022-00149-w
13. Pathri R, Jha S, Tandon S, GangaShetty S. Acoustic Epidemiology of Pulmonary Tuberculosis (TB) & Covid19 leveraging AI/ML. Published online February 8, 2022. doi:10.1101/2022.02.05.22269707
14. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. *Nat Med*. 2022;28(1):31-38. doi:10.1038/s41591-021-01614-0

15. Song WJ, Chang YS, Faruqi S, et al. The global epidemiology of chronic cough in adults: a systematic review and meta-analysis. *Eur Respir J*. 2015;45(5):1479-1481. doi:10.1183/09031936.00218714
16. Bergmann M, Haasenritter J, Beidatsch D, et al. Prevalence, aetiologies and prognosis of the symptom cough in primary care: a systematic review and meta-analysis. *BMC Fam Pract*. 2021;22(1):151. doi:10.1186/s12875-021-01501-0
17. Dicpinigaitis PV, Rauf K. The Influence of Gender on Cough Reflex Sensitivity. *Chest*. 1998;113(5):1319-1321. doi:10.1378/chest.113.5.1319
18. Dicpinigaitis PV, Allusson VRC, Baldanti A, Nalamati JR. Ethnic and Gender Differences in Cough Reflex Sensitivity. *Respiration*. 2001;68(5):480-482. doi:10.1159/000050554
19. Morice AH, Jakes AD, Faruqi S, et al. A worldwide survey of chronic cough: a manifestation of enhanced somatosensory response. *Eur Respir J*. 2014;44(5):1149-1155. doi:10.1183/09031936.00217813
20. Momtazmanesh S, Moghaddam SS, Ghamari SH, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. *eClinicalMedicine*. 2023;59:101936. doi:10.1016/j.eclinm.2023.101936
21. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1204-1222. doi:10.1016/S0140-6736(20)30925-9
22. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. *Acta Bio Medica Atenei Parmensis*. 2020;91(1):157-160. doi:10.23750/abm.v91i1.9397
23. World Health Organization. WHO COVID-19 dashboard. datadot. Accessed October 15, 2024. https://covid19.who.int/
24. Duroseau B, Kipshidze N, Limaye RJ. The impact of delayed access to COVID-19 vaccines in low- and lower-middle-income countries. *Front Public Health*. 2022;10:1087138. doi:10.3389/fpubh.2022.1087138
25. Ferrari AJ, Santomauro DF, Aali A, et al. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. *The Lancet*. 2024;403(10440):2133-2161. doi:10.1016/S0140-6736(24)00757-8
26. Andre M, Lau LS, Pokharel MD, et al. From Alpha to Omicron: How Different Variants of Concern of the SARS-Coronavirus-2 Impacted the World. *Biology*. 2023;12(9):1267. doi:10.3390/biology12091267
27. Dhama K, Nainu F, Frediansyah A, et al. Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies. *Journal of Infection and Public Health*. 2023;16(1):4-14. doi:10.1016/j.jiph.2022.11.024
28. Torabi SH, Riahi SM, Ebrahimzadeh A, Salmani F. Changes in symptoms and characteristics of COVID-19 patients across different variants: two years study using neural network analysis. *BMC Infect Dis*. 2023;23(1):838. doi:10.1186/s12879-023-08813-9

29. Viner RM, Ward JL, Hudson LD, et al. Systematic review of reviews of symptoms and signs of COVID-19 in children and adolescents. *Arch Dis Child*. 2021;106(8):802-807. doi:10.1136/archdischild-2020-320972
30. Cabrera Martimbianco AL, Pacheco RL, Bagattini ÂM, Riera R. Frequency, signs and symptoms, and criteria adopted for long COVID-19: A systematic review. *Int J Clin Pract*. 2021;75(10). doi:10.1111/ijcp.14357
31. World Health Organization. *Global Tuberculosis Report 2024*.; 2024. Accessed November 18, 2024. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024
32. Bhargava A, Bhargava M, Juneja A. Social determinants of tuberculosis: context, framework, and the way forward to ending TB in India. *Expert Review of Respiratory Medicine*. 2021;15(7):867-883. doi:10.1080/17476348.2021.1832469
33. Sossen B, Richards AS, Heinsohn T, et al. The natural history of untreated pulmonary tuberculosis in adults: a systematic review and meta-analysis. *The Lancet Respiratory Medicine*. 2023;11(4):367-379. doi:10.1016/S2213-2600(23)00097-8
34. Horton KC, Richards AS, Emery JC, Esmail H, Houben RMGJ. Reevaluating progression and pathways following *Mycobacterium tuberculosis* infection within the spectrum of tuberculosis. *Proc Natl Acad Sci USA*. 2023;120(47):e2221186120. doi:10.1073/pnas.2221186120
35. Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. *Breathe*. 2021;17(3):210079. doi:10.1183/20734735.0079-2021
36. Kendall EA, Shrestha S, Dowdy DW. The Epidemiological Importance of Subclinical Tuberculosis. A Critical Reappraisal. *Am J Respir Crit Care Med*. 2021;203(2):168-174. doi:10.1164/rccm.202006-2394PP
37. World Health Organization. *WHO Operational Handbook on Tuberculosis - Module 2: Systematic Screening for Tuberculosis Disease*. WHO; 2022. https://www.who.int/publications/i/item/9789240022614
38. Frascella B, Richards AS, Sossen B, et al. Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. *Clinical Infectious Diseases*. 2021;73(3):e830-e841. doi:10.1093/cid/ciaa1402
39. Ngang PN, Ntaganira J, Kalk A, Wolter S, Ecks S. Perceptions and beliefs about cough and tuberculosis and implications for TB control in rural Rwanda. *Int J Tuberc Lung Dis*. 2007;11(10):1108-1113.
40. Field SK, Escalante P, Fisher DA, et al. Cough Due to TB and Other Chronic Infections. *Chest*. 2018;153(2):467-497. doi:10.1016/j.chest.2017.11.018
41. Cho SH, Lin HC, Ghoshal AG, et al. Respiratory disease in the Asia-Pacific region: Cough as a key symptom. *allergy asthma proc*. 2016;37(2):131-140. doi:10.2500/aap.2016.37.3925
42. Sun X, Luo S, Lou L, et al. Health seeking behavior and associated factors among individuals with cough in Yiwu, China: a population-based study. *BMC Public Health*. 2021;21(1):1157. doi:10.1186/s12889-021-11250-5
43. Blaiss MS, Dicpinigaitis PV, Eccles R, Wingertzahn MA. Consumer attitudes on cough and cold: US (ACHOO) survey results. *Current Medical Research and Opinion*. 2015;31(8):1527-1538. doi:10.1185/03007995.2014.1002558

44. Allan W. The symptoms of newly diagnosed pulmonary tuberculosis and patients' attitudes to the disease and to its treatment in Kong Hong. *Tubercle*. 1979;60(4):211-223. doi:10.1016/0041-3879(79)90002-3
45. Chang AB, Glomb WB. Guidelines for Evaluating Chronic Cough in Pediatrics. *Chest*. 2006;129(1):260S-283S. doi:10.1378/chest.129.1_suppl.260S
46. Morice AH, Fontana GA, Belvisi MG, et al. ERS guidelines on the assessment of cough. *European Respiratory Journal*. 2007;29(6):1256-1276. doi:10.1183/09031936.00101006
47. Morice AH, Millqvist E, Bieksiene K, et al. ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. *Eur Respir J*. 2020;55(1):1901136. doi:10.1183/13993003.01136-2019
48. Pauwels RA, Buist AS, Calverley PMA, Jenkins CR, Hurd SS. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. *Am J Respir Crit Care Med*. 2001;163(5):1256-1276. doi:10.1164/ajrccm.163.5.2101039
49. Grief SN, Loza JK. Guidelines for the Evaluation and Treatment of Pneumonia. *Prim Care*. 2018;45(3):485-503. doi:10.1016/j.pop.2018.04.001
50. Hughes R, Rapsomaniki E, Janson C, et al. Frequent productive cough: Symptom burden and future exacerbation risk among patients with asthma and/or COPD in the NOVELTY study. *Respiratory Medicine*. 2022;200:106921. doi:10.1016/j.rmed.2022.106921
51. Smith J, Woodcock A. Cough and its importance in COPD. *International Journal of COPD*. 2006;1(3):305-314. doi:10.2147/copd.2006.1.3.305
52. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention. Global Initiative for Asthma - GINA. 2024. Accessed October 9, 2024. https://ginasthma.org/2024-report/
53. Parker D, Picone J, Harati A, Lu S, Jenkyns MH, Polgreen PM. Detecting Paroxysmal Coughing from Pertussis Cases Using Voice Recognition Technology. Hozbor DF, ed. *PLoS ONE*. 2013;8(12):e82971. doi:10.1371/journal.pone.0082971
54. Wood N, McIntyre P. Pertussis: review of epidemiology, diagnosis, management and prevention. *Paediatric Respiratory Reviews*. 2008;9(3):201-212. doi:10.1016/j.prrv.2008.05.010
55. Boujaoude ZC, Pratter MR. Clinical Approach to Acute Cough. *Lung*. 2010;188(S1):41-46. doi:10.1007/s00408-009-9170-6
56. Ryerson CJ, Abbritti M, Ley B, Elicker BM, Jones KD, Collard HR. Cough predicts prognosis in idiopathic pulmonary fibrosis. *Respirology*. 2011;16(6):969-975. doi:10.1111/j.1440-1843.2011.01996.x
57. Marsden PA, Satia I, Ibrahim B, et al. Objective Cough Frequency, Airway Inflammation, and Disease Control in Asthma. *Chest*. 2016;149(6):1460-1466. doi:10.1016/j.chest.2016.02.676
58. Smith JA, Ashurst HL, Jack S, Woodcock AA, Earis JE. The description of cough sounds by healthcare professionals. *Cough*. 2006;2:1. doi:10.1186/1745-9974-2-1
59. Jakusova J, Brozmanova M. Methods of Cough Assessment and Objectivization. *Physiol Res*. Published online 2023:687-700. doi:10.33549/physiolres.935062
60. Wang Z, Wang M, Wen S, Yu L, Xu X. Types and applications of cough-related questionnaires. *J Thorac Dis*. 2019;11(10):4379-4388. doi:10.21037/jtd.2019.09.62

61. Spinou A, Birring SS. An update on measurement and monitoring of cough: what are the important study endpoints? *J Thorac Dis*. 2014;6(Suppl 7):S728-734. doi:10.3978/j.issn.2072-1439.2014.10.08
62. Su X, Huang Y, Dai W, et al. Recall Bias in the Assessment of Cough for Patients Discharged from Lung Surgery. *Patient Prefer Adherence*. 2023;17:1561-1572. doi:10.2147/PPA.S399635
63. Martin Nguyen A, Bacci ED, Vernon M, et al. Validation of a visual analog scale for assessing cough severity in patients with chronic cough. *Ther Adv Respir Dis*. 2021;15:175346662110497. doi:10.1177/17534666211049743
64. Irwin RS, Madison JM. The Diagnosis and Treatment of Cough. *N Engl J Med*. 2000;343(23):1715-1721. doi:10.1056/NEJM200012073432308
65. Korpáš J, Sadloňová J, Vrabec M. Analysis of the Cough Sound: an Overview. *Pulmonary Pharmacology*. 1996;9(5-6):261-268. doi:10.1006/pulp.1996.0034
66. Irwin RS, French CL, Chang AB, et al. Classification of Cough as a Symptom in Adults and Management Algorithms. *Chest*. 2018;153(1):196-209. doi:10.1016/j.chest.2017.10.016
67. Gibson PG, Chang AB, Glasgow NJ, et al. CICADA: Cough in Children and Adults: Diagnosis and Assessment. Australian Cough Guidelines summary statement. *Medical Journal of Australia*. 2010;192(5):265-271. doi:10.5694/j.1326-5377.2010.tb03504.x
68. Bösmüller H, Matter M, Fend F, Tzankov A. The pulmonary pathology of COVID-19. *Virchows Arch*. 2021;478(1):137-150. doi:10.1007/s00428-021-03053-1
69. Tian S, Hu W, Niu L, Liu H, Xu H, Xiao SY. Pulmonary Pathology of Early-Phase 2019 Novel Coronavirus (COVID-19) Pneumonia in Two Patients With Lung Cancer. *Journal of Thoracic Oncology*. 2020;15(5):700-704. doi:10.1016/j.jtho.2020.02.010
70. Bai HX, Hsieh B, Xiong Z, et al. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. *Radiology*. 2020;296(2):E46-E54. doi:10.1148/radiol.2020200823
71. Alshoabi SA, Almas KM, Aldofri SA, et al. The Diagnostic Deceiver: Radiological Pictorial Review of Tuberculosis. *Diagnostics*. 2022;12(2):306. doi:10.3390/diagnostics12020306
72. Khan FA, Majidulla A, Tavaziva G, et al. Chest x-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: a prospective study of diagnostic accuracy for culture-confirmed disease. *The Lancet Digital Health*. 2020;2(11):e573-e581. doi:10.1016/S2589-7500(20)30221-1
73. Bhalla AS, Goyal A, Guleria R, Gupta AK. Chest tuberculosis: Radiological review and imaging recommendations. *Indian Journal of Radiology and Imaging*. 2015;25(03):213-225. doi:10.4103/0971-3026.161431
74. Brown C, Chauhan J, Grammenos A, et al. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. Published online 2020. doi:10.48550/ARXIV.2006.05919
75. Banerjee A, Nilhani A. A Residual Network based Deep Learning Model for Detection of COVID-19 from Cough Sounds. Published online 2021. doi:10.48550/ARXIV.2106.02348
76. Ghrabli S, Elgendi M, Menon C. Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments. *Sci Rep*. 2024;14(1):593. doi:10.1038/s41598-023-50371-2

77. Zhang J, Chen S, Wang J, et al. Research on the Identification Method of Respiratory Tract Infections Based on Multi-parameter Fusion. Published online May 31, 2024. doi:10.21203/rs.3.rs-4430123/v1
78. Huddart S, Yadav V, Sieberts SK, et al. Solicited Cough Sound Analysis for Tuberculosis Triage Testing: The CODA TB DREAM Challenge Dataset. Published online March 28, 2024. doi:10.1101/2024.03.27.24304980
79. Orlandic L, Teijeiro T, Atienza D. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Sci Data*. 2021;8(1):156. doi:10.1038/s41597-021-00937-4
80. Bhattacharya D, Sharma NK, Dutta D, et al. Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection. *Sci Data*. 2023;10(1):397. doi:10.1038/s41597-023-02266-0
81. Xiong H, Berkovsky S, Kâafar MA, Jaffe A, Coiera E, Sharan RV. Reliability of crowdsourced data and patient-reported outcome measures in cough-based COVID-19 screening. *Sci Rep*. 2022;12(1):21990. doi:10.1038/s41598-022-26492-5
82. Liu J, Li J, Li W, Wu J. Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2016;115:134-142. doi:10.1016/j.isprsjprs.2015.11.006
83. Coppock H, Akman A, Bergler C, et al. A summary of the ComParE COVID-19 challenges. *Front Digit Health*. 2023;5:1058163. doi:10.3389/fdgth.2023.1058163
84. Pizzo DT, Esteban S. IATos: AI-powered pre-screening tool for COVID-19 from cough audio samples. Published online 2021. doi:10.48550/ARXIV.2104.13247
85. Cohen-McFarlane M, Goubran R, Knoefel F. Novel Coronavirus Cough Database: NoCoCoDa. *IEEE Access*. 2020;8:154087-154094. doi:10.1109/ACCESS.2020.3018028
86. Zarkogianni K, Dervakos E, Filandrianos G, et al. The smarty4covid dataset and knowledge base as a framework for interpretable physiological audio data analysis. *Sci Data*. 2023;10(1):770. doi:10.1038/s41597-023-02646-6
87. Chaudhari G, Jiang X, Fakhry A, et al. Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough. Published online 2020. doi:10.48550/ARXIV.2011.13320
88. Galvosas M, Gabaldón-Figueira JC, Keen EM, et al. Performance evaluation of the smartphone-based AI cough monitoring app - Hyfe Cough Tracker against solicited respiratory sounds. *F1000Res*. 2023;11:730. doi:10.12688/f1000research.122597.2
89. Sanchez-Morillo D, Sales-Lerida D, Priego-Torres B, León-Jiménez A. Cough Detection Using Acceleration Signals and Deep Learning Techniques. *Electronics*. 2024;13(12):2410. doi:10.3390/electronics13122410
90. Muguli A, Pinto L, R. N, et al. DiCOVA Challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics. Published online 2021. doi:10.48550/ARXIV.2103.09148
91. Bagad P, Dalmia A, Doshi J, et al. Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds. Published online 2020. doi:10.48550/ARXIV.2009.08790
92. Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*. 2020;20:100378. doi:10.1016/j.imu.2020.100378

93. Tena A, Clarià F, Solsona F. Automated detection of COVID-19 cough. *Biomedical Signal Processing and Control*. 2022;71:103175. doi:10.1016/j.bspc.2021.103175
94. Ponomarchuk A, Burenko I, Malkin E, et al. Project Achoo: A Practical Model and Application for COVID-19 Detection From Recordings of Breath, Voice, and Cough. IEEE J Sel Top Signal Process. 2022;16(2):175-187. doi:10.1109/JSTSP.2022.3142514
95. Chowdhury NK, Kabir MA, Rahman MdM, Islam SMS. Machine learning for detecting COVID-19 from cough sounds: An ensemble-based MCDM method. Computers in Biology and Medicine. 2022;145:105405. doi:10.1016/j.compbiomed.2022.105405
96. Hamdi S, Oussalah M, Moussaoui A, Saidi M. Attention-based hybrid CNN-LSTM and spectral data augmentation for COVID-19 diagnosis from cough sound. J Intell Inf Syst. 2022;59(2):367-389. doi:10.1007/s10844-022-00707-7
97. Kapoor T, Pandhi T, Gupta B. Cough Audio Analysis for COVID-19 Diagnosis. SN COMPUT SCI. 2022;4(2):125. doi:10.1007/s42979-022-01522-1
98. Melek M. Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound. Neural Comput & Applic. 2021;33(24):17621-17632. doi:10.1007/s00521-021-06346-3
99. Nguyen LH, Pham NT, Do VH, et al. Fruit-CoV: An efficient vision-based framework for speedy detection and diagnosis of SARS-CoV-2 infections through recorded cough sounds. Expert Systems with Applications. 2023;213:119212. doi:10.1016/j.eswa.2022.119212
100. Nguyen-Trong K, Nguyen-Hoang K. Multi-modal approach for COVID-19 detection using coughs and self-reported symptoms. IFS. 2023;44(3):3501-3513. doi:10.3233/JIFS-222863
101. Ren Z, Chang Y, Bartl-Pokorny KD, Pokorny FB, Schuller BW. The Acoustic Dissection of Cough: Diving Into Machine Listening-based COVID-19 Analysis and Detection. Journal of Voice. Published online June 2022:S0892199722001667. doi:10.1016/j.jvoice.2022.06.011
102. Sunitha G, Arunachalam R, Abd-Elnaby M, Eid MMA, Rashed ANZ. A comparative analysis of deep neural network architectures for the dynamic diagnosis of COVID-19 based on acoustic cough features. Int J Imaging Syst Tech. 2022;32(5):1433-1446. doi:10.1002/ima.22749
103. Akgün D, Kabakuş AT, Karapinar Şentürk Z, Şentürk A, Küçükkülahli E. A transfer learning-based deep learning approach for automated COVID-19 diagnosis with audio data. Turk J Elec Eng & Comp Sci. 2021;29(SI-1):2807-2823. doi:10.3906/elk-2105-64
104. Islam R, Abdel-Raheem E, Tarique M. A study of using cough sounds and deep neural networks for the early detection of Covid-19. Biomedical Engineering Advances. 2022;3:100025. doi:10.1016/j.bea.2022.100025
105. Najaran MHT. An evolutionary ensemble learning for diagnosing COVID-19 via cough signals. Intelligent Medicine. 2023;3(3):200-212. doi:10.1016/j.imed.2023.01.001
106. Ulukaya S, Sarıca AA, Erdem O, Karaali A. MSCCov19Net: multi-branch deep learning model for COVID-19 detection from cough sounds. Med Biol Eng Comput. 2023;61(7):1619-1629. doi:10.1007/s11517-023-02803-4
107. Hemdan EED, El-Shafai W, Sayed A. CR19: a framework for preliminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications. J Ambient Intell Human Comput. 2023;14(9):11715-11727. doi:10.1007/s12652-022-03732-0

108. Alaaeldein O, El Ahl OS, Elmahy L, Ihab M, Gomaa W. Analysis of COVID-19 Using Imaging and Audio Modalities. In: 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE; 2022:1-10. doi:10.1109/IMCOM53663.2022.9721730
109. Elrahman SMA, Abraham A. A Review of Class Imbalance Problem. *Journal of Network and Innovative Computing*. 2013;1:9-9.
110. Pahar M, Klopper M, Warren R, Niesler T. COVID-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine*. 2021;135:104572. doi:10.1016/j.compbiomed.2021.104572
111. Akman A, Coppock H, Gaskell A, Tzirakis P, Jones L, Schuller BW. Evaluating the COVID-19 Identification ResNet (CIdeR) on the INTERSPEECH COVID-19 From Audio Challenges. *Front Digit Health*. 2022;4:789980. doi:10.3389/fdgth.2022.789980
112. Awais M, Bhuva A, Bhuva D, Fatima S, Sadiq T. Optimized DEC: An effective cough detection framework using optimal weighted Features-aided deep Ensemble classifier for COVID-19. *Biomedical Signal Processing and Control*. 2023;86:105026. doi:10.1016/j.bspc.2023.105026
113. Celik G. CovidCoughNet: A new method based on convolutional neural networks and deep feature extraction using pitch-shifting data augmentation for covid-19 detection from cough, breath, and voice signals. *Computers in Biology and Medicine*. 2023;163:107153. doi:10.1016/j.compbiomed.2023.107153
114. Chang Y, Jing X, Ren Z, Schuller BW. CovNet: A Transfer Learning Framework for Automatic COVID-19 Detection From Crowd-Sourced Cough Sounds. *Front Digit Health*. 2022;3:799067. doi:10.3389/fdgth.2021.799067
115. Despotovic V, Ismael M, Cornil M, Call RM, Fagherazzi G. Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results. *Computers in Biology and Medicine*. 2021;138:104944. doi:10.1016/j.compbiomed.2021.104944
116. Elizalde B, Tompkins D. COVID-19 Detection Using Recorded Coughs in the 2021 DiCOVA Challenge. Published online 2021. doi:10.48550/ARXIV.2105.10619
117. Erdoğan YE, Narin A. COVID-19 detection with traditional and deep features on cough acoustic signals. *Computers in Biology and Medicine*. 2021;136:104765. doi:10.1016/j.compbiomed.2021.104765
118. Hoang T, Pham L, Ngo D, Nguyen HD. A Cough-based deep learning framework for detecting COVID-19. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE; 2022:3422-3425. doi:10.1109/EMBC48229.2022.9871179
119. Kamble MR, Gonzalez-Lopez JA, Grau T, et al. PANACEA cough sound-based diagnosis of COVID-19 for the DiCOVA 2021 Challenge. Published online 2021. doi:10.48550/ARXIV.2106.04423
120. Laguarta J, Hueto F, Subirana B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. *IEEE Open J Eng Med Biol*. 2020;1:275-281. doi:10.1109/OJEMB.2020.3026928
121. Lella KK, Pja A. Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath. *Alexandria Engineering Journal*. 2022;61(2):1319-1334. doi:10.1016/j.aej.2021.06.024

122. Liu S, Mallol-Ragolta A, Schuller BW. COVID-19 Detection with a Novel Multi-Type Deep Fusion Method using Breathing and Coughing Information. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE; 2021:1840-1843. doi:10.1109/EMBC46164.2021.9630050
123. Loey M, Mirjalili S. COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models. *Computers in Biology and Medicine*. 2021;139:105020. doi:10.1016/j.compbiomed.2021.105020
124. Mitrofanova A, Mikhaylov D, Shaznaev I, Chumanskaia V, Saveliev V. Acoustery System for Differential Diagnosing of Coronavirus COVID-19 Disease. *IEEE Open J Eng Med Biol*. 2021;2:299-303. doi:10.1109/OJEMB.2021.3127078
125. Mohammed EA, Keyhani M, Sanati-Nezhad A, Hejazi SH, Far BH. An ensemble learning approach to digital corona virus preliminary screening from cough sounds. *Sci Rep*. 2021;11(1):15404. doi:10.1038/s41598-021-95042-2
126. Mouawad P, Dubnov T, Dubnov S. Robust Detection of COVID-19 in Cough Sounds: Using Recurrence Dynamics and Variable Markov Model. *SN COMPUT SCI*. 2021;2(1):34. doi:10.1007/s42979-020-00422-6
127. Askari Nasab K, Mirzaei J, Zali A, Gholizadeh S, Akhlaghdoust M. Coronavirus diagnosis using cough sounds: Artificial intelligence approaches. *Front Artif Intell*. 2023;6:1100112. doi:10.3389/frai.2023.1100112
128. Rahman T, Ibtehaz N, Khandakar A, et al. QUCoughScope: An Intelligent Application to Detect COVID-19 Patients Using Cough and Breath Sounds. *Diagnostics*. 2022;12(4):920. doi:10.3390/diagnostics12040920
129. Nessiem MA, Mohamed MM, Coppock H, Gaskell A, Schuller BW. Detecting COVID-19 from Breathing and Coughing Sounds using Deep Neural Networks. In: *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE; 2021:183-188. doi:10.1109/CBMS52027.2021.00069
130. Sobahi N, Atila O, Deniz E, Sengur A, Acharya UR. Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds. *Biocybernetics and Biomedical Engineering*. 2022;42(3):1066-1080. doi:10.1016/j.bbe.2022.08.005
131. Hope TMH. Linear regression. In: *Machine Learning*. Elsevier; 2020:67-81. doi:10.1016/B978-0-12-815739-8.00004-3
132. Liang D, Frederick DA, Lledo EE, et al. Examining the utility of nonlinear machine learning approaches versus linear regression for predicting body image outcomes: The U.S. Body Project I. *Body Image*. 2022;41:32-45. doi:10.1016/j.bodyim.2022.01.013
133. Peng Y, Nagata MH. An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. *Chaos, Solitons & Fractals*. 2020;139:110055. doi:10.1016/j.chaos.2020.110055
134. Dietterich TG. Ensemble Methods in Machine Learning. In: *Multiple Classifier Systems*. Vol 1857. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2000:1-15. doi:10.1007/3-540-45014-9_1
135. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks*. 2015;61:85-117. doi:10.1016/j.neunet.2014.09.003
136. O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. Published online 2015. doi:10.48550/ARXIV.1511.08458

137. Cabitza F, Campagner A, Soares F, et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Computer Methods and Programs in Biomedicine*. 2021;208:106288. doi:10.1016/j.cmpb.2021.106288
138. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. *Nat Med*. 2023;29(11):2686-2687. doi:10.1038/s41591-023-02540-z
139. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med*. 2021;181(8):1065. doi:10.1001/jamainternmed.2021.2626
140. Habib AR, Lin AL, Grant RW. The Epic Sepsis Model Falls Short—The Importance of External Validation. *JAMA Intern Med*. 2021;181(8):1040. doi:10.1001/jamainternmed.2021.3333
141. Khine M. Editorial: Clinical validation of digital health technologies for personalized medicine. *Front Digit Health*. 2022;4:831517. doi:10.3389/fdgth.2022.831517
142. Makri A. Bridging the digital divide in health care. *The Lancet Digital Health*. 2019;1(5):e204-e205. doi:10.1016/S2589-7500(19)30111-6
143. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and Facilitators of Digital Health Among Culturally and Linguistically Diverse Populations: Qualitative Systematic Review. *J Med Internet Res*. 2023;25:e42719. doi:10.2196/42719

Complete reference list

1. Bouayed J. Sorry, I am sneezing and coughing but I do not have COVID-19. Brain, Behavior, and Immunity. 2022;101:57-58. doi:10.1016/j.bbi.2021.12.018
2. Ghrabli S, Elgendi M, Menon C. Challenges and Opportunities of Deep Learning for Cough-Based COVID-19 Diagnosis: A Scoping Review. Diagnostics. 2022;12(9):2142. doi:10.3390/diagnostics12092142
3. Pitts T. Airway Protective Mechanisms. Lung. 2014;192(1):27-31. doi:10.1007/s00408-013-9540-y
4. Morice AH. Epidemiology of Cough. Pulmonary Pharmacology & Therapeutics. 2002;15(3):253-259. doi:10.1006/pupt.2002.0352
5. Sharma S, Hashmi MF, Alhajjaj MS. Cough. In: StatPearls. StatPearls Publishing; 2023. Accessed November 13, 2024. http://www.ncbi.nlm.nih.gov/books/NBK493221/
6. Melek Manshouri N. Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. Cogn Neurodyn. 2022;16(1):239-253. doi:10.1007/s11571-021-09695-w
7. Haritaoglu ED, Rasmussen N, Tan DCH, et al. Using Deep Learning with Large Aggregated Datasets for COVID-19 Classification from Cough. Published online 2022. doi:10.48550/ARXIV.2201.01669

8. Chandy ST, Thangakunam B, Rudraraju G, et al. Identification of Respiratory Diseases using Peak Energy Analysis of Acoustic Cough. Published online May 31, 2024. doi:10.1101/2024.05.29.24308077
9. Fukuhara A, Saito J, Birring SS, et al. Clinical Characteristics of Cough Frequency Patterns in Patients with and without Asthma. The Journal of Allergy and Clinical Immunology: In Practice. 2020;8(2):654-661. doi:10.1016/j.jaip.2019.08.053
10. Huddart S, Asege L, Jaganath D, et al. Continuous cough monitoring: a novel digital biomarker for TB diagnosis and treatment response monitoring. int j tuberc lung dis. 2023;27(3):221-222. doi:10.5588/ijtld.22.0511
11. Biga R, Nottebaum S, Kozlakidis Z, Psomiadis S. Digitalization of Healthcare in LMICs: Digital Health and the Digital Divide Based on Technological Availability and Development. In: Kozlakidis Z, Muradyan A, Sargsyan K, eds. Digitalization of Medicine in Low- and Middle-Income Countries. Sustainable Development Goals Series. Springer International Publishing; 2024:185-193. doi:10.1007/978-3-031-62332-5_18
12. Zimmer AJ, Ugarte-Gil C, Pathri R, et al. Making cough count in tuberculosis care. Commun Med. 2022;2(1):83. doi:10.1038/s43856-022-00149-w
13. Pathri R, Jha S, Tandon S, GangaShetty S. Acoustic Epidemiology of Pulmonary Tuberculosis (TB) & Covid19 leveraging AI/ML. Published online February 8, 2022. doi:10.1101/2022.02.05.22269707
14. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28(1):31-38. doi:10.1038/s41591-021-01614-0
15. Song WJ, Chang YS, Faruqi S, et al. The global epidemiology of chronic cough in adults: a systematic review and meta-analysis. Eur Respir J. 2015;45(5):1479-1481. doi:10.1183/09031936.00218714
16. Bergmann M, Haasenritter J, Beidatsch D, et al. Prevalence, aetiologies and prognosis of the symptom cough in primary care: a systematic review and meta-analysis. BMC Fam Pract. 2021;22(1):151. doi:10.1186/s12875-021-01501-0
17. Dicpinigaitis PV, Rauf K. The Influence of Gender on Cough Reflex Sensitivity. Chest. 1998;113(5):1319-1321. doi:10.1378/chest.113.5.1319
18. Dicpinigaitis PV, Allusson VRC, Baldanti A, Nalamati JR. Ethnic and Gender Differences in Cough Reflex Sensitivity. Respiration. 2001;68(5):480-482. doi:10.1159/000050554
19. Morice AH, Jakes AD, Faruqi S, et al. A worldwide survey of chronic cough: a manifestation of enhanced somatosensory response. Eur Respir J. 2014;44(5):1149-1155. doi:10.1183/09031936.00217813
20. Momtazmanesh S, Moghaddam SS, Ghamari SH, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the Global Burden of Disease Study 2019. eClinicalMedicine. 2023;59:101936. doi:10.1016/j.eclinm.2023.101936
21. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1204-1222. doi:10.1016/S0140-6736(20)30925-9

22. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Bio Medica Atenei Parmensis. 2020;91(1):157-160. doi:10.23750/abm.v91i1.9397
23. World Health Organization. WHO COVID-19 dashboard. datadot. Accessed October 15, 2024. https://covid19.who.int/
24. Duroseau B, Kipshidze N, Limaye RJ. The impact of delayed access to COVID-19 vaccines in low- and lower-middle-income countries. Front Public Health. 2022;10:1087138. doi:10.3389/fpubh.2022.1087138
25. Ferrari AJ, Santomauro DF, Aali A, et al. Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. 2024;403(10440):2133-2161. doi:10.1016/S0140-6736(24)00757-8
26. Andre M, Lau LS, Pokharel MD, et al. From Alpha to Omicron: How Different Variants of Concern of the SARS-Coronavirus-2 Impacted the World. Biology. 2023;12(9):1267. doi:10.3390/biology12091267
27. Dhama K, Nainu F, Frediansyah A, et al. Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies. Journal of Infection and Public Health. 2023;16(1):4-14. doi:10.1016/j.jiph.2022.11.024
28. Torabi SH, Riahi SM, Ebrahimzadeh A, Salmani F. Changes in symptoms and characteristics of COVID-19 patients across different variants: two years study using neural network analysis. BMC Infect Dis. 2023;23(1):838. doi:10.1186/s12879-023-08813-9
29. Viner RM, Ward JL, Hudson LD, et al. Systematic review of reviews of symptoms and signs of COVID-19 in children and adolescents. Arch Dis Child. 2021;106(8):802-807. doi:10.1136/archdischild-2020-320972
30. Cabrera Martimbianco AL, Pacheco RL, Bagattini ÂM, Riera R. Frequency, signs and symptoms, and criteria adopted for long COVID-19: A systematic review. Int J Clin Pract. 2021;75(10). doi:10.1111/ijcp.14357
31. World Health Organization. Global Tuberculosis Report 2024.; 2024. Accessed November 18, 2024. https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2024
32. Bhargava A, Bhargava M, Juneja A. Social determinants of tuberculosis: context, framework, and the way forward to ending TB in India. Expert Review of Respiratory Medicine. 2021;15(7):867-883. doi:10.1080/17476348.2021.1832469
33. Sossen B, Richards AS, Heinsohn T, et al. The natural history of untreated pulmonary tuberculosis in adults: a systematic review and meta-analysis. The Lancet Respiratory Medicine. 2023;11(4):367-379. doi:10.1016/S2213-2600(23)00097-8
34. Horton KC, Richards AS, Emery JC, Esmail H, Houben RMGJ. Reevaluating progression and pathways following Mycobacterium tuberculosis infection within the spectrum of tuberculosis. Proc Natl Acad Sci USA. 2023;120(47):e2221186120. doi:10.1073/pnas.2221186120
35. Migliori GB, Ong CWM, Petrone L, D'Ambrosio L, Centis R, Goletti D. The definition of tuberculosis infection based on the spectrum of tuberculosis disease. Breathe. 2021;17(3):210079. doi:10.1183/20734735.0079-2021

36. Kendall EA, Shrestha S, Dowdy DW. The Epidemiological Importance of Subclinical Tuberculosis. A Critical Reappraisal. Am J Respir Crit Care Med. 2021;203(2):168-174. doi:10.1164/rccm.202006-2394PP
37. World Health Organization. WHO Operational Handbook on Tuberculosis - Module 2: Systematic Screening for Tuberculosis Disease. WHO; 2022. https://www.who.int/publications/i/item/9789240022614
38. Frascella B, Richards AS, Sossen B, et al. Subclinical Tuberculosis Disease—A Review and Analysis of Prevalence Surveys to Inform Definitions, Burden, Associations, and Screening Methodology. Clinical Infectious Diseases. 2021;73(3):e830-e841. doi:10.1093/cid/ciaa1402
39. Ngang PN, Ntaganira J, Kalk A, Wolter S, Ecks S. Perceptions and beliefs about cough and tuberculosis and implications for TB control in rural Rwanda. Int J Tuberc Lung Dis. 2007;11(10):1108-1113.
40. Field SK, Escalante P, Fisher DA, et al. Cough Due to TB and Other Chronic Infections. Chest. 2018;153(2):467-497. doi:10.1016/j.chest.2017.11.018
41. Cho SH, Lin HC, Ghoshal AG, et al. Respiratory disease in the Asia-Pacific region: Cough as a key symptom. allergy asthma proc. 2016;37(2):131-140. doi:10.2500/aap.2016.37.3925
42. Sun X, Luo S, Lou L, et al. Health seeking behavior and associated factors among individuals with cough in Yiwu, China: a population-based study. BMC Public Health. 2021;21(1):1157. doi:10.1186/s12889-021-11250-5
43. Blaiss MS, Dicpinigaitis PV, Eccles R, Wingertzahn MA. Consumer attitudes on cough and cold: US (ACHOO) survey results. Current Medical Research and Opinion. 2015;31(8):1527-1538. doi:10.1185/03007995.2014.1002558
44. Allan W. The symptoms of newly diagnosed pulmonary tuberculosis and patients' attitudes to the disease and to its treatment in Kong Hong. Tubercle. 1979;60(4):211-223. doi:10.1016/0041-3879(79)90002-3
45. Chang AB, Glomb WB. Guidelines for Evaluating Chronic Cough in Pediatrics. Chest. 2006;129(1):260S-283S. doi:10.1378/chest.129.1_suppl.260S
46. Morice AH, Fontana GA, Belvisi MG, et al. ERS guidelines on the assessment of cough. European Respiratory Journal. 2007;29(6):1256-1276. doi:10.1183/09031936.00101006
47. Morice AH, Millqvist E, Bieksiene K, et al. ERS guidelines on the diagnosis and treatment of chronic cough in adults and children. Eur Respir J. 2020;55(1):1901136. doi:10.1183/13993003.01136-2019
48. Pauwels RA, Buist AS, Calverley PMA, Jenkins CR, Hurd SS. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease: NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop Summary. Am J Respir Crit Care Med. 2001;163(5):1256-1276. doi:10.1164/ajrccm.163.5.2101039
49. Grief SN, Loza JK. Guidelines for the Evaluation and Treatment of Pneumonia. Prim Care. 2018;45(3):485-503. doi:10.1016/j.pop.2018.04.001

50. Hughes R, Rapsomaniki E, Janson C, et al. Frequent productive cough: Symptom burden and future exacerbation risk among patients with asthma and/or COPD in the NOVELTY study. Respiratory Medicine. 2022;200:106921. doi:10.1016/j.rmed.2022.106921
51. Smith J, Woodcock A. Cough and its importance in COPD. International Journal of COPD. 2006;1(3):305-314. doi:10.2147/copd.2006.1.3.305
52. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention. Global Initiative for Asthma - GINA. 2024. Accessed October 9, 2024. https://ginasthma.org/2024-report/
53. Parker D, Picone J, Harati A, Lu S, Jenkyns MH, Polgreen PM. Detecting Paroxysmal Coughing from Pertussis Cases Using Voice Recognition Technology. Hozbor DF, ed. PLoS ONE. 2013;8(12):e82971. doi:10.1371/journal.pone.0082971
54. Wood N, McIntyre P. Pertussis: review of epidemiology, diagnosis, management and prevention. Paediatric Respiratory Reviews. 2008;9(3):201-212. doi:10.1016/j.prrv.2008.05.010
55. Boujaoude ZC, Pratter MR. Clinical Approach to Acute Cough. Lung. 2010;188(S1):41-46. doi:10.1007/s00408-009-9170-6
56. Ryerson CJ, Abbritti M, Ley B, Elicker BM, Jones KD, Collard HR. Cough predicts prognosis in idiopathic pulmonary fibrosis. Respirology. 2011;16(6):969-975. doi:10.1111/j.1440-1843.2011.01996.x
57. Marsden PA, Satia I, Ibrahim B, et al. Objective Cough Frequency, Airway Inflammation, and Disease Control in Asthma. Chest. 2016;149(6):1460-1466. doi:10.1016/j.chest.2016.02.676
58. Smith JA, Ashurst HL, Jack S, Woodcock AA, Earis JE. The description of cough sounds by healthcare professionals. Cough. 2006;2:1. doi:10.1186/1745-9974-2-1
59. Jakusova J, Brozmanova M. Methods of Cough Assessment and Objectivization. Physiol Res. Published online 2023:687-700. doi:10.33549/physiolres.935062
60. Wang Z, Wang M, Wen S, Yu L, Xu X. Types and applications of cough-related questionnaires. J Thorac Dis. 2019;11(10):4379-4388. doi:10.21037/jtd.2019.09.62
61. Spinou A, Birring SS. An update on measurement and monitoring of cough: what are the important study endpoints? J Thorac Dis. 2014;6(Suppl 7):S728-734. doi:10.3978/j.issn.2072-1439.2014.10.08
62. Su X, Huang Y, Dai W, et al. Recall Bias in the Assessment of Cough for Patients Discharged from Lung Surgery. Patient Prefer Adherence. 2023;17:1561-1572. doi:10.2147/PPA.S399635
63. Martin Nguyen A, Bacci ED, Vernon M, et al. Validation of a visual analog scale for assessing cough severity in patients with chronic cough. Ther Adv Respir Dis. 2021;15:175346662110497. doi:10.1177/17534666211049743
64. Irwin RS, Madison JM. The Diagnosis and Treatment of Cough. N Engl J Med. 2000;343(23):1715-1721. doi:10.1056/NEJM200012073432308
65. Korpáš J, Sadloňová J, Vrabec M. Analysis of the Cough Sound: an Overview. Pulmonary Pharmacology. 1996;9(5-6):261-268. doi:10.1006/pulp.1996.0034

66. Irwin RS, French CL, Chang AB, et al. Classification of Cough as a Symptom in Adults and Management Algorithms. Chest. 2018;153(1):196-209. doi:10.1016/j.chest.2017.10.016
67. Gibson PG, Chang AB, Glasgow NJ, et al. CICADA: Cough in Children and Adults: Diagnosis and Assessment. Australian Cough Guidelines summary statement. Medical Journal of Australia. 2010;192(5):265-271. doi:10.5694/j.1326-5377.2010.tb03504.x
68. Bösmüller H, Matter M, Fend F, Tzankov A. The pulmonary pathology of COVID-19. Virchows Arch. 2021;478(1):137-150. doi:10.1007/s00428-021-03053-1
69. Tian S, Hu W, Niu L, Liu H, Xu H, Xiao SY. Pulmonary Pathology of Early-Phase 2019 Novel Coronavirus (COVID-19) Pneumonia in Two Patients With Lung Cancer. Journal of Thoracic Oncology. 2020;15(5):700-704. doi:10.1016/j.jtho.2020.02.010
70. Bai HX, Hsieh B, Xiong Z, et al. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. Radiology. 2020;296(2):E46-E54. doi:10.1148/radiol.2020200823
71. Alshoabi SA, Almas KM, Aldofri SA, et al. The Diagnostic Deceiver: Radiological Pictorial Review of Tuberculosis. Diagnostics. 2022;12(2):306. doi:10.3390/diagnostics12020306
72. Khan FA, Majidulla A, Tavaziva G, et al. Chest x-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: a prospective study of diagnostic accuracy for culture-confirmed disease. The Lancet Digital Health. 2020;2(11):e573-e581. doi:10.1016/S2589-7500(20)30221-1
73. Bhalla AS, Goyal A, Guleria R, Gupta AK. Chest tuberculosis: Radiological review and imaging recommendations. Indian Journal of Radiology and Imaging. 2015;25(03):213-225. doi:10.4103/0971-3026.161431
74. Brown C, Chauhan J, Grammenos A, et al. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. Published online 2020. doi:10.48550/ARXIV.2006.05919
75. Banerjee A, Nilhani A. A Residual Network based Deep Learning Model for Detection of COVID-19 from Cough Sounds. Published online 2021. doi:10.48550/ARXIV.2106.02348
76. Ghrabli S, Elgendi M, Menon C. Identifying unique spectral fingerprints in cough sounds for diagnosing respiratory ailments. Sci Rep. 2024;14(1):593. doi:10.1038/s41598-023-50371-2
77. Zhang J, Chen S, Wang J, et al. Research on the Identification Method of Respiratory Tract Infections Based on Multi-parameter Fusion. Published online May 31, 2024. doi:10.21203/rs.3.rs-4430123/v1
78. Huddart S, Yadav V, Sieberts SK, et al. Solicited Cough Sound Analysis for Tuberculosis Triage Testing: The CODA TB DREAM Challenge Dataset. Published online March 28, 2024. doi:10.1101/2024.03.27.24304980
79. Orlandic L, Teijeiro T, Atienza D. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. Sci Data. 2021;8(1):156. doi:10.1038/s41597-021-00937-4

80. Bhattacharya D, Sharma NK, Dutta D, et al. Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection. Sci Data. 2023;10(1):397. doi:10.1038/s41597-023-02266-0
81. Xiong H, Berkovsky S, Kâafar MA, Jaffe A, Coiera E, Sharan RV. Reliability of crowdsourced data and patient-reported outcome measures in cough-based COVID-19 screening. Sci Rep. 2022;12(1):21990. doi:10.1038/s41598-022-26492-5
82. Liu J, Li J, Li W, Wu J. Rethinking big data: A review on the data quality and usage issues. ISPRS Journal of Photogrammetry and Remote Sensing. 2016;115:134-142. doi:10.1016/j.isprsjprs.2015.11.006
83. Coppock H, Akman A, Bergler C, et al. A summary of the ComParE COVID-19 challenges. Front Digit Health. 2023;5:1058163. doi:10.3389/fdgth.2023.1058163
84. Pizzo DT, Esteban S. IATos: AI-powered pre-screening tool for COVID-19 from cough audio samples. Published online 2021. doi:10.48550/ARXIV.2104.13247
85. Cohen-McFarlane M, Goubran R, Knoefel F. Novel Coronavirus Cough Database: NoCoCoDa. IEEE Access. 2020;8:154087-154094. doi:10.1109/ACCESS.2020.3018028
86. Zarkogianni K, Dervakos E, Filandrianos G, et al. The smarty4covid dataset and knowledge base as a framework for interpretable physiological audio data analysis. Sci Data. 2023;10(1):770. doi:10.1038/s41597-023-02646-6
87. Chaudhari G, Jiang X, Fakhry A, et al. Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough. Published online 2020. doi:10.48550/ARXIV.2011.13320
88. Galvosas M, Gabaldón-Figueira JC, Keen EM, et al. Performance evaluation of the smartphone-based AI cough monitoring app - Hyfe Cough Tracker against solicited respiratory sounds. F1000Res. 2023;11:730. doi:10.12688/f1000research.122597.2
89. Sanchez-Morillo D, Sales-Lerida D, Priego-Torres B, León-Jiménez A. Cough Detection Using Acceleration Signals and Deep Learning Techniques. Electronics. 2024;13(12):2410. doi:10.3390/electronics13122410
90. Muguli A, Pinto L, R. N, et al. DiCOVA Challenge: Dataset, task, and baseline system for COVID-19 diagnosis using acoustics. Published online 2021. doi:10.48550/ARXIV.2103.09148
91. Bagad P, Dalmia A, Doshi J, et al. Cough Against COVID: Evidence of COVID-19 Signature in Cough Sounds. Published online 2020. doi:10.48550/ARXIV.2009.08790
92. Imran A, Posokhova I, Qureshi HN, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Informatics in Medicine Unlocked. 2020;20:100378. doi:10.1016/j.imu.2020.100378
93. Tena A, Clarià F, Solsona F. Automated detection of COVID-19 cough. Biomedical Signal Processing and Control. 2022;71:103175. doi:10.1016/j.bspc.2021.103175
94. Ponomarchuk A, Burenko I, Malkin E, et al. Project Achoo: A Practical Model and Application for COVID-19 Detection From Recordings of Breath, Voice, and Cough. IEEE J Sel Top Signal Process. 2022;16(2):175-187. doi:10.1109/JSTSP.2022.3142514
95. Chowdhury NK, Kabir MA, Rahman MdM, Islam SMS. Machine learning for detecting COVID-19 from cough sounds: An ensemble-based MCDM method. Computers in Biology and Medicine. 2022;145:105405. doi:10.1016/j.compbiomed.2022.105405

96. Hamdi S, Oussalah M, Moussaoui A, Saidi M. Attention-based hybrid CNN-LSTM and spectral data augmentation for COVID-19 diagnosis from cough sound. J Intell Inf Syst. 2022;59(2):367-389. doi:10.1007/s10844-022-00707-7
97. Kapoor T, Pandhi T, Gupta B. Cough Audio Analysis for COVID-19 Diagnosis. SN COMPUT SCI. 2022;4(2):125. doi:10.1007/s42979-022-01522-1
98. Melek M. Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound. Neural Comput & Applic. 2021;33(24):17621-17632. doi:10.1007/s00521-021-06346-3
99. Nguyen LH, Pham NT, Do VH, et al. Fruit-CoV: An efficient vision-based framework for speedy detection and diagnosis of SARS-CoV-2 infections through recorded cough sounds. Expert Systems with Applications. 2023;213:119212. doi:10.1016/j.eswa.2022.119212
100. Nguyen-Trong K, Nguyen-Hoang K. Multi-modal approach for COVID-19 detection using coughs and self-reported symptoms. IFS. 2023;44(3):3501-3513. doi:10.3233/JIFS-222863
101. Ren Z, Chang Y, Bartl-Pokorny KD, Pokorny FB, Schuller BW. The Acoustic Dissection of Cough: Diving Into Machine Listening-based COVID-19 Analysis and Detection. Journal of Voice. Published online June 2022:S0892199722001667. doi:10.1016/j.jvoice.2022.06.011
102. Sunitha G, Arunachalam R, Abd-Elnaby M, Eid MMA, Rashed ANZ. A comparative analysis of deep neural network architectures for the dynamic diagnosis of COVID-19 based on acoustic cough features. Int J Imaging Syst Tech. 2022;32(5):1433-1446. doi:10.1002/ima.22749
103. Akgün D, Kabakuş AT, Karapinar Şentürk Z, Şentürk A, Küçükkülahli E. A transfer learning-based deep learning approach for automated COVID-19 diagnosis with audio data. Turk J Elec Eng & Comp Sci. 2021;29(SI-1):2807-2823. doi:10.3906/elk-2105-64
104. Islam R, Abdel-Raheem E, Tarique M. A study of using cough sounds and deep neural networks for the early detection of Covid-19. Biomedical Engineering Advances. 2022;3:100025. doi:10.1016/j.bea.2022.100025
105. Najaran MHT. An evolutionary ensemble learning for diagnosing COVID-19 via cough signals. Intelligent Medicine. 2023;3(3):200-212. doi:10.1016/j.imed.2023.01.001
106. Ulukaya S, Sarıca AA, Erdem O, Karaali A. MSCCov19Net: multi-branch deep learning model for COVID-19 detection from cough sounds. Med Biol Eng Comput. 2023;61(7):1619-1629. doi:10.1007/s11517-023-02803-4
107. Hemdan EED, El-Shafai W, Sayed A. CR19: a framework for preliminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications. J Ambient Intell Human Comput. 2023;14(9):11715-11727. doi:10.1007/s12652-022-03732-0
108. Alaaeldein O, El Ahl OS, Elmahy L, Ihab M, Gomaa W. Analysis of COVID-19 Using Imaging and Audio Modalities. In: 2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE; 2022:1-10. doi:10.1109/IMCOM53663.2022.9721730
109. Elrahman SMA, Abraham A. A Review of Class Imbalance Problem. Journal of Network and Innovative Computing. 2013;1:9-9.

110. Pahar M, Klopper M, Warren R, Niesler T. COVID-19 cough classification using machine learning and global smartphone recordings. Computers in Biology and Medicine. 2021;135:104572. doi:10.1016/j.compbiomed.2021.104572
111. Akman A, Coppock H, Gaskell A, Tzirakis P, Jones L, Schuller BW. Evaluating the COVID-19 Identification ResNet (CIdeR) on the INTERSPEECH COVID-19 From Audio Challenges. Front Digit Health. 2022;4:789980. doi:10.3389/fdgth.2022.789980
112. Awais M, Bhuva A, Bhuva D, Fatima S, Sadiq T. Optimized DEC: An effective cough detection framework using optimal weighted Features-aided deep Ensemble classifier for COVID-19. Biomedical Signal Processing and Control. 2023;86:105026. doi:10.1016/j.bspc.2023.105026
113. Celik G. CovidCoughNet: A new method based on convolutional neural networks and deep feature extraction using pitch-shifting data augmentation for covid-19 detection from cough, breath, and voice signals. Computers in Biology and Medicine. 2023;163:107153. doi:10.1016/j.compbiomed.2023.107153
114. Chang Y, Jing X, Ren Z, Schuller BW. CovNet: A Transfer Learning Framework for Automatic COVID-19 Detection From Crowd-Sourced Cough Sounds. Front Digit Health. 2022;3:799067. doi:10.3389/fdgth.2021.799067
115. Despotovic V, Ismael M, Cornil M, Call RM, Fagherazzi G. Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results. Computers in Biology and Medicine. 2021;138:104944. doi:10.1016/j.compbiomed.2021.104944
116. Elizalde B, Tompkins D. COVID-19 Detection Using Recorded Coughs in the 2021 DiCOVA Challenge. Published online 2021. doi:10.48550/ARXIV.2105.10619
117. Erdoğan YE, Narin A. COVID-19 detection with traditional and deep features on cough acoustic signals. Computers in Biology and Medicine. 2021;136:104765. doi:10.1016/j.compbiomed.2021.104765
118. Hoang T, Pham L, Ngo D, Nguyen HD. A Cough-based deep learning framework for detecting COVID-19. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2022:3422-3425. doi:10.1109/EMBC48229.2022.9871179
119. Kamble MR, Gonzalez-Lopez JA, Grau T, et al. PANACEA cough sound-based diagnosis of COVID-19 for the DiCOVA 2021 Challenge. Published online 2021. doi:10.48550/ARXIV.2106.04423
120. Laguarta J, Hueto F, Subirana B. COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings. IEEE Open J Eng Med Biol. 2020;1:275-281. doi:10.1109/OJEMB.2020.3026928
121. Lella KK, Pja A. Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath. Alexandria Engineering Journal. 2022;61(2):1319-1334. doi:10.1016/j.aej.2021.06.024
122. Liu S, Mallol-Ragolta A, Schuller BW. COVID-19 Detection with a Novel Multi-Type Deep Fusion Method using Breathing and Coughing Information. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE; 2021:1840-1843. doi:10.1109/EMBC46164.2021.9630050

123. Loey M, Mirjalili S. COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models. Computers in Biology and Medicine. 2021;139:105020. doi:10.1016/j.compbiomed.2021.105020
124. Mitrofanova A, Mikhaylov D, Shaznaev I, Chumanskaia V, Saveliev V. Acoustery System for Differential Diagnosing of Coronavirus COVID-19 Disease. IEEE Open J Eng Med Biol. 2021;2:299-303. doi:10.1109/OJEMB.2021.3127078
125. Mohammed EA, Keyhani M, Sanati-Nezhad A, Hejazi SH, Far BH. An ensemble learning approach to digital corona virus preliminary screening from cough sounds. Sci Rep. 2021;11(1):15404. doi:10.1038/s41598-021-95042-2
126. Mouawad P, Dubnov T, Dubnov S. Robust Detection of COVID-19 in Cough Sounds: Using Recurrence Dynamics and Variable Markov Model. SN COMPUT SCI. 2021;2(1):34. doi:10.1007/s42979-020-00422-6
127. Askari Nasab K, Mirzaei J, Zali A, Gholizadeh S, Akhlaghdoust M. Coronavirus diagnosis using cough sounds: Artificial intelligence approaches. Front Artif Intell. 2023;6:1100112. doi:10.3389/frai.2023.1100112
128. Rahman T, Ibtehaz N, Khandakar A, et al. QUCoughScope: An Intelligent Application to Detect COVID-19 Patients Using Cough and Breath Sounds. Diagnostics. 2022;12(4):920. doi:10.3390/diagnostics12040920
129. Nessiem MA, Mohamed MM, Coppock H, Gaskell A, Schuller BW. Detecting COVID-19 from Breathing and Coughing Sounds using Deep Neural Networks. In: 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2021:183-188. doi:10.1109/CBMS52027.2021.00069
130. Sobahi N, Atila O, Deniz E, Sengur A, Acharya UR. Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds. Biocybernetics and Biomedical Engineering. 2022;42(3):1066-1080. doi:10.1016/j.bbe.2022.08.005
131. Hope TMH. Linear regression. In: Machine Learning. Elsevier; 2020:67-81. doi:10.1016/B978-0-12-815739-8.00004-3
132. Liang D, Frederick DA, Lledo EE, et al. Examining the utility of nonlinear machine learning approaches versus linear regression for predicting body image outcomes: The U.S. Body Project I. Body Image. 2022;41:32-45. doi:10.1016/j.bodyim.2022.01.013
133. Peng Y, Nagata MH. An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. Chaos, Solitons & Fractals. 2020;139:110055. doi:10.1016/j.chaos.2020.110055
134. Dietterich TG. Ensemble Methods in Machine Learning. In: Multiple Classifier Systems. Vol 1857. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2000:1-15. doi:10.1007/3-540-45014-9_1
135. Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks. 2015;61:85-117. doi:10.1016/j.neunet.2014.09.003
136. O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. Published online 2015. doi:10.48550/ARXIV.1511.08458
137. Cabitza F, Campagner A, Soares F, et al. The importance of being external. methodological insights for the external validation of machine learning models in medicine. Computer

Methods and Programs in Biomedicine. 2021;208:106288. doi:10.1016/j.cmpb.2021.106288

138. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. Nat Med. 2023;29(11):2686-2687. doi:10.1038/s41591-023-02540-z
139. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. JAMA Intern Med. 2021;181(8):1065. doi:10.1001/jamainternmed.2021.2626
140. Habib AR, Lin AL, Grant RW. The Epic Sepsis Model Falls Short—The Importance of External Validation. JAMA Intern Med. 2021;181(8):1040. doi:10.1001/jamainternmed.2021.3333
141. Khine M. Editorial: Clinical validation of digital health technologies for personalized medicine. Front Digit Health. 2022;4:831517. doi:10.3389/fdgth.2022.831517
142. Makri A. Bridging the digital divide in health care. The Lancet Digital Health. 2019;1(5):e204-e205. doi:10.1016/S2589-7500(19)30111-6
143. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and Facilitators of Digital Health Among Culturally and Linguistically Diverse Populations: Qualitative Systematic Review. J Med Internet Res. 2023;25:e42719. doi:10.2196/42719
144. World Health Organization. Global Tuberculosis Report 2020. World Health Organization (WHO); 2020. https://www.who.int/publications/i/item/9789240013131
145. Zimmer AJ, Klinton JS, Oga-Omenka C, et al. Tuberculosis in times of COVID-19. J Epidemiol Community Health. 2022;76(3):310-316. doi:10.1136/jech-2021-217529
146. Pai M, Behr MA, Dowdy D, et al. Tuberculosis. Nat Rev Dis Primers. 2016;2(1):16076. doi:10.1038/nrdp.2016.76
147. Sharma SK, Mohan A, Kohli M. Extrapulmonary tuberculosis. Expert Review of Respiratory Medicine. 2021;15(7):931-948. doi:10.1080/17476348.2021.1927718
148. Vongthilath-Moeung R, Poncet A, Renzi G, Schrenzel J, Janssens JP. Time to Detection of Growth for Mycobacterium tuberculosis in a Low Incidence Area. Front Cell Infect Microbiol. 2021;11:704169. doi:10.3389/fcimb.2021.704169
149. Kik SV, Denkinger CM, Chedore P, Pai M. Replacing smear microscopy for the diagnosis of tuberculosis: what is the market potential? Eur Respir J. 2014;43(6):1793-1796. doi:10.1183/09031936.00217313
150. Steingart KR, Schiller I, Horne DJ, Pai M, Boehme CC, Dendukuri N. Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Infectious Diseases Group, ed. Cochrane Database of Systematic Reviews. Published online January 21, 2014. doi:10.1002/14651858.CD009593.pub3
151. Acharya B, Acharya A, Gautam S, et al. Advances in diagnosis of Tuberculosis: an update into molecular diagnosis of Mycobacterium tuberculosis. Mol Biol Rep. 2020;47(5):4065-4075. doi:10.1007/s11033-020-05413-7
152. Nahid P, Dorman SE, Alipanah N, et al. Executive Summary: Official American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America Clinical Practice Guidelines: Treatment of Drug-Susceptible Tuberculosis. Clinical Infectious Diseases. 2016;63(7):853-867. doi:10.1093/cid/ciw566

153. Forget EJ, Menzies D. Adverse reactions to first-line antituberculosis drugs. Expert Opinion on Drug Safety. 2006;5(2):231-249. doi:10.1517/14740338.5.2.231
154. Proaño A, Bui DP, López JW, et al. Cough Frequency During Treatment Associated With Baseline Cavitary Volume and Proximity to the Airway in Pulmonary TB. Chest. 2018;153(6):1358-1367. doi:10.1016/j.chest.2018.03.006
155. Proaño A, Bravard MA, López JW, et al. Dynamics of Cough Frequency in Adults Undergoing Treatment for Pulmonary Tuberculosis. Clinical Infectious Diseases. 2017;64(9):1174-1181. doi:10.1093/cid/cix039
156. Lee GO, Comina G, Hernandez-Cordova G, et al. Cough dynamics in adults receiving tuberculosis treatment. Ehtesham HS, ed. PLoS ONE. 2020;15(6):e0231167. doi:10.1371/journal.pone.0231167
157. Williams CM, Abdulwhhab M, Birring SS, et al. Exhaled Mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: prospective observational studies. The Lancet Infectious Diseases. 2020;20(5):607-617. doi:10.1016/S1473-3099(19)30707-8
158. World Health Organization. High-Priority Target Product Profiles for New Tuberculosis Diagnostics.; 2014. https://www.who.int/publications/i/item/WHO-HTM-TB-2014.18
159. Cho PSP, Birring SS, Fletcher HV, Turner RD. Methods of Cough Assessment. The Journal of Allergy and Clinical Immunology: In Practice. 2019;7(6):1715-1723. doi:10.1016/j.jaip.2019.01.049
160. Feikin DR, Audi A, Olack B, et al. Evaluation of the optimal recall period for disease symptoms in home-based morbidity surveillance in rural and urban Kenya. International Journal of Epidemiology. 2010;39(2):450-458. doi:10.1093/ije/dyp374
161. Helm JM, Swiergosz AM, Haeberle HS, et al. Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. Curr Rev Musculoskelet Med. 2020;13(1):69-76. doi:10.1007/s12178-020-09600-8
162. Suzuki T, Shirai T, Hirai K, et al. Improved cough- and sputum-related quality of life after initiation of treatment in pulmonary tuberculosis. Respiratory Investigation. 2019;57(3):252-259. doi:10.1016/j.resinv.2019.01.002
163. Turner R, Bothamley G, Birring S. P240 Validation of the Leicester Cough Questionnaire in pulmonary tuberculosis. Thorax. 2015;70(Suppl 3):A197.2-A198. doi:10.1136/thoraxjnl-2015-207770.376
164. Birring SS. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). Thorax. 2003;58(4):339-343. doi:10.1136/thorax.58.4.339
165. Loudon RG, Spohn SK. Cough frequency and infectivity in patients with pulmonary tuberculosis. Am Rev Respir Dis. 1969;99(1):109-111. doi:10.1164/arrd.1969.99.1.109
166. Hsu J, Stone R, Logan-Sinclair R, Worsdell M, Busst C, Chung K. Coughing frequency in patients with persistent cough: assessment using a 24 hour ambulatory recorder. Eur Respir J. 1994;7(7):1246-1253. doi:10.1183/09031936.94.07071246
167. Paul IM, Wai K, Jewell SJ, Shaffer ML, Varadan VV. Evaluation of a new self-contained, ambulatory, objective cough monitor. Cough. 2006;2:7. doi:10.1186/1745-9974-2-7

168. Hall JI, Lozano M, Estrada-Petrocelli L, Birring S, Turner R. The present and future of cough counting tools. J Thorac Dis. 2020;12(9):5207-5223. doi:10.21037/jtd-2020-icc-003
169. Smith J, Woodcock A. New Developments in the Objective Assessment of Cough. Lung. 2008;186(S1):48-54. doi:10.1007/s00408-007-9059-1
170. Matos S, Birring SS, Pavord ID, Evans DH. An Automated System for 24-h Monitoring of Cough Frequency: The Leicester Cough Monitor. IEEE Trans Biomed Eng. 2007;54(8):1472-1479. doi:10.1109/TBME.2007.900811
171. Birring SS, Fleming T, Matos S, Raj AA, Evans DH, Pavord ID. The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough. European Respiratory Journal. 2008;31(5):1013-1018. doi:10.1183/09031936.00057407
172. Turner R, Repossi A, Matos S, Birring S, Bothamley G. S79 Cough Prevalence And Frequency In Pulmonary Tuberculosis. Thorax. 2014;69(Suppl 2):A43-A44. doi:10.1136/thoraxjnl-2014-206260.85
173. Tracey BH, Comina G, Larson S, Bravard M, Lopez JW, Gilman RH. Cough detection algorithm for monitoring patient recovery from pulmonary tuberculosis. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2011:6017-6020. doi:10.1109/IEMBS.2011.6091487
174. Larson S, Comina G, Gilman RH, Tracey BH, Bravard M, López JW. Validation of an Automated Cough Detection Algorithm for Tracking Recovery of Pulmonary Tuberculosis Patients. Pai M, ed. PLoS ONE. 2012;7(10):e46229. doi:10.1371/journal.pone.0046229
175. Cremers AL, De Laat MM, Kapata N, Gerrets R, Klipstein-Grobusch K, Grobusch MP. Assessing the Consequences of Stigma for Tuberculosis Patients in Urban Zambia. Kumar A, ed. PLoS ONE. 2015;10(3):e0119861. doi:10.1371/journal.pone.0119861
176. Gabaldon-Figueira JC, Brew J, Doré DH, et al. Digital acoustic surveillance for early detection of respiratory disease outbreaks in Spain: a protocol for an observational study. BMJ Open. 2021;11(7):e051278. doi:10.1136/bmjopen-2021-051278
177. Moschovis PP, Sampayo EM, Cook A, et al. The diagnosis of respiratory disease in children using a phone-based cough and symptom analysis algorithm: The smartphone recordings of cough sounds 2 (SMARTCOUGH-C 2) trial design. Contemporary Clinical Trials. 2021;101:106278. doi:10.1016/j.cct.2021.106278
178. Bales C, Nabeel M, John CN, et al. Can Machine Learning Be Used to Recognize and Diagnose Coughs? Published online 2020. doi:10.48550/ARXIV.2004.01495
179. Coppock H, Gaskell A, Tzirakis P, Baird A, Jones L, Schuller B. End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: a pilot study. BMJ Innov. 2021;7(2):356-362. doi:10.1136/bmjinnov-2021-000668
180. Cesnakova L, Boza V, Dubec P, et al. Continuous Sound Collection Using Smartphones and Machine Learning to Measure Cough. Digit Biomark. 2019;3(3):166-175. doi:10.1159/000504666
181. Pahar M, Klopper M, Reeve B, Warren R, Theron G, Niesler T. Automatic cough classification for tuberculosis screening in a real-world environment. Physiol Meas. 2021;42(10):105014. doi:10.1088/1361-6579/ac2fb8
182. Botha GHR, Theron G, Warren RM, et al. Detection of tuberculosis by automatic cough sound analysis. Physiol Meas. 2018;39(4):045005. doi:10.1088/1361-6579/aab6d0

183. Topol EJ. Is my cough COVID-19? The Lancet. 2020;396(10266):1874. doi:10.1016/S0140-6736(20)32589-7
184. Rutjes AWS, Reitsma JB, Di Nisio M, Smidt N, Van Rijn JC, Bossuyt PMM. Evidence of bias and variation in diagnostic accuracy studies. CMAJ. 2006;174(4):469-476. doi:10.1503/cmaj.050090
185. Subbaraman R, Nathavitharana RR, Mayer KH, et al. Constructing care cascades for active tuberculosis: A strategy for program monitoring and identifying gaps in quality of care. PLoS Med. 2019;16(2):e1002754. doi:10.1371/journal.pmed.1002754
186. Qin ZZ, Ahmed S, Sarker MS, et al. Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. The Lancet Digital Health. 2021;3(9):e543-e554. doi:10.1016/S2589-7500(21)00116-3
187. Satyanarayana S, Subbaraman R, Shete P, et al. Quality of tuberculosis care in India: a systematic review. Int J Tuberc Lung Dis. 2015;19(7):751-763. doi:10.5588/ijtld.15.0186
188. Das J, Kwan A, Daniels B, et al. Use of standardised patients to assess quality of tuberculosis care: a pilot, cross-sectional study. The Lancet Infectious Diseases. 2015;15(11):1305-1313. doi:10.1016/S1473-3099(15)00077-8
189. Nguyen TA, Pham MT, Nguyen TL, et al. Video Directly Observed Therapy to support adherence with treatment for tuberculosis in Vietnam: A prospective cohort study. International Journal of Infectious Diseases. 2017;65:85-89. doi:10.1016/j.ijid.2017.09.029
190. Velayutham B, Chadha VK, Singla N, et al. Recurrence of tuberculosis among newly diagnosed sputum positive pulmonary tuberculosis patients treated under the Revised National Tuberculosis Control Programme, India: A multi-centric prospective study. Cardona PJ, ed. PLoS ONE. 2018;13(7):e0200150. doi:10.1371/journal.pone.0200150
191. Meghji J, Simpson H, Squire SB, Mortimer K. A Systematic Review of the Prevalence and Pattern of Imaging Defined Post-TB Lung Disease. Hill PC, ed. PLoS ONE. 2016;11(8):e0161176. doi:10.1371/journal.pone.0161176
192. Davies G, Boeree M, Hermann D, Hoelscher M. Accelerating the transition of new tuberculosis drug combinations from Phase II to Phase III trials: New technologies and innovative designs. PLoS Med. 2019;16(7):e1002851. doi:10.1371/journal.pmed.1002851
193. Cruz Rivera S, McMullan C, Jones L, Kyte D, Slade A, Calvert M. The impact of patient-reported outcome data from clinical trials: perspectives from international stakeholders. J Patient Rep Outcomes. 2020;4(1):51. doi:10.1186/s41687-020-00219-4
194. Insmed Incorporated. A Randomized, Double-Blind, Placebo-Controlled, Active Comparator, Multicenter Study to Validate Patient-Reported Outcome Instruments in Adult Subjects With Newly Diagnosed Nontuberculous Mycobacterial (NTM) Lung Infection Caused by Mycobacterium Avium Complex (MAC). clinicaltrials.gov; 2021. https://clinicaltrials.gov/study/NCT04677543
195. Jaeger S, Candemir S, Antani S, Wáng YXJ, Lu PX, Thoma G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant Imaging Med Surg. 2014;4(6):475-477. doi:10.3978/j.issn.2223-4292.2014.11.20
196. The CRyPTIC Consortium and the 100,000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med. 2018;379(15):1403-1415. doi:10.1056/NEJMoa1800474

197. McGuinness K, Holt K, Dockry R, Smith J. P159 Validation of the VitaloJAKTM 24 Hour Ambulatory Cough Monitor: Abstract P159 Table 1. Thorax. 2012;67(Suppl 2):A131.1-A131. doi:10.1136/thoraxjnl-2012-202678.220
198. Rashid HA, Mazumder AN, Niyogi UPK, Mohsenin T. CoughNet: A Flexible Low Power CNN-LSTM Processor for Cough Sound Detection. In: 2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS). IEEE; 2021:1-4. doi:10.1109/AICAS51828.2021.9458509
199. Sharan RV. Cough sound detection from raw waveform using SincNet and bidirectional GRU. Biomedical Signal Processing and Control. 2023;82:104580. doi:10.1016/j.bspc.2023.104580
200. Alqudaihi KS, Aslam N, Khan IU, et al. Cough Sound Detection and Diagnosis Using Artificial Intelligence Techniques: Challenges and Opportunities. IEEE Access. 2021;9:102327-102344. doi:10.1109/ACCESS.2021.3097559
201. Keen EM, True EJ, Summers AR, Smith EC, Brew J, Grandjean Lapierre S. High-throughput digital cough recording on a university campus: A SARS-CoV-2-negative curated open database and operational template for acoustic screening of respiratory diseases. DIGITAL HEALTH. 2022;8:205520762210975. doi:10.1177/20552076221097513
202. Friedrich C, Francke K, Birring SS, et al. The P2X3 receptor antagonist filapixant in patients with refractory chronic cough: a randomized controlled trial. Respir Res. 2023;24(1):109. doi:10.1186/s12931-023-02384-8
203. Dutta P, Funston W, Mossop H, et al. Randomised, double-blind, placebo-controlled pilot trial of omeprazole in idiopathic pulmonary fibrosis. Thorax. 2019;74(4):346-353. doi:10.1136/thoraxjnl-2018-212102
204. Ludbrook VJ, Hanrott KE, Kreindler JL, et al. Adaptive study design to assess effect of TRPV4 inhibition in patients with chronic cough. ERJ Open Res. 2021;7(3):00269-02021. doi:10.1183/23120541.00269-2021
205. Turner RD, Birring SS. Measuring cough: what really matters? J Thorac Dis. 2023;15(4):2288-2299. doi:10.21037/jtd-23-230
206. Schmit KM, Coeytaux RR, Goode AP, et al. Evaluating Cough Assessment Tools. Chest. 2013;144(6):1819-1826. doi:10.1378/chest.13-0310
207. Tricco AC, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169(7):467-473. doi:10.7326/M18-0850
208. The Joanna Briggs Institute. The Joanna Briggs Institute Reviewers' Manual - Methodology for JBI Scoping Review.; 2015. https://reben.com.br/revista/wp-content/uploads/2020/10/Scoping.pdf
209. Zürcher K, Riou J, Morrow C, et al. Estimating Tuberculosis Transmission Risks in a Primary Care Clinic in South Africa: Modeling of Environmental and Clinical Data. The Journal of Infectious Diseases. 2022;225(9):1642-1652. doi:10.1093/infdis/jiab534
210. Bisballe-Müller N, Chang AB, Plumb EJ, Oguoma VM, Halken S, McCallum GB. Can Acute Cough Characteristics From Sound Recordings Differentiate Common Respiratory Illnesses in Children? Chest. 2021;159(1):259-269. doi:10.1016/j.chest.2020.06.067

211. Crooks MG, Hayman Y, Innes A, Williamson J, Wright CE, Morice AH. Objective Measurement of Cough Frequency During COPD Exacerbation Convalescence. Lung. 2016;194(1):117-120. doi:10.1007/s00408-015-9782-y
212. Doenges J, Kuckuck E, Cassel W, et al. Disease control in patients with asthma and respiratory symptoms (wheezing, cough) during sleep. asthma res and pract. 2020;6(1):9. doi:10.1186/s40733-020-00062-w
213. Elghamoudi D, McGuinness K, Smith J, Murray C. S73 Cough frequency and diurnal patterns in children with asthma. In: New Approaches to Characterising Paediatric Respiratory Diseases. BMJ Publishing Group Ltd and British Thoracic Society; 2017:A46-A47. doi:10.1136/thoraxjnl-2017-210983.79
214. Fischer P, Gross V, Kroenig J, et al. Description of nighttime cough epochs in patients with stable COPD GOLD II–IV. COPD. 2018;Volume 13:1071-1078. doi:10.2147/COPD.S154539
215. Fletcher H, Cho PSP, Matos S, et al. Chronic cough: objective cough frequency in a large cough clinic cohort. In: Clinical Problems Asthma. European Respiratory Society; 2017:OA4677. doi:10.1183/1393003.congress-2017.OA4677
216. Grosse-Onnebrink J, Radine A, Werner C, Omran H. Objectification of cough in cystic fibrosis. Journal of Cystic Fibrosis. 2016;15:S47. doi:10.1016/S1569-1993(16)30230-2
217. Harle ASM, Blackhall FH, Molassiotis A, et al. Cough in Patients With Lung Cancer: A Longitudinal Observational Study of Characterization and Clinical Associations. Chest. 2019;155(1):103-113. doi:10.1016/j.chest.2018.10.003
218. Hirai K, Enseki M, Tabata H, et al. Objective measurement of frequency and pattern of nocturnal cough in children with asthma exacerbation. Annals of Allergy, Asthma & Immunology. 2016;117(2):169-174. doi:10.1016/j.anai.2016.06.007
219. Hirai K, Nukaga M, Tabata H, et al. Objective measurement of nocturnal cough in infants with acute bronchiolitis. Respiratory Investigation. 2019;57(6):605-610. doi:10.1016/j.resinv.2019.06.005
220. Hirai K, Kaiga C, Otomo T, et al. Effect of Post-nasal Drip on Overnight-cough Frequency and Cough Pattern in Children with Asthma. Tokai J Exp Clin Med. 2022;47(2):79-84.
221. Holmes J, McGarvey LPA, Birring SS, Fletcher H, Heaney LG. An observational study to determine the relationship between cough frequency and markers of inflammation in severe asthma. Eur Respir J. 2022;60(6):2103205. doi:10.1183/13993003.03205-2021
222. Imai E, Hirai K, Mikami Y, et al. An objective evaluation of nocturnal cough count and cough pattern in children with psychogenic cough. Respiratory Investigation. 2017;55(6):334-337. doi:10.1016/j.resinv.2017.08.008
223. Key A, Holt K, Smith J, Walker P. Objective cough count in bronchiectasis patients is reproducible and relates to quality of life. In: Monitoring Airway Disease. European Respiratory Society; 2019:PA2672. doi:10.1183/13993003.congress-2019.PA2672
224. Koehler U, Hildebrandt O, Fischer P, et al. Time course of nocturnal cough and wheezing in children with acute bronchitis monitored by lung sound analysis. Eur J Pediatr. 2019;178(9):1385-1394. doi:10.1007/s00431-019-03426-4
225. Krönig J, Hildebrandt O, Weissflog A, et al. Long-term Recording of Night-Time Respiratory Symptoms in Patients with Stable COPD II–IV. COPD: Journal of Chronic

Obstructive Pulmonary Disease. 2017;14(5):498-503. doi:10.1080/15412555.2017.1338681

226. Kulnik S, Birring S, Rafferty G, Moxham J, Kalra L. P237 Cough frequency in acute stroke: Abstract P237 Table 1. Thorax. 2015;70(Suppl 3):A196.1-A196. doi:10.1136/thoraxjnl-2015-207770.373
227. Lindenhofer M, Roth L, Mädel C, et al. Wheeze and cough measurements at night in children with respiratory symptoms. BMC Pediatr. 2020;20(1):556. doi:10.1186/s12887-020-02455-5
228. Lodhi S, Smith JA, Satia I, Holt KJ, Maidstone RJ, Durrington HJ. Cough rhythms in asthma: Potential implication for management. The Journal of Allergy and Clinical Immunology: In Practice. 2019;7(6):2024-2027. doi:10.1016/j.jaip.2018.12.020
229. Mackay A, Donaldson G, Singh R, et al. Utility of Objective Cough Monitoring in Stable COPD and at Exacerbation. American Thoracic Society. Published online 2015:A2539-A2539.
230. Ovsyannikov E, Avdeev S, Budnevsky A, Shkatova Y. Influence of Anxiety/Depression on the Subjective Evaluation of Cough in Patients with Chronic Obstructive Pulmonary Disease and Obesity. Medicina. 2019;55(5):134. doi:10.3390/medicina55050134
231. Radine A, Werner C, Raidt J, et al. Comparison of Nocturnal Cough Analysis in Healthy Subjects and in Patients with Cystic Fibrosis and Primary Ciliary Dyskinesia: A Prospective Observational Study. Respiration. 2019;97(1):60-69. doi:10.1159/000493323
232. Rassouli F, Tinschert P, Barata F, et al. Characteristics of Asthma-related Nocturnal Cough: A Potential New Digital Biomarker. JAA. 2020;Volume 13:649-657. doi:10.2147/JAA.S278119
233. Rhee H, Belyea MJ, Sterling M, Bocko MF. Evaluating the Validity of an Automated Device for Asthma Monitoring for Adolescents: Correlational Design. J Med Internet Res. 2015;17(10):e234. doi:10.2196/jmir.4975
234. Schwarz SB, Windisch W, Majorski DS, Callegari J, Pläcking M, Magnet FS. Long-Term Auscultation in Chronic Obstructive Pulmonary Disease: Renaissance of an Ideograph of Medical Care. Respiration. 2021;100(3):201-208. doi:10.1159/000513439
235. Sinha A, Lee KK, Rafferty GF, et al. Predictors of objective cough frequency in pulmonary sarcoidosis. Eur Respir J. 2016;47(5):1461-1471. doi:10.1183/13993003.01369-2015
236. Spinou A, Lee KK, Sinha A, et al. The Objective Assessment of Cough Frequency in Bronchiectasis. Lung. 2017;195(5):575-585. doi:10.1007/s00408-017-0038-x
237. Sumner H, Woodcock A, Kolsum U, et al. Predictors of Objective Cough Frequency in Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med. 2013;187(9):943-949. doi:10.1164/rccm.201211-2000OC
238. Turner R, Birring S, Matos S, Bothamley G. S29 Cough frequency and morbidity in inpatients with acute respiratory disease. Thorax. 2013;68(Suppl 3):A17.2-A18. doi:10.1136/thoraxjnl-2013-204457.36
239. Turner RD, Birring SS, Darmalingam M, et al. Daily cough frequency in tuberculosis and association with household infection. Int J Tuberc Lung Dis. 2018;22(8):863-870. doi:10.5588/ijtld.17.0652

240. Vertigan AE, Kapela SM, Kearney EK, Gibson PG. Laryngeal Dysfunction in Cough Hypersensitivity Syndrome: A Cross-Sectional Observational Study. J. Allergy Clin. Immunol.: In Practice. 2018;6(6):2087-2095. doi:10.1016/j.jaip.2018.04.015
241. Vertigan AE, Kapela SL, Gibson PG. Chronic cough in Vocal Cord Dysfunction: Description of a clinical entity. Respir. Med. 2020;168:105990. doi:10.1016/j.rmed.2020.105990
242. Wang K, Bettiol S, Thompson MJ, et al. Symptomatic treatment of the cough in whooping cough. Cochrane Acute Respiratory Infections Group, ed. Cochrane Database of Systematic Reviews. Published online September 22, 2014. doi:10.1002/14651858.CD003257.pub5
243. Weisser S, Lex C, Roth L, et al. Evaluation of nocturnal asthma control (NAC-study): Comparative multicenter analysis of subjective and objective methods. In: 07.02 - Paediatric Asthma and Allergy. ERS; 2022:1588. doi:10.1183/13993003.congress-2022.1588
244. Winders T, Nemala S, De Gardner D, Stempel D. Novel Technology to Identify Times of Peak Nighttime Cough Frequency in Severe Asthma. Journal of Allergy and Clinical Immunology. 2023;151(2):AB334. doi:10.1016/j.jaci.2022.12.775
245. Yousaf N, Monteiro W, Matos S, Birring SS, Pavord ID. Cough frequency in health and disease. Eur. Respir. J. 2013;41(1):241-243. doi:10.1183/09031936.00089312
246. Faruqi S, Shiferaw D, Morice AH. Effect of Ivacaftor on Objective and Subjective Measures of Cough in Patients with Cystic Fibrosis. TORMJ. 2016;10(1):105-108. doi:10.2174/1874306401610010105
247. Faruqi S, Sykes DL, Crooks MG, Brindle K, Thompson J, Morice AH. Objective Assessment of Cough: An Early Marker of Response to Biological Therapies in Asthma? Lung. 2020;198(5):767-770. doi:10.1007/s00408-020-00391-w
248. Jung J, Lee K, Song C, Sun Park J. ACC2308. Exploration of Cough Counts and Epochs in Patients with Asthma and Interstitial Lung Disease. Lung. 2023;201(S1):1-6. doi:10.1007/s00408-023-00634-6
249. Kang YR, Oh JY, Lee JH, Small PM, Chung KF, Song WJ. Long-COVID severe refractory cough: discussion of a case with 6-week longitudinal cough characterization. Asia Pacific Allergy. 2022;12(2):e19. doi:10.5415/apallergy.2022.12.e19
250. Lee SE, Rudd M, Kim TH, et al. Feasibility and Utility of a Smartphone Application-Based Longitudinal Cough Monitoring in Chronic Cough Patients in a Real-World Setting. Lung. 2023;201(6):555-564. doi:10.1007/s00408-023-00647-1
251. Shim JS, Kim BK, Kim SH, et al. A smartphone-based application for cough counting in patients with acute asthma exacerbation. J Thorac Dis. 2023;15(7):4053-4065. doi:10.21037/jtd-22-1492
252. Turner R, Birring S, Bothamley G. S37 The response of objectively-measured cough to treatment in tuberculosis: an exploratory study. Thorax. 2015;70(Suppl 3):A25.2-A26. doi:10.1136/thoraxjnl-2015-207770.43
253. Vertigan AE, Kapela SL, Birring SS, Gibson PG. Feasibility and clinical utility of ambulatory cough monitoring in an outpatient clinical setting: a real-world retrospective evaluation. ERJ Open Res. 2021;7(4):00319-02021. doi:10.1183/23120541.00319-2021

254. Zhang M, Brindle K, Robinson M, Ingram D, Cavany T, Morice A. Chronic cough in cystic fibrosis: the effect of modulator therapy on objective 24-h cough monitoring. ERJ Open Res. 2022;8(2):00031-02022. doi:10.1183/23120541.00031-2022
255. Altshuler E, Tannir B, Jolicoeur G, et al. Digital cough monitoring – A potential predictive acoustic biomarker of clinical outcomes in hospitalized COVID-19 patients. Journal of Biomedical Informatics. 2023;138:104283. doi:10.1016/j.jbi.2023.104283
256. Boesch M, Rassouli F, Baty F, et al. Smartphone-based cough monitoring as a near real-time digital pneumonia biomarker. ERJ Open Res. 2023;9(3):00518-02022. doi:10.1183/23120541.00518-2022
257. Crooks MG, Den Brinker AC, Thackray-Nocera S, Van Dinther R, Wright CE, Morice AH. Domiciliary Cough Monitoring for the Prediction of COPD Exacerbations. Lung. 2021;199(2):131-137. doi:10.1007/s00408-021-00435-9
258. Den Brinker AC, Van Dinther R, Crooks MG, Thackray-Nocera S, Morice AH. Alert system design based on experimental findings from long-term unobtrusive monitoring in COPD. Biomed. Signal Process. Control. 2021;63:102205. doi:10.1016/j.bspc.2020.102205
259. Pekacka-Egli AM, Kazmierski R, Lutz D, et al. Predictive Value of Cough Frequency in Addition to Aspiration Risk for Increased Risk of Pneumonia in Dysphagic Stroke Survivors: A Clinical Pilot Study. Brain Sciences. 2021;11(7):847. doi:10.3390/brainsci11070847
260. Tinschert P, Rassouli F, Barata F, et al. Nocturnal Cough and Sleep Quality to Assess Asthma Control and Predict Attacks. JAA. 2020;Volume 13:669-678. doi:10.2147/JAA.S278155
261. Al Hossain F, Lover AA, Corey GA, Reich NG, Rahman T. FluSense: A Contactless Syndromic Surveillance Platform for Influenza-Like Illness in Hospital Waiting Areas. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2020;4(1):1-28. doi:10.1145/3381014
262. Gabaldón-Figueira JC, Keen E, Giménez G, et al. Acoustic surveillance of cough for detecting respiratory disease using artificial intelligence. ERJ Open Res. 2022;8(2):00053-02022. doi:10.1183/23120541.00053-2022
263. Rahman T, Hossain F, Tonmoy TH, et al. Passive Monitoring of Crowd-Level Cough Counts in Waiting Areas Produces Reliable Syndromic Indicator for Total COVID-19 Burden in a Hospital Emergency Clinic. In Review; 2023. doi:10.21203/rs.3.rs-3084318/v1
264. Deblej Elghamoudi D, Sumner H, McGuiness K, Smith J, Murray C. P241 The feasibility and validity of objective cough monitoring in children using an adult cough detection system. Thorax. 2015;70(Suppl 3):A198.1-A198. doi:10.1136/thoraxjnl-2015-207770.377
265. Gross V, Scholtes M, Sohrabi K, Weissflog A. Validation of the LEOSound®-monitor for standardized detection of cough and wheezing in children. In: World Sleep. 2015. http://www.congrex-switzerland.com/fileadmin/files/2015/worldsleep2015/pdf/Worldsleep_2015_-_Online_Abstract_Book_by_session__date_and_time.pdf
266. Huddart S, Asege L, Dang H, et al. Feasibility and Analysis Considerations of Continuous Community-Based Cough Monitoring in Low- and Middle-Income Settings. Public and Global Health; 2022. doi:10.1101/2022.07.06.22277224

267. Rhee H, Miner S, Sterling M, Halterman JS, Fairbanks E. The Development of an Automated Device for Asthma Monitoring for Adolescents: Methodologic Approach and User Acceptability. JMIR mHealth uHealth. 2014;2(2):e27. doi:10.2196/mhealth.3118
268. Urban C, Kiefer A, Conradt R, et al. Validation of the LEOSound® monitor for standardized detection of wheezing and cough in children. Pediatric Pulmonology. 2022;57(2):551-559. doi:10.1002/ppul.25768
269. Shim J, Kim M, Lee SM, et al. An artificial intelligence algorithm-based smartphone application for daily cough monitoring. Allergy. 2023;78(5):1378-1380. doi:10.1111/all.15632
270. Zimmer AJ, Lainati F, Aguilera Vasquez N, et al. Biomarkers That Correlate with Active Pulmonary Tuberculosis Treatment Response: a Systematic Review and Meta-analysis. McAdam AJ, ed. J Clin Microbiol. 2022;60(2):e01859-21. doi:10.1128/jcm.01859-21
271. Ulrik CS, Ringbæk T, Laursen LC, Green A, Brøndum E, Frausing E. Effect of tele health care on exacerbations and hospital admissions in patients with chronic obstructive pulmonary disease: a randomized clinical trial. COPD. Published online September 2015:1801. doi:10.2147/COPD.S85596
272. Al Rajeh A, Steiner MC, Aldabayan Y, et al. Use, utility and methods of telehealth for patients with COPD in England and Wales: a healthcare provider survey. BMJ Open Resp Res. 2019;6(1):e000345. doi:10.1136/bmjresp-2018-000345
273. Anguera Miro X, Bozonnet S, Evans N, Fredouille C, Friedland G, Vinyals O. Speaker Diarization: A Review of Recent Research. IEEE Trans Audio Speech Lang Process. 2012;20(2):356-370. doi:10.1109/TASL.2011.2125954
274. Santosh K, Rasmussen N, Mamun M, Aryal S. A systematic review on cough sound analysis for Covid-19 diagnosis and screening: is my cough sound COVID-19? PeerJ Computer Science. 2022;8:e958. doi:10.7717/peerj-cs.958
275. Ming DK, Sangkaew S, Chanh HQ, et al. Continuous physiological monitoring using wearable technology to inform individual management of infectious diseases, public health and outbreak responses. International Journal of Infectious Diseases. 2020;96:648-654. doi:10.1016/j.ijid.2020.05.086
276. Gauthier-Beaupré A, Grosjean S. Understanding acceptability of digital health technologies among francophone-speaking communities across the world: a meta-ethnographic study. Front Commun. 2023;8:1230015. doi:10.3389/fcomm.2023.1230015
277. Turner RD, Bothamley GH. Cough and the Transmission of Tuberculosis. The Journal of Infectious Diseases. 2015;211(9):1367-1372. doi:10.1093/infdis/jiu625
278. Stadler J a. M. Updated WHO definitions for tuberculosis outcomes: Simplified, unified and future-proofed. Afr J Thorac Crit Care Med. 2022;28(2). doi:10.7196/AJTCCM.2022.v28i2.224
279. Horne DJ, Royce SE, Gooze L, et al. Sputum monitoring during tuberculosis treatment for predicting outcome: systematic review and meta-analysis. The Lancet Infectious Diseases. 2010;10(6):387-394. doi:10.1016/S1473-3099(10)70071-2
280. Gupta-Wright A, Den Boon S, MacLean E, et al. Target product profiles: tests for tuberculosis treatment monitoring and optimization. Bull World Health Org. 2023;101(11):730-737. doi:10.2471/BLT.23.290901

281. Zimmer AJ, Lainati F, Aguilera Vasquez N, et al. Biomarkers That Correlate with Active Pulmonary Tuberculosis Treatment Response: a Systematic Review and Meta-analysis. McAdam AJ, ed. J Clin Microbiol. 2022;60(2):e01859-21. doi:10.1128/jcm.01859-21
282. Heyckendorf J, Georghiou SB, Frahm N, et al. Tuberculosis Treatment Monitoring and Outcome Measures: New Interest and New Strategies. Clin Microbiol Rev. 2022;35(3):e00227-21. doi:10.1128/cmr.00227-21
283. Wejse C, Gustafson P, Nielsen J, et al. TBscore: Signs and symptoms from tuberculosis patients in a low-resource setting have predictive value and may be used to assess clinical course. Scandinavian Journal of Infectious Diseases. 2008;40(2):111-120. doi:10.1080/00365540701558698
284. Turner RD, Birring SS, Bothamley GH. Measuring cough-related quality of life and cough frequency in pulmonary TB. int j tuberc lung dis. 2024;28(2):93-98. doi:10.5588/ijtld.23.0312
285. Turner R, Repossi A, Matos S, Birring S, Bothamley G. S79 Cough Prevalence And Frequency In Pulmonary Tuberculosis. Thorax. 2014;69(Suppl 2):A43-A44. doi:10.1136/thoraxjnl-2014-206260.85
286. Sharma M, Nduba V, Njagi LN, et al. TBscreen: A passive cough classifier for tuberculosis screening with a controlled dataset. Sci Adv. 2024;10(1):eadi0282. doi:10.1126/sciadv.adi0282
287. Rudd M, Song WJ, Small PM. The Statistics of Counting Coughs: Easy as 1, 2, 3? Lung. 2022;200(5):531-537. doi:10.1007/s00408-022-00559-6
288. Proaño A, Bui DP, López JW, et al. Cough Frequency During Treatment Associated With Baseline Cavitary Volume and Proximity to the Airway in Pulmonary TB. Chest. 2018;153(6):1358-1367. doi:10.1016/j.chest.2018.03.006
289. Magombedze G, Pasipanodya JG, Gumbo T. Bacterial load slopes represent biomarkers of tuberculosis therapy success, failure, and relapse. Commun Biol. 2021;4(1):664. doi:10.1038/s42003-021-02184-0
290. R2D2 TB Network. R2D2 TB Network. Accessed October 24, 2024. https://www.r2d2tbnetwork.org
291. Martin-Higuera MC, Rivas G, Rolo M, Muñoz-Gallego I, Lopez-Roa P. Xpert MTB/RIF Ultra CT value provides a rapid measure of sputum bacillary burden and predicts smear status in patients with pulmonary tuberculosis. Sci Rep. 2023;13(1):1591. doi:10.1038/s41598-023-28869-6
292. Opota O, Mazza-Stalder J, Greub G, Jaton K. The rapid molecular test Xpert MTB/RIF ultra: towards improved tuberculosis diagnosis and rifampicin resistance detection. Clinical Microbiology and Infection. 2019;25(11):1370-1376. doi:10.1016/j.cmi.2019.03.021
293. World Health Organization. WHO Meeting Report of a Technical Expert Consultation: Non-Inferiority Analysis of Xpert MTB/RIF Ultra Compared to Xpert MTB/RIF.; 2017.
294. Nielsen MC, Clarner P, Paroha R, Lee S, Thwe PM, Ren P. Comparison of Analytical Sensitivity (Limit of Detection) of Xpert MTB/RIF and Xpert MTB/RIF Ultra for Non-Sputum Specimens. Pathogens. 2023;12(2):157. doi:10.3390/pathogens12020157

295. Chilukutu L, Mwanza W, Kerkhoff AD, Somwe P, Kagujje M, Muyoyeta M. Prevalence and interpretation of Xpert® Ultra trace results among presumptive TB patients. public health action. 2022;12(1):28-33. doi:10.5588/pha.21.0066
296. Murphy K, Habib SS, Zaidi SMA, et al. Computer aided detection of tuberculosis on chest radiographs: An evaluation of the CAD4TB v6 system. Sci Rep. 2020;10(1):5492. doi:10.1038/s41598-020-62148-y
297. Worodria W, Castro R, Kik SV, et al. An independent, multi-country head-to-head accuracy comparison of automated chest x-ray algorithms for the triage of pulmonary tuberculosis. Published online June 19, 2024. doi:10.1101/2024.06.19.24309061
298. Soares TR, Oliveira RD de, Liu YE, et al. Evaluation of chest X-ray with automated interpretation algorithms for mass tuberculosis screening in prisons: a cross-sectional study. Lancet Reg Health Am. 2023;17:100388. doi:10.1016/j.lana.2022.100388
299. Buuren SV, Groothuis-Oudshoorn K. mice : Multivariate Imputation by Chained Equations in R. J Stat Soft. 2011;45(3). doi:10.18637/jss.v045.i03
300. Robitzsch A, Grund S, Thorsten H. Package 'miceadds.' Published online January 9, 2024. https://cran.r-project.org/web/packages/miceadds/miceadds.pdf
301. Raberahona M, Zimmer A, Rakotoarivelo RA, et al. Continuous digital cough monitoring during 6-month pulmonary tuberculosis treatment. ERJ Open Research. Published online January 1, 2024. doi:10.1183/23120541.00655-2024
302. Gabaldón-Figueira JC, Keen E, Rudd M, et al. Longitudinal passive cough monitoring and its implications for detecting changes in clinical status. ERJ Open Res. 2022;8(2):00001-02022. doi:10.1183/23120541.00001-2022
303. Hedeker D, Gibbons RD. Longitudinal Data Analysis. John Wiley & Sons; 2006.
304. Hartig F, Lohse L. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. Published online September 8, 2022. Accessed July 4, 2024. https://CRAN.R-project.org/package=DHARMa
305. Brooks M, Bolker B, Kristensen K, Maechler M, Magnusson A, McGillycuddy M. glmmTMB: Generalized Linear Mixed Models using Template Model Builder. Published online March 20, 2024. https://cran.uni-muenster.de/web/packages/glmmTMB/glmmTMB.pdf
306. Zimmer AJ, Tsang LY, Jolicoeur G, et al. Incidence of cough from acute exposure to fine particulate matter (PM2.5) in Madagascar: A pilot study. Puttaswamy N, ed. PLOS Glob Public Health. 2024;4(7):e0003530. doi:10.1371/journal.pgph.0003530
307. Zhang M, Chen Y, Li L, Wang D. Speaker Recognition with Cough, Laugh and "Wei." Published online 2017. doi:10.48550/ARXIV.1706.07860
308. Sheridan C. COVID-19 spurs wave of innovative diagnostics. Nat Biotechnol. 2020;38(7):769-772. doi:10.1038/s41587-020-0597-x
309. Aly M, Alotaibi NS. A novel deep learning model to detect COVID-19 based on wavelet features extracted from Mel-scale spectrogram of patients' cough and breathing sounds. Informatics in Medicine Unlocked. 2022;32:101049. doi:10.1016/j.imu.2022.101049
310. Schuller BW, Coppock H, Gaskell A. Detecting COVID-19 from Breathing and Coughing Sounds using Deep Neural Networks. Published online 2020. doi:10.48550/ARXIV.2012.14553

311. Anupam A, Mohan NJ, Sahoo S, Chakraborty S. Preliminary Diagnosis of COVID-19 Based on Cough Sounds Using Machine Learning Algorithms. In: 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE; 2021:1391-1397. doi:10.1109/ICICCS51141.2021.9432324
312. Vrindavanam J, Srinath R, Shankar HH, Nagesh G. Machine Learning based COVID-19 Cough Classification Models - A Comparative Analysis. In: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE; 2021:420-426. doi:10.1109/ICCMC51019.2021.9418358
313. Adadi A, Berrada M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access. 2018;6:52138-52160. doi:10.1109/ACCESS.2018.2870052
314. Xia T, Spathis D, Brown C, et al. COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. In: ; 2021. Accessed September 3, 2024. https://openreview.net/forum?id=9KArJb4r5ZQ
315. MacLean ELH, Villa-Castillo L, Espinoza-Lopez P, et al. Integrating tuberculosis and COVID-19 molecular testing in Lima, Peru: a cross-sectional, diagnostic accuracy study. The Lancet Microbe. 2023;4(6):e452-e460. doi:10.1016/S2666-5247(23)00042-3
316. Barletta F, Otero L, de Jong BC, et al. Predominant Mycobacterium tuberculosis Families and High Rates of Recent Transmission among New Cases Are Not Associated with Primary Multidrug Resistance in Lima, Peru. J Clin Microbiol. 2015;53(6):1854-1863. doi:10.1128/JCM.03585-14
317. Institut national de santé publique du Québec. Données COVID-19 au Québec. Institut national de santé publique du Québec. Accessed November 4, 2024. https://www.inspq.qc.ca/covid-19/donnees
318. Schmidt K, Ben Moussa M, Buckrell S, et al. National Influenza Annual Report, Canada, 2022-2023: Canada's first fall epidemic since the 2019-2020 season. Can Commun Dis Rep. 2023;49(10):413-424. doi:10.14745/ccdr.v49i10a02
319. Public Health Agency of Canada. Tuberculosis in Canada: Infographic (2022). March 4, 2024. Accessed November 4, 2024. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/tuberculosis-canada-infographic-2022.html
320. Cao XJ, Fang KY, Li YP, Zhou J, Guo XG. The Diagnostic Accuracy of Xpert Xpress to SARS-CoV-2: A systematic review. J Virol Methods. 2022;301:114460. doi:10.1016/j.jviromet.2022.114460
321. Mostafa HH, Carroll KC, Hicken R, et al. Multicenter Evaluation of the Cepheid Xpert Xpress SARS-CoV-2/Flu/RSV Test. J Clin Microbiol. 2021;59(3):e02955-20. doi:10.1128/JCM.02955-20
322. Cepheid. Cepheid Receives Emergency Use Authorization from FDA for Rapid SARS-CoV-2 Test. News Release Archive. Accessed November 7, 2024. http://cepheid.mediaroom.com/2020-03-21-Cepheid-Receives-Emergency-Use-Authorization-from-FDA-for-Rapid-SARS-CoV-2-Test
323. Cepheid. Cepheid Receives Emergency Use Authorization for Xpert Xpress CoV-2/Flu/RSV plus. News Release Archive. Accessed November 7, 2024.

http://cepheid.mediaroom.com/2021-09-14-Cepheid-Receives-Emergency-Use-Authorization-for-Xpert-Xpress-CoV-2-Flu-RSV-plus

324. World Health Organization. WHO Meeting Report of a Technical Expert Consultation: Non-Inferiority Analysis of Xpert MTB/RIF Ultra Compared to Xpert MTB/RIF.; 2017.
325. Creager HM, Cabrera B, Schnaubelt A, et al. Clinical evaluation of the BioFire® Respiratory Panel 2.1 and detection of SARS-CoV-2. Journal of Clinical Virology. 2020;129:104538. doi:10.1016/j.jcv.2020.104538
326. Peeters, G. A large set of audio features for sound description (similarity and classification) in the cuidado project. CUIDADO Ist Project Report. 2004;54(0):1-25.
327. Sreeram ASK, Ravishankar U, Sripada NR, Mamidgi B. Investigating the potential of MFCC features in classifying respiratory diseases. In: 2020 7th International Conference on Internet of Things: Systems, Management and Security (IOTSMS). IEEE; 2020:1-7. doi:10.1109/IOTSMS52051.2020.9340166
328. Podder P, Zaman Khan T, Haque Khan M, Muktadir Rahman M. Comparative Performance Analysis of Hamming, Hanning and Blackman Window. IJCA. 2014;96(18):1-7. doi:10.5120/16891-6927
329. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Published online 2016. doi:10.48550/ARXIV.1603.02754
330. Sagi O, Rokach L. Ensemble learning: A survey. WIREs Data Min & Knowl. 2018;8(4):e1249. doi:10.1002/widm.1249
331. Irawati ME, Zakaria H. Classification Model for Covid-19 Detection Through Recording of Cough Using XGboost Classifier Algorithm. In: 2021 International Symposium on Electronics and Smart Devices (ISESD). IEEE; 2021:1-5. doi:10.1109/ISESD53023.2021.9501695
332. Deng H, Zhou Y, Wang L, Zhang C. Ensemble learning for the early prediction of neonatal jaundice with genetic features. BMC Med Inform Decis Mak. 2021;21(1):338. doi:10.1186/s12911-021-01701-9
333. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair. 2002;16:321-357. doi:10.1613/jair.953
334. Shapley LS. 17. A Value for n-Person Games. In: Kuhn HW, Tucker AW, eds. Contributions to the Theory of Games (AM-28), Volume II. Princeton University Press; 1953:307-318. doi:10.1515/9781400881970-018
335. Štrumbelj E, Kononenko I. Explaining prediction models and individual predictions with feature contributions. Knowl Inf Syst. 2014;41(3):647-665. doi:10.1007/s10115-013-0679-x
336. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. Published online November 25, 2017. doi:10.48550/arXiv.1705.07874
337. Ravanelli M, Parcollet T, Plantinga P, et al. SpeechBrain: A General-Purpose Speech Toolkit. Published online 2021. doi:10.48550/ARXIV.2106.04624
338. Owens FJ, Murphy MS. A short-time Fourier transform. Signal Processing. 1988;14(1):3-10. doi:10.1016/0165-1684(88)90040-0

339. Guo J, Zhu J, Lin S, Shi F. ECAPA-TDNN Embeddings for Speaker Recognition. In: 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT). IEEE; 2024:1488-1491. doi:10.1109/AINIT61980.2024.10581514
340. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12(1):77. doi:10.1186/1471-2105-12-77
341. Melek Manshouri N. Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. Cogn Neurodyn. 2022;16(1):239-253. doi:10.1007/s11571-021-09695-w
342. Nayak SS, Darji AD, Shah PK. Machine learning approach for detecting Covid-19 from speech signal using Mel frequency magnitude coefficient. SIViP. 2023;17(6):3155-3162. doi:10.1007/s11760-023-02537-8
343. Akman A, Coppock H, Gaskell A, Tzirakis P, Jones L, Schuller BW. Evaluating the COVID-19 Identification ResNet (CIdeR) on the INTERSPEECH COVID-19 From Audio Challenges. Front Digit Health. 2022;4:789980. doi:10.3389/fdgth.2022.789980
344. Amin SU, Taj S, Hussain A, Seo S. An automated chest X-ray analysis for COVID-19, tuberculosis, and pneumonia employing ensemble learning approach. Biomedical Signal Processing and Control. 2024;87:105408. doi:10.1016/j.bspc.2023.105408
345. Kumar Acharya A, Satapathy R. A Deep Learning Based Approach towards the Automatic Diagnosis of Pneumonia from Chest Radio-Graphs. Biomed Pharmacol J. 2020;13(1):449-455. doi:10.13005/bpj/1905
346. Naji O, Darwish I, Bessame K, et al. A Comparison of the Epidemiological Characteristics Between Influenza and COVID-19 Patients: A Retrospective, Observational Cohort Study. Cureus. 2023;15(11):e49280. doi:10.7759/cureus.49280
347. Venkataramana L, Prasad DVV, Saraswathi S, Mithumary CM, Karthikeyan R, Monika N. Classification of COVID-19 from tuberculosis and pneumonia using deep learning techniques. Med Biol Eng Comput. 2022;60(9):2681-2691. doi:10.1007/s11517-022-02632-x
348. Goel S, Kipp A, Goel N, Kipp J. COVID-19 vs. Influenza: A Chest X-ray Comparison. Cureus. 2022;14(11):e31794. doi:10.7759/cureus.31794
349. Saleh Y. Synthesizing Cough Audio with GAN for COVID-19 Detection. Published online 2023. doi:10.48550/ARXIV.2305.04810
350. Tavaziva G, Harris M, Abidi SK, et al. Chest X-ray Analysis With Deep Learning-Based Software as a Triage Test for Pulmonary Tuberculosis: An Individual Patient Data Meta-Analysis of Diagnostic Accuracy. Clinical Infectious Diseases. 2022;74(8):1390-1400. doi:10.1093/cid/ciab639
351. Kapetanidis P, Kalioras F, Tsakonas C, et al. Respiratory Diseases Diagnosis Using Audio Analysis and Artificial Intelligence: A Systematic Review. Sensors. 2024;24(4):1173. doi:10.3390/s24041173
352. Yellapu GD, Rudraraju G, Sripada NR, et al. Development and clinical validation of Swaasa AI platform for screening and prioritization of pulmonary TB. Sci Rep. 2023;13(1):4740. doi:10.1038/s41598-023-31772-9

353. Huddart S, Yadav V, Sieberts SK, et al. A dataset of Solicited Cough Sound for Tuberculosis Triage Testing. Sci Data. 2024;11(1):1149. doi:10.1038/s41597-024-03972-z
354. Jaganath D, Sieberts SK, Raberahona M, et al. Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM Challenge. Published online May 14, 2024. doi:10.1101/2024.05.13.24306584
355. Nihues SDSE, Mancuzo EV, Sulmonetti N, et al. Chronic symptoms and pulmonary dysfunction in post-tuberculosis Brazilian patients. The Brazilian Journal of Infectious Diseases. 2015;19(5):492-497. doi:10.1016/j.bjid.2015.06.005
356. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A Survey on Bias and Fairness in Machine Learning. ACM Comput Surv. 2022;54(6):1-35. doi:10.1145/3457607
357. Khine M. Editorial: Clinical validation of digital health technologies for personalized medicine. Front Digit Health. 2022;4:831517. doi:10.3389/fdgth.2022.831517
358. Makri A. Bridging the digital divide in health care. The Lancet Digital Health. 2019;1(5):e204-e205. doi:10.1016/S2589-7500(19)30111-6
359. Whitehead L, Talevski J, Fatehi F, Beauchamp A. Barriers to and Facilitators of Digital Health Among Culturally and Linguistically Diverse Populations: Qualitative Systematic Review. J Med Internet Res. 2023;25:e42719. doi:10.2196/42719

Doctoral training publication list

1. Raberahona M*, **Zimmer A***, Rakotoarivelo RA, Randrianarisoa PA, Rambeloson G, Rakotomijoro E, Andry CE, Razafindrakoto HA, Andriantahiana D, Randria M, Rakotosamimanana N, Grandjean Lapierre S. Continuous digital cough monitoring during 6-month pulmonary tuberculosis treatment. ERJ Open Research. 2024. doi:10.1183/23120541.00655-2024
2. **Zimmer AJ**, Tsang LY, Jolicoeur G, Tannir B, Batisse E, Pando C, Sadananda G, McKinney J, Ambinintsoa IV, Rabetombosoa RM, Knoblauch AM, Rakotosamimanana N, Chartier R, Diachenko A, Small P, Grandjean Lapierre S. Incidence of cough from acute exposure to fine particulate matter (PM2.5) in Madagascar: A pilot study. Puttaswamy N, ed. PLOS Glob Public Health. 2024;4(7):e0003530. doi:10.1371/journal.pgph.0003530
3. MacLean ELH, **Zimmer AJ**, Den Boon S, Gupta-Wright A, Cirillo DM, Cobelens F, Gillespie SH, Nahid P, Phillips P, Ruhwald M, Denkinger CM. Tuberculosis treatment monitoring tests during routine practice: study design guidance. Clinical Microbiology and Infection. 2024;30(4):481-488. doi:10.1016/j.cmi.2023.12.027
4. Thompson RR, Nalugwa T, Oyuku D, Tucker A, Nantale M, Nakaweesa A, Musinguzi J, Reza TF, **Zimmer AJ**, Ferguson O, Turyahabwe S, Joloba M, Cattamanchi A, Katamba A, Dowdy DW, Sohn H. Multicomponent strategy with decentralised molecular testing for tuberculosis in Uganda: a cost and cost-effectiveness analysis. The Lancet Global Health. 2023;11(2):e278-e286. doi:10.1016/S2214-109X(22)00509-5
5. McGrath S, Katzenschlager S, **Zimmer AJ**, Seitel A, Steele R, Benedetti A. Standard error estimation in meta-analysis of studies reporting medians. Stat Methods Med Res. 2023;32(2):373-388. doi:10.1177/09622802221139233

6. **Zimmer AJ**, Ugarte-Gil C, Pathri R, Dewan P, Jaganath D, Cattamanchi A, Pai M, Grandjean Lapierre S. Making cough count in tuberculosis care. Commun Med. 2022;2(1):83. doi:10.1038/s43856-022-00149-w
7. **Zimmer AJ**, Klinton JS, Oga-Omenka C, Heitkamp P, Nyirenda CN, Furin J, Pai M. Tuberculosis in times of COVID-19. J Epidemiol Community Health. 2022;76(3):310-316. doi:10.1136/jech-2021-217529
8. **Zimmer AJ**, Lainati F, Aguilera Vasquez N, Chedid C, McGrath S, Benedetti A, Maclean E, Ruhwald M, Denkinger CM. Biomarkers That Correlate with Active Pulmonary Tuberculosis Treatment Response: a Systematic Review and Meta-analysis. McAdam AJ, ed. J Clin Microbiol. 2022;60(2):e01859-21. doi:10.1128/jcm.01859-21
9. **Zimmer AJ**, Heitkamp P, Malar J, Malar J, Dantas C, O'Brien K, Pandita A, Waite RC. Facility-based directly observed therapy (DOT) for tuberculosis during COVID-19: A community perspective. Journal of Clinical Tuberculosis and Other Mycobacterial Diseases. 2021;24:100248. doi:10.1016/j.jctube.2021.100248
10. Katzenschlager S*, **Zimmer AJ***, Gottschalk C, Grafeneder J, Schmitz S, Kraker S, Ganslmeier M, Muth A, Seitel A, Maier-Hein L, Benedetti A, Larmann J, Weigand MA, McGrath S, Denkinger CM. Can we predict the severe course of COVID-19 - a systematic review and meta-analysis of indicators of clinical outcome? Lazzeri C, ed. PLoS ONE. 2021;16(7):e0255154. doi:10.1371/journal.pone.0255154
11. **Zimmer AJ**, Schumacher SG, Södersten E, Mantsoki A, Wyss R, Persing DH, Banderby S, Strömqvist Meuzelaar L, Prieto J, Gnanashanmugam D, Khatri P, Ongarello S, Ruhwald M, Denkinger CM. A novel blood-based assay for treatment monitoring of tuberculosis. BMC Res Notes. 2021;14(1):247. doi:10.1186/s13104-021-05663-z
12. Broger T, Nicol MP, Sigal GB, Gotuzzo E, **Zimmer AJ**, Surtie S, Caceres-Nakiche T, Mantsoki A, Ivanova Reipold E, Székely R, Tsionsky M, van Heerden J, Plisova T, Chikamatsu K, Lowary TL, Pinter A, Mitarai S, Moreau E, Schumacher SG, Denkinger CM. Diagnostic accuracy of 3 urine lipoarabinomannan tuberculosis assays in HIV-negative outpatients. Journal of Clinical Investigation. 2020;130(11):5756-5764. doi:10.1172/JCI140461
13. Faust L*, **Zimmer AJ***, Kohli M, Saha S, Boffa J, Bayot ML, Nsofor I, Campos L, Mashamba-Thompson T, Herrera R, Emeka E, Shrestha S, Ugarte-Gil C, Katamba A, Pambudi I, Bichara D, Calderon RI, Ahmadzada N, Safdar MA, Nikam C, Dos Santos Lázari C, Hussain H, Win MM, Than KZ, Ahumibe A, Waning B, Pai M. SARS-CoV-2 testing in low- and middle-income countries: availability and affordability in the private health sector. Microbes and Infection. 2020;22(10):511-514. doi:10.1016/j.micinf.2020.10.005

*co-first authors