Addis Ababa University

College of Technology and Built Environment

School of Biomedical Engineering

Prescreening of Tuberculosis Based on Cough Sound Analysis and Clinical Information Using Machine Learning

by

Woineshet Habtom Bekele

In partial fulfilment of the requirements for the degree of

Master of Engineering in Biomedical Engineering

Supervisors

Dr. Dawit Assefa Haile (BSc, MSc, PhD)

Addis Ababa, Ethiopia

October 13, 2025

Declaration

I, the undersigned, declare that this Master of Engineering Industry Project is my original work, has not been presented to fulfill a degree in this or any other University, and all sources and materials used for the industry project have been acknowledged.

Signed: _____

Date: _____

Approval

This industry project, titled Prescreening Tuberculosis Based on Cough Sound Analysis and Clinical Information Using Machine Learning, has been approved by the following supervisors.

Supervisors:

Dr. Dawit Assefa Haile (BSc, MSc, PhD) *Sign:*_____*Date:*_____

Certificate of Examination

This is to certify that the industry project prepared by Woineshet Habtom entitled "Pre-screening of Tuberculosis Based on Cough Sound Analysis and Clinical Information Using Machine Learning" submitted in partial fulfillment of the requirements for the degree of Master of Engineering in Biomedical Engineering (Bioinstrumentation and Imaging) complies with the regulations of the University and meets the accepted standards concerning originality and quality.

Signed by the examining committee:

Examiner:_____ *Sign:*_____ *Date:*_____

Examiner:_____ *Sign:*_____ *Date:*_____

Advisor:_____ *Sign:*_____ *Date:*_____

Chief of Department or Graduate Program Coordinator

Dedication

First and foremost, I dedicate this work to God, for giving me strength and purpose throughout this journey. To my father, my husband and all my family, your faith in me, your sacrifices, and your unwavering support carried me through every challenge. I am forever grateful.

Acknowledgements

First and foremost, I thank God for granting me strength throughout this academic journey. Without his guidance and grace, this achievement would not have been possible.

I would like to express my deepest appreciation to my supervisor, Dr. Dawit Assefa Haile, whose expertise, guidance, and patience have greatly enriched my graduate experience to the completion of this research.

I am also sincerely grateful to my family for their unwavering support, love, and encouragement throughout my studies.

Special thanks go to ALERT Comprehensive Specialized Hospital, St. Peter Specialized Hospital, and Dilfre Health Center, and their dedicated staff for their support during the data collection phase.

Lastly, I extend my heartfelt thanks to my colleagues and the university staff who have contributed to my journey. Their consistent support and cooperation has been truly invaluable.

Abstract

Tuberculosis (TB) remains a major global health concern, particularly in low and middle-income countries (LMICs) where diagnostic resources are limited. TB is an infectious disease caused by bacteria that most often affects the lungs. Early diagnosis of TB is critical for treatment as well as to reduce its transmission. Current TB diagnosis techniques include chest radiography, sputum smear microscopy, tuberculin skin tests, sputum culture, and Gene-Xpert assays, which often require specialized facilities and trained personnel, posing challenges in resource-limited settings. This study develops a prescreening tool for TB detection that leverages cough sound analysis and patient-specific clinical information through a deep learning approach. The dataset used includes 29,768 cough sounds and clinical information from 1,105 individuals collected across seven countries. The developed model, a convolutional neural network (CNN)-based deep learning architecture for binary classification, employs a two-stage design: a base CNN model (CNNbasedNet121) trained on 1D convolutional layers of cough audio records and an enhanced multimodal model (CNNbasedNet) that fuses learned audio features with clinical metadata through dense layers to improve performance. Preprocessing was performed, including scaling, categorical encoding, data imputation for clinical data, as well as clipping and padding of audio waveforms, and memory optimization for both clinical metadata and audio recordings. Data visualization techniques such as histograms, box plots, and correlation heatmaps were used to validate data distribution. Evaluation of the model showed promising classification results with AUC-ROC of 98.59%, 94.22% accuracy, 91.58% recall, 0.8843 Precision, and 0.8998 F1-score. SHAP analysis was applied to interpret model decisions as part of the explainable AI, which revealed that weight, heart rate, smoking history, and night sweat were among the most influential features. In addition, a user-friendly web-based interface was developed using Streamlit to allow real-time TB prediction based on audio uploads and clinical information input. This interface aims to simplify the process for users, making it more automatic, rapid, cost-effective, and accurate to diagnose TB from cough sound and clinical information. To further assess real-world applicability, the deployed model was tested using a separate dataset collected from selected Ethiopian healthcare facilities. Patient cough audio was recorded using a mobile application called Cough Detector and Recorder App, and clinical data was collected through a structured questionnaire. The deployment results confirmed the model's practical effectiveness in real-world LMIC settings, achieving an accuracy of 83.33% and a sensitivity of 77.78%.

Keywords: TB, Cough sound, Clinical information, CNN

Contents

List of Figures

List of Tables

List of Acronyms

1D	One-Dimensional
2D	Two-Dimensional
AFB	Acid-Fast Bacillus
AI	Artificial Intelligence
AIDS	Acquired Immunodeficiency Syndrome
ANN	Artificial Neural Network
AUC	Area Under the Curve
BCG	Bacille Calmette-Guérin
Bi-GRU	Bidirectional Gated Recurrent Unit
BiLSTM	Bidirectional Long Short-Term Memory
BMI	Body Mass Index
CNN	Convolutional Neural Network
CODA	Cough Detection Algorithm
COVID-19	Coronavirus Disease 2019
CSV	Comma-Separated Values
DMRNet	Deep Multi-Resolution Network
FCNN	Fully Connected Neural Network
FFANN	Feedforward Artificial Neural Network
FN	False Negative
FP	False Positive
GoogLeNet	Google Inception Network
GUI	Graphical User Interface
HIV	Human Immunodeficiency Virus
HOG	Histogram of Oriented Gradients
HR	Heart Rate
IGRAs	Interferon Gamma Release Assays

KNN	k-Nearest Neighbors
LFB	Local Feature Block
LR	Logistic Regression
LSTM	Long Short-Term Memory
LTBI	Latent Tuberculosis Infection
MFB	Multi-scale Feature Block
MFCCs	Mel-Frequency Cepstral Coefficients
MHSA	Multi-Head Self Attention
ML	Machine Learning
MLP	Multilayer Perceptron
MSE	Mean Squared Error
MTB	Mycobacterium tuberculosis
MUAC	Mid-Upper Arm Circumference
NAA	Nucleic Acid Amplification
NLP	Natural Language Processing
ORCID	Open Researcher and Contributor ID
PPD	Purified Protein Derivative
PR	Precision-Recall
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSA	Position-Sensitive Attention
ReLU	Rectified Linear Unit
ResNet50	50-layer Residual Network
RF	Random Forest
RMSprop	Root Mean Square Propagation
ROC	Receiver Operating Characteristic
SFS	Sequential Forward Selection
SGD	Stochastic Gradient Descent
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine

TB	Tuberculosis
TN	True Negative
TP	True Positive
TST	Mantoux Tuberculin Skin Test
VGG16	Visual Geometry Group 16-layer Network
WHO	World Health Organization
ZCR	Zero-Crossing Rate

Chapter 1

Introduction

1.1 Background

Tuberculosis (TB) remains a global health concern, with approximately 1.25 million deaths and an estimated 10.8 million new cases in 2023, of which about 8.2 million were newly diagnosed and reported, according to the 2024 World Health Organization (WHO) global tuberculosis report [1]. Figure 1.1 illustrates the global distribution of TB incidence [2]. Ethiopia is one of the 30 highest TB burden countries in the world. Based on WHO's 2020 report, the number of deaths due to tuberculosis in Ethiopia reached 21,455, which accounts for 3.81% of total deaths [3]. Before the onset of the Covid-19 pandemic, TB ranked as the foremost infectious cause of death globally, exceeding that of HIV/AIDS [4].

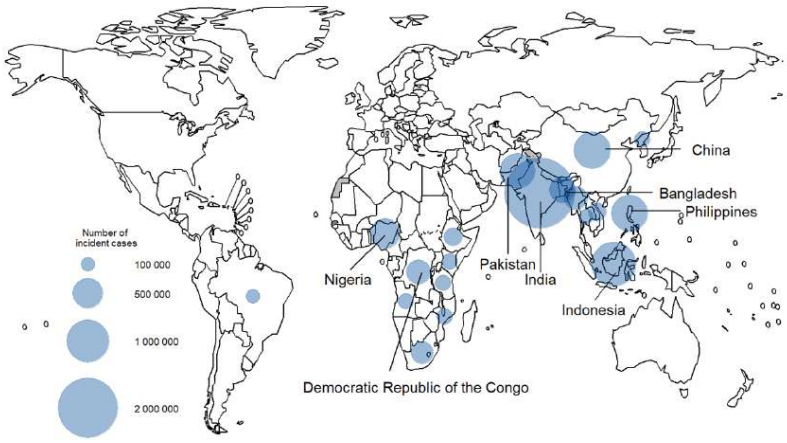


Figure 1.1: Global distribution of tuberculosis (TB) incidence in 2023.

TB is caused by a bacterium called mycobacterium tuberculosis, which is primarily transmitted through air, particularly via coughing. Thus, the infection begins when an individual inhales this bacterium. Existing TB tests encompass different clinical and laboratory techniques, which include sputum smear microscopy, chest radiography, tuberculin skin tests, sputum culture, and Gene-Xpert assays. While these diagnostic techniques are effective, they often require specialized facilities and trained personnel, posing significant challenges in resource-limited settings [5, 6]. Consequently, there is a critical need for an affordable, rapid, and accessible TB pre-screening method, particularly in low and middle income countries (LMICs), to prevent delayed diagnosis, ongoing transmission, and mortality [7, 8].

Coughing, characterized by an explosive expulsion of air from the airways, is a common symptom of respiratory diseases and serves as a valuable indicator for diagnosis [9, 10]. The acoustic properties of coughs can vary significantly depending on the underlying condition,

offering essential insights into someone's respiratory health that may aid clinical decision-making [11].

Recently, advanced computational techniques have been developed and have played a significant role across many sectors. Among these, machine learning has become particularly dominant, increasingly replacing manual work in different domains [12]. Accordingly, the utilization of machine learning algorithms to analyze and classify cough sound patterns introduces a promising and non-invasive approach for pre-screening of TB and other respiratory diseases. These coughs frequently display alterations in their sound characteristics referred to as acoustic variations that reflect underlying respiratory conditions and provide a non-invasive pre-screening techniques, enabling to distinguish TB and non-TB cases. The detection and classification of cough sounds have been active areas of research since as early as 1989, contributing significantly to the diagnosis of various pulmonary diseases [7].

To support this approach, the current study utilizes a comprehensive dataset consisting of 29,768 cough sounds and 16 corresponding clinical data collected from 1,105 individuals across seven countries. The dataset provides diversity in demographics and symptom profiles, reflecting real-world variability in TB and non-TB cases [13]. The developed model employs a two-stage deep learning architecture based on Convolutional Neural Networks (CNNs). The first stage (Base model) processes audio features extracted from cough sounds using 1D convolutional layers, and the second stage (Enhanced Model) processes the standardized clinical features and fuses them with the learned audio features through fully connected dense layers to generate binary predictions showing the likelihood of TB. This design leverages the complementary strengths of audio and clinical data, enhancing diagnostic accuracy while maintaining accessibility and ease of use.

To maximize the practical utility of this model, a user-friendly web-based interface was developed using Streamlit, enabling real-time TB prediction based on uploaded cough audio and inputted patient-specific clinical information. This interface is designed to simplify the diagnostic process and provide an accessible, rapid, and cost-effective prescreening tool for healthcare professionals and community health workers. This study aims to bridge the gap in TB screening in low-resource countries such as Ethiopia, where timely diagnosis remains a critical challenge by leveraging digital health solutions that do not require laboratory infrastructure. It does so by combining deep learning, acoustic analysis, and clinical data in a deployable format.

1.1.1 Tuberculosis in the World and in Ethiopia

In 2023, TB continued to pose a major health challenge worldwide, with an estimated 10.8 million individuals affected by the disease around the world [1]. This figure underscores the persistent burden of TB, particularly in regions with limited healthcare infrastructure. The distribution of cases across countries highlights the urgent need for international collabora-

tion in disease surveillance, prevention, and control. Identifying the most affected regions and investigating the contributing factors is essential for a comprehensive understanding of the global TB burden. Such insights are critical for informing targeted strategies to reduce transmission and mortality, especially in high-incidence regions such as those within the African continent [2].

Ethiopia is among the 30 countries with the highest incidence rates of TB, as identified by the WHO [2]. Figure 1.2 illustrates the estimated incidence rates (shown in green solid lines) and the notifications of new and relapse cases (shown in black solid lines). The green shaded areas represent 95% uncertainty intervals. The horizontal dashed line shows the 2025 milestone of the End TB Strategy, which was a 50% reduction between the years 2015 and 2025 [2].

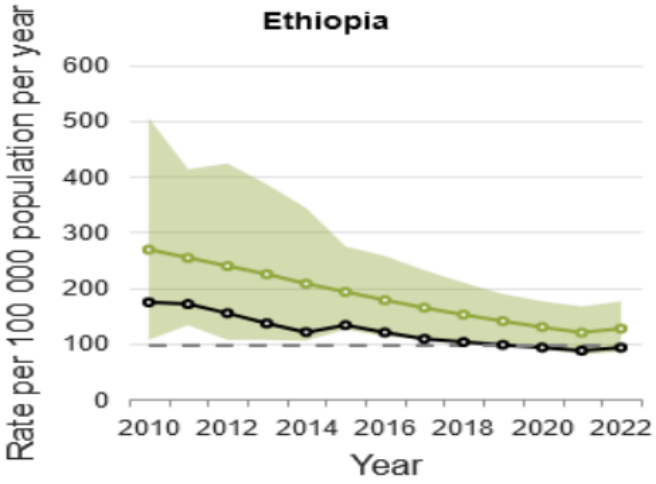


Figure 1.2: Trends in TB incidence rates and notifications of new and relapse cases in Ethiopia

1.2 Problem Statement

TB is one of the leading causes of death worldwide, despite being preventable and curable through early diagnosis and appropriate treatment [14]. However, in low- and middle-income countries, access to reliable and timely diagnostic tools is often limited: some methods are expensive, others require trained personnel or laboratory infrastructure, some involve radiation exposure, and certain ones depend on invasive sample collection. Additionally, certain tests take a long time to give results. Because of these limitations, many patients are not diagnosed or treated in time, which increases disease transmission and the emergence of multidrug-resistant TB (MDR-TB). Moreover, commonly used diagnostic methods like chest X-rays and conventional sputum smear microscopy (e.g., Ziehl–Neelsen staining) have limited sensitivity and specificity, which can result in inaccurate TB diagnoses. Accessibility to these diagnostic tools also remains a major challenge in low-resource environments [15, 16].

Although recent studies have explored AI-based TB detection using either cough sounds or clinical information, most rely on a single data source, which may limit diagnostic accuracy

by failing to capture the complementary information that each modality provides. Additionally, many existing models are complex for real-time deployment in low-resource settings and lack interpretability, making it difficult for healthcare workers to understand or trust the predictions.

Therefore, it is essential to develop a prescreening tool that addresses these gaps, especially in remote and underserved regions. This study presents a practical and multimodal prescreening tool that combines cough sound analysis with clinical information using a custom-designed two-stage deep CNN architecture, offering a promising solution for early identification of individuals at risk of TB. A web-based application has been developed to ensure the tool is accessible, low-cost, non-invasive, rapid, automated, and suitable for real-time use in low-resource settings. The model incorporates interpretability to enhance transparency and was tested on data collected in Ethiopia to evaluate its practical relevance and potential as a prescreening tool in resource-limited environments.

1.3 Objectives

1.3.1 General Objective

The general objective of the study is to develop web-based CNN model for prescreening of TB using cough sounds and patient-specific clinical information.

1.3.2 Specific Objectives

- To develop a CNN-based model for classifying TB from cough sound data.
- To integrate clinical information with the model to enhance classification accuracy.
- To analyze the impact of clinical information on the model's prediction ability.
- To evaluate the model's performance using different classification metrics.
- To develop and implement a web-based TB prescreening application and evaluate its practical relevance using data collected in Ethiopia.

1.4 Research Question

Is it possible to accurately differentiate between TB and non-TB cases using cough sounds and other relevant clinical information taken from patients through CNN technique?

1.5 Significance of the Study

This study presents a cost-effective and accessible method for early TB prescreening using cough sound records by a smartphone, enhancing diagnostic accuracy by integrating clinical metadata, particularly in resource-limited communities like Ethiopia.

1.6 Scope of the Study

This study focuses on the development and evaluation of a CNN-based prescreening model for TB using cough sound and clinical information. The study does not include other respiratory diseases.

1.7 Structure of the Project Report

The rest of this project document is organized into five chapters. Chapter 2 discusses the theoretical background of the study. The third chapter attempts to address related works pertinent to the study, while the methods employed in this study are outlined in the fourth chapter. The fifth and sixth chapters present the results and discussion, and the conclusion, respectively.

Chapter 2

Theoretical Background

2.1 Tuberculosis

TB is an infectious disease caused by the bacteria called Mycobacterium tuberculosis (MTB). Although it mainly affects the lungs, the bacterium can also infect other parts of the body, such as kidneys, spine, and the brain [25]. The disease is transmitted through the air when an infected individual coughs or sneezes, releasing respiratory droplets that contain the bacteria [17]. While not everyone infected with TB will develop the disease, it is important to understand the two main conditions related to TB: latent TB infection (LTBI) and active TB disease [18].

Latent TB infection (LTBI) is a condition when someone is infected with Mycobacterium TB but shows no symptoms of illness. In this state, the immune system controls the bacteria without completely eliminating them. The bacteria remain dormant in the body, and individuals with LTBI are non-contagious and show no radiographic or microbiological evidence of active disease [17, 19]. Active TB disease, on the other hand, occurs when the immune system fails to control the bacteria, allowing them to multiply and cause illness. Individuals with active TB often have symptoms like persistent cough, coughing up blood or sputum, fever, loss of appetite, chest pain, and weight loss. In this condition, they become contagious and can spread the bacteria to others through droplets in the air. If not treated, active TB can lead to serious health issues and sometimes cause death [15, 20].

2.2 Signs and Symptoms of TB

TB shows many signs and symptoms that change based on the stage and severity of the disease. The most common symptoms include persistent cough, coughing up blood or sputum, fever, loss of appetite, chest pain, and weight loss, weakness or fatigue, and sweating at night [15, 21, 22].

2.3 Cause and Pathogenesis of TB

As mentioned above, TB is caused by MTBs, also known as the tubercle bacillus. This pathogen is highly contagious and is primarily transmitted through droplets in the air, especially in a poorly ventilated environment [21]. Once inhaled, the bacteria travel to the lungs and start to multiply. MTB replicates slowly, dividing approximately once every 16 to 20 hours, significantly slower than most bacteria [15]. Because of its slow growth, culturing of the bacteria takes several weeks. This delay makes it challenging for a timely diagnosis of TB [16, 23, 24].

The bacterium also has a special cell wall rich in lipids, particularly mycolic acids. It is

a waxy layer which makes it resistant to many disinfectants and antibiotics [25]. Beyond its structural resistance, the thick, lipid-rich cell wall allows the bacterium to survive, grow, and replicate inside immune cells called macrophages. This immune cell typically destroys pathogens. In this protected intracellular environment, the bacterium may remain dormant for a long time. It may later reactivate when the immune system becomes weakened, causing active TB [26].

2.4 Transmission of Tuberculosis

TB is most commonly transmitted through the air when an infected person coughs, sneezes, or speaks, releasing tiny particles known as droplet nuclei. These particles can remain suspended in the air for many hours, especially in enclosed or inadequately ventilated environments, which increases the chances of others inhaling them. This airborne transmission pathway is illustrated in Figure 2.1 [27]. The longer someone stays in a space containing these droplets, the higher their risk of being exposed to TB [28, 29]. These droplet nuclei that contain the MTB bacteria become airborne when moisture in larger droplets evaporates. Once inhaled, the bacteria can either settle on the upper airways, where the likelihood of infection is reduced, or travel deeper into the lungs. If they reach the alveoli, the tiny air sacs in the lungs, the body's immune system responds by sending macrophages to engulf the bacteria [30]. At this point, the bacteria may be eradicated or killed, remain dormant in latent state within the body, or begin to multiply, leading to the onset of active TB disease. The progression of infection depends upon various factors such as the ability of the bacterial pathogen to cause damage to a host, the infectiousness of the initial case, the immune response of the individual, living conditions, and environmental influences [17, 31].

In the majority of instances, the immune system effectively manages the infection, leading to latent TB infection. However, in persons with weakened immunity, such as those suffering with HIV, malnutrition, or chronic disease the bacteria may become active and cause illness [26]. While TB predominantly affects the lungs (pulmonary TB), it can also spread to other organs such as the lymph nodes, spine, or brain, which is referred to as extrapulmonary TB [32].

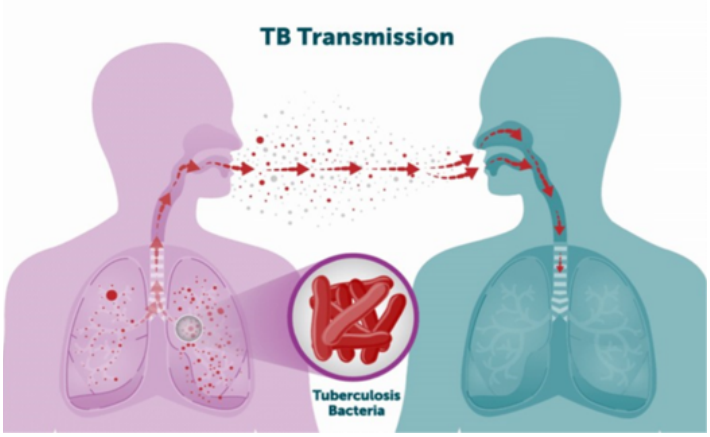


Figure 2.1: Illustration of how tuberculosis is transmitted through the air

After the initial exposure and inhalation of MTB, the first stage of infection is referred to as the primary infection. In certain individuals, the immune system eliminates the bacteria at this stage. Nevertheless, in many cases, some bacteria survive in a dormant, slow-growing state and become trapped within immune structures called granulomas. This defines the onset of latent TB infection, during which the infected person has no symptoms. The bacteria may be reactivated if the immune system is weakened. This phenomenon, referred to as post-primary TB, can occur months or even years later, leading to active and potentially transmissible disease. [26, 33]. The progression from primary infection to either latent or active TB is summarized in Figure 2.2 [34].

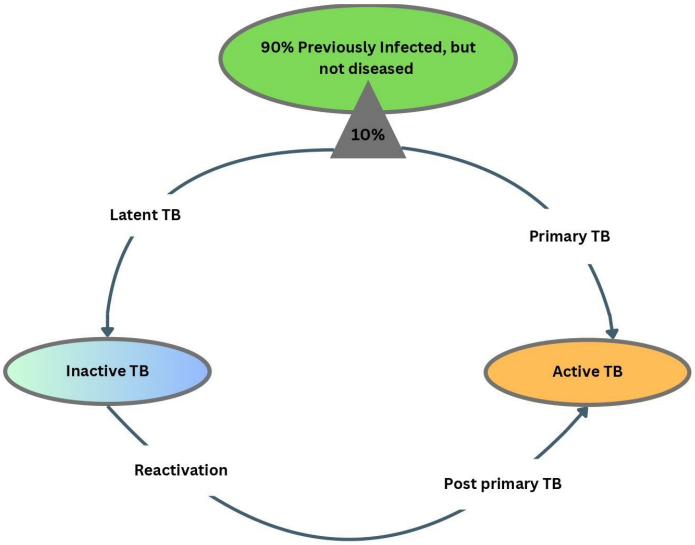


Figure 2.2: Natural history of tuberculosis infection. Approximately 90% of individuals infected with MBT do not develop active disease and enter a latent TB phase. About 10% progress to active TB, either shortly after primary infection or through later reactivation (post-primary TB). Source: *The Radiology Assistant*

2.5 Diagnostic Methods of TB

To identify both latent TB infection and active TB disease, a combination of physical examinations such as symptom assessment and medical history, imaging and laboratory tests are involved [15].

2.5.1 Physical Examinations

Common symptoms of TB such as persistent cough, fever, night sweats and weight lose are very important for physicians during initial diagnosis. Medical history is also an important step in diagnosing TB. It gives general information about potential exposure to TB, evaluating risk factors such as working or living in high-risk environments and other medical conditions. It is also necessary to assess for underlying conditions like HIV or diabetes that may increase the likelihood of progression from latent to active TB. Although a physical examination can reveal general signs of illness and help assess overall health, it may also identify indicators suggestive of extra-pulmonary TB, such as swelling over lymph nodes

(scrofula). However, these signs are not specific to TB. Therefore, a physical examination alone is not sufficient to confirm or rule out TB and should always be supported by further diagnostic testing and clinical assessment for appropriate treatment planning [15, 35].

2.5.2 Laboratory Tests

TB blood test (Interferon Gamma Release Assays-IGRAs): This test measures the immune response of the body to TB bacteria by mixing a blood sample with TB protein. This test is favored for individuals who have previously received the Bacillus Calmette-Guérin (BCG) vaccine. Additional tests may be required because this test alone cannot identify whether the TB infection is latent (inactive) or active [15].

Tuberculin skin test (Mantoux): Tuberculin skin test (TST) or Mantoux is used to assess TB infection by injecting 0.1 ml of tuberculin purified protein derivative (PPD), which is derived from tuberculin, under the skin to evaluate reaction after 48-72 hours, which determine if TB bacteria are present in the body (see Figure 2.3) [36]. Induration result 5mm or greater: considered positive for individuals with HIV infection, recent contact with TB case, fibrotic changes on chest x-ray, or organ transplant recipients. Induration result 10mm or greater: considered positive for recent arrival from high-prevalence countries, injection drug users, residents or employees of high-risk congregate settings, Mycobactriology lab personnel, children less than 4 years old, infants, children, and adolescents exposed to high-risk categories. Whereas, induration greater than 15mm: considered positive with no known risk factors for TB [36]. This test also required further examination to identify whether the TB disease is active or inactive [37, 38].

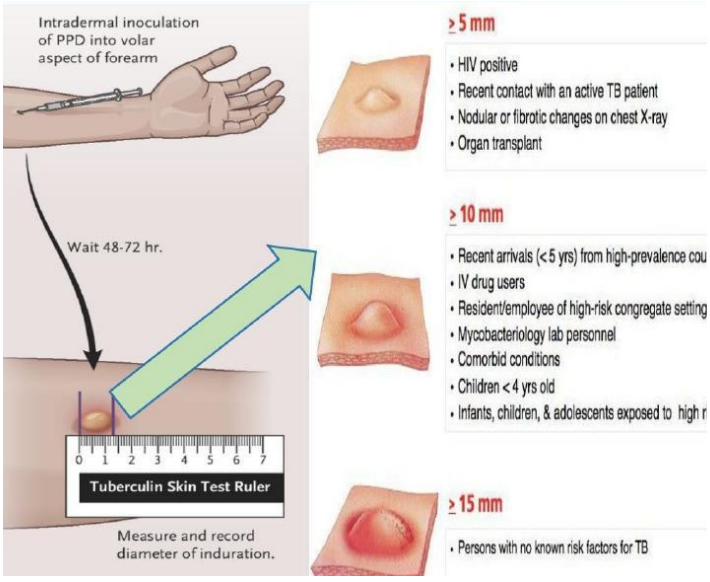


Figure 2.3: Tuberculin skin Test. It involves injecting a small amount of PPD, tuberculin, into the skin of the forearm and evaluating the reaction after 48-72 hours to confirm the presence of TB bacteria.

Bacteriological examination: Bacteriologic examination uses testing specimens such as sputum, urine, or cerebrospinal fluid. These tests include smear microscopy for detecting

acid-fast bacilli (AFB), nucleic acid amplification (NAA) for rapid identification of MTB and drug resistance, culture confirmation, and drug susceptibility testing through both molecular and growth-based techniques. These tests are critical for confirming TB and guiding appropriate treatment decisions [15].

Sputum and Xpert MTB/RIF: Smear microscopy, culture and Polymerase Chain Reaction (PCR) are used to examine sputum sample to confirm the presence of TB bacteria. Sputum smear microscopy has been the primary method for diagnosis of pulmonary TB in low and middle-income countries. A sputum sample is examined under a microscope to detect TB. Ziehl-Neelsen (ZN) or fluorescence staining techniques can be used to detect bacilli in the sputum sample. ZN staining is the most commonly used stain type in sputum smear, which is performed under bright field microscopy for identifying acid-fast bacilli (AFB) in sputum samples. Due to its unique lipid-rich cell wall, MTB does not absorb dyes used in standard Gram staining, making it difficult to detect using this common method [39]. Instead, MTB is identified using the Ziehl-Neelsen staining method, which highlights the bacterium in bright red against a blue background. This technique classifies it as an acid-fast bacillus (AFB) [16]. An example of this staining method is shown in Figure 2.4 [36].

Xpert MTB/RIF is a nucleic acid amplification test (NAAT) that works by targeting specific genetic sequences contained within the TB mycobacterium and amplifying them. During the amplification process, a fluorescent molecule is released and measured to determine the amount of initial DNA, which helps confirm a TB diagnosis [40].

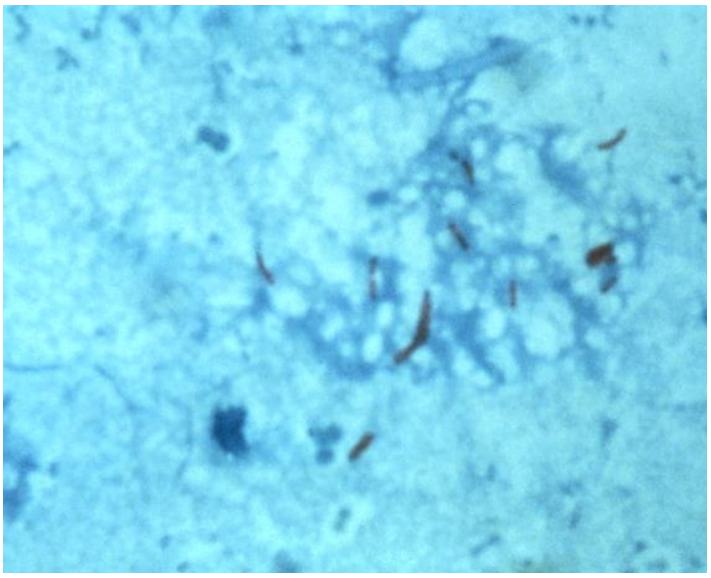


Figure 2.4: Acid-fast Ziehl-Neelsen stained specimen

2.5.3 X-ray

Chest x-ray, particularly posterior-anterior views, helps distinguish between latent infection and active pulmonary TB. Irregular patches in lungs such as lung lesions or cavities may suggest active TB. However, chest X-rays alone cannot confirm the diagnosis [35]. A nor-

mal X-ray can help suggest that a person without symptoms probably doesn't have active TB, even if they tested positive on a TB skin or blood test. Still, this approach isn't always reliable, especially in people with HIV or weakened immune systems, who might still have TB even if their X-ray appears normal. Despite these limitations, chest X-ray remains an important tool for screening of individuals from pulmonary TB. It helps guide the selection of candidates for bacteriologic testing and supports a clinical diagnosis in cases where bacteriologic confirmation is not possible [35].

2.6 Treatment of Tuberculosis

To reduce the burden of illness and death caused by TB, it is essential to prevent new infections with MTB and to stop the progression from infection to active disease. TB is a curable disease, typically treated with a standard six-month course of four antibiotics. Common first-line drugs include rifampicin and isoniazid [41]. The choice of treatment depends on the type of TB bacteria. TB can be caused by drug-susceptible or drug-resistant strains. Accordingly, there are two main categories of treatment: first-line treatment for drug-susceptible TB, and second-line treatment for drug-resistant TB [42].

In cases of latent TB infection, where a person is infected but not yet sick, TB preventive treatment can be given to stop the disease from developing. This treatment uses the same medications but over a shorter period. With new treatment options, the duration has been reduced to just 1 to 3 months, compared to the traditional 6-month course [41, 42].

When TB bacteria are resistant to standard drugs, the patient is diagnosed with drug-resistant TB. This form of TB requires longer and more complex treatment. If treatment is not properly completed, TB can become drug-resistant and continue to spread [42]. To improve treatment success, patients are provided with supervision, information, and support by trained health workers or volunteers. Without such support, completing the full course of TB treatment can be difficult [41, 42].

2.7 Cough Sound Analysis for TB Detection

Cough is a primary and clinically significant symptom of most respiratory diseases, and changes in cough characteristics provide valuable information on the structural nature of the respiratory airways for diagnosing respiratory diseases, particularly TB. It is a reflex action characterized by a rapid expulsion of air from the lungs to clear the airways, typically in response to irritation or inflammation in the respiratory tract. The acoustic properties of a cough, such as its duration, frequency content, intensity, and temporal structure, are influenced by the underlying pathology of the respiratory system. As a result, cough sounds can serve as non-invasive biomarkers for disease screening and monitoring [43].

Recent advances in machine learning have enabled the use of cough audio recordings for automated disease detection. In particular, digital cough recordings can be captured using standard microphones or smartphones, offering a low-cost and scalable diagnostic alterna-

tive, especially in low-resource settings. Analyzing cough sounds through computational models allows for the extraction of disease-specific acoustic patterns that may not be distinguishable by the human ear [44].

In the current study, cough sound is treated as a one-dimensional time-series signal sampled at 44.1 kHz. A deep learning architecture based on 1D CNNs is employed to automatically learn acoustic features directly from raw cough waveforms. Unlike traditional approaches that rely on manually engineered features such as mel-frequency cepstral coefficients (MFCCs), the developed model leverages end-to-end learning to preserve the full temporal and amplitude dynamics of the signal. This is particularly advantageous in distinguishing TB-induced coughs from those caused by other conditions.

The integration of cough analysis into TB screening methods provides a promising direction for remote and rapid diagnostic support. When combined with clinical metadata, as explored in this work, it has the potential to improve diagnostic accuracy and prioritize testing in resource-limited healthcare environments [43].

2.7.1 Basic Properties of Cough Sound

A. Sampling Rate

The sampling rate is defined as the number of samples captured per second from a continuous audio signal, measured in hertz (Hz) as defined by Equation 2.1. It determines the temporal resolution of the signal and the maximum frequency that can be accurately represented, according to the Nyquist theorem. A higher sampling rate enables better preservation of high-frequency content but results in increased data size [45].

The total number of samples N in a recording is calculated as:

$$N = f_s \times T \tag{2.1}$$

where f_s is the sampling rate in hertz (Hz), and T is the duration of the signal in seconds.

B. Nyquist Frequency

The Nyquist frequency is the highest sound frequency that can be accurately captured, and it is half of the sampling rate as indicated by Equation 2.2. This concept is critical because any frequency component above the Nyquist limit will be misrepresented in the digital signal [45].

$$f_{\text{Nyquist}} = \frac{f_s}{2} \tag{2.2}$$

where f_s is the sampling rate in hertz (Hz).

Example: If an audio with a sampling rate of 16 kHz, the recording can only accurately capture sounds up to 8 kHz. This means that any sound above 8 kHz is either lost or recorded incorrectly. So, in our study, the model is trained on audio recorded at a higher rate (44.1

kHz); using lower-rate recordings (16 kHz) may reduce performance because some important sound details are missing.

C. Time Resolution (Δt)

The time resolution, Δt, is the time interval between two consecutive audio samples as defined by Equation 2.3. It indicates how finely the audio signal is sampled in time [45]. A smaller Δt (higher sampling rate) means more frequent measurements, which provide greater detail about frequent sound changes such as sharp coughs or speech transitions. In contrast, a larger Δt (lower sampling rate) means less frequent measurements per second, which may miss fine details.

$$\Delta t = \frac{1}{f_s} \tag{2.3}$$

where f_s is the sampling rate in hertz (Hz). In the current study, cough audio is used at a sampling rate of 44.1 kHz, the time between samples is approximately $\Delta t \approx 22.68\,\mu$s.

D. Duration

Duration is the total length of the audio signal, typically measured in seconds. Duration directly affects the number of samples and is important for defining the input size of machine learning models. Fixed-length inputs, commonly ranging from 0.5 to 2 seconds, are often used for consistency in training [45]. The relation to Sampling Rate is defined by Equation 2.1.

2.8 Convolutional Neural Networks (CNNs)

Deep learning is a specialized subset of machine learning, which itself falls under the broader domain of artificial intelligence (AI). AI encompasses techniques and systems that aim to replicate or augment human tasks. Machine learning focuses on enabling systems to learn patterns and make decisions from data, while deep learning uses layered neural networks to automatically extract and transform complex features from large datasets [46].

CNNs represent a class of deep learning models widely used for analyzing structured data. While CNNs are best known for their success in image classification, their architecture is highly flexible and can be adapted to 1D or 2D inputs. In the context of audio processing, CNNs are commonly applied using one-dimensional convolutions (Conv1D) to extract features from raw waveforms inputs or preprocessed signals such as mel-spectrograms [47, 48]. CNNs are designed to automatically and hierarchically learn local patterns in data. In audio, this may include low-level features such as energy or pitch, and high-level patterns like phonemes or cough characteristics. A typical CNN is composed of a sequence of layers that extract, compress, and interpret features across time steps (see Figure 2.5) [46].

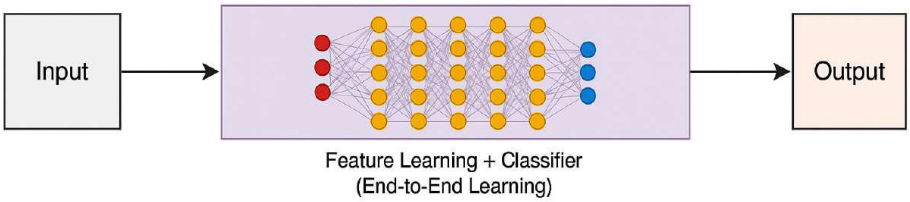


Figure 2.5: An end-to-end CNN pipeline where feature learning and classification are performed within a single model. The input is processed through multiple hidden layers for hierarchical feature extraction and ends with output prediction.

The basic operation of a 1D convolution layer can be expressed as in Equation 2.4:

$$y(t) = \sum_{k=0}^{K-1} x(t+k) \cdot w(k) \tag{2.4}$$

where $x(t)$ is the input signal, $w(k)$ is the kernel (filter), K is the kernel size, and $y(t)$ is the output feature map. This operation slides the kernel across the input and computes a weighted sum at each position, enabling sparse connectivity (local connections) and parameter sharing (same kernel applied everywhere).

2.8.1 The CNN Architecture

A basic CNN architecture consists of two important components, as shown in Figure 2.6), **feature extraction** and **classification** [49–51].

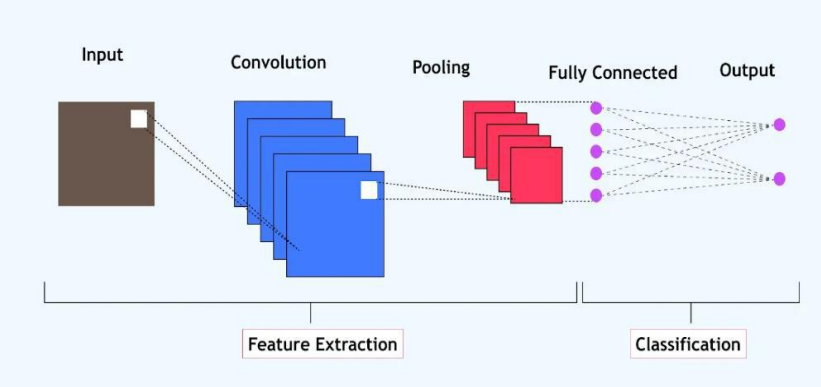


Figure 2.6: Basic CNN architecture is made up of two important components, feature extraction and classification. Feature extraction process consists of input, convolution, and pooling layers, and the second CNN architecture component, classification, consists fully connected layer and an output layer.

CNN Layers In CNN convolutional layer, pooling layer, fully connected layer, dropout layer, and activation functions work together to extract features and classify data efficiently [51].

- **Convolutional layers** use small, trainable filters that scan across the input to detect

features. These layers reduce the need for manual feature engineering and learn patterns directly from the data [46].

- **Pooling layers** are used to reduce the spatial or temporal size of feature maps (the output of a convolutional layer), improving robustness and reducing computation. MaxPooling selects the highest value from local regions, while GlobalAveragePooling averages the entire feature map to produce a compact representation, often used before dense layers [46, 51].
- **Fully connected layers** are typically used at the end of the network to integrate learned features and produce predictions [46].
- **Dropoutlayer** is a regularization method that randomly disables a portion of neurons during training, reducing the risk of overfitting [46, 52].
- **Activation functions** such as LeakyReLU, ReLU, sigmoid, or tanh introduce nonlinearity, enabling the network to learn complex mappings [46].
- **Batch normalization** stabilizes and accelerates training by normalizing the activations of each layer [46].

2.8.2 Training of CNN Model

Training of CNN model requires labeled input data, followed by a process of iteratively modifying its internal parameters to minimize the difference between the prediction and the actual label. The process of supervised learning allows the model to learn to recognize patterns and features in the data. It involves learning optimal weights for the filters and neurons to minimize the error between predicted and actual labels. This is done using backpropagation, combined with an optimization algorithm such as gradient descent [52].

A loss function is used to measure the discrepancy between predicted outputs and ground truth labels. Common loss functions include cross-entropy (for classification) and mean squared error (MSE) (for regression). The model computes the gradient of the loss with respect to each parameter and updates them as follows (Equation 2.5):

$$W_{t+1} = W_t - \alpha \cdot \frac{\partial L}{\partial W} \tag{2.5}$$

where W is a learnable weight, α is the learning rate, and L is the loss function.
In practice, mini-batch gradient descent is commonly used, which updates parameters based on small batches of training data. This improves training speed and stability. Optimizers such as Adam, SGD with momentum, and RMSprop are widely used to enhance convergence and generalization [53].

Chapter 3

Literature Review

Cough is one of the most prominent symptoms of respiratory diseases, including TB. Given the wide range of pathological and physiological conditions that can cause coughing, it is important to distinguish between the characteristic cough patterns associated with each illness. Conventional TB diagnostic methods tend to be costly and necessitate specialized knowledge and frameworks, which restricts their availability in resource-constrained environments. Therefore, there is a growing demand for alternative diagnostic approaches that support early and accessible TB detection. Recent advancements in deep learning and audio signal processing have led to the development of TB screening systems based on the analysis of patients' cough sounds. In addition to audio signal analysis methods, some studies have explored the integration of clinical metadata with cough sound features to improve diagnostic accuracy. This literature review presents an overview of key studies in both categories which are only cough sound analysis method and those integrate both audio analysis and clinical metadata for TB prediction using machine learning and deep learning approaches.

The literature review of related works was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The initial search used a combination of relevant keywords to identify studies published since 2018. Research papers were collected from PubMed, Google Scholar, and IEEE Xplore, resulting in an initial collection of 288 records. After removing 14 duplicates, 29 papers were screened based on their titles and abstracts. A total of 16 studies were finally selected based on inclusion and exclusion criteria (the type of respiratory disorder, the dataset used, and the algorithms used). Figure 3.1 provides a visual summary of this article selection process.

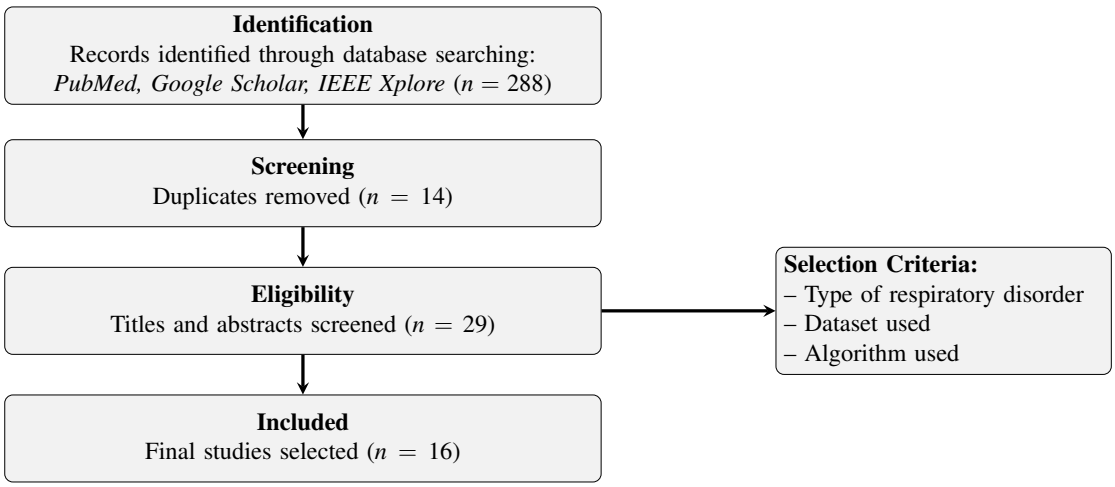


Figure 3.1: PRISMA Flow Diagram for Article Selection.

Jember *et al.* conducted a method for TB screening based on the analysis of patients' cough sounds. Their study utilized a dataset comprising 6,476 cough sound events to classify coughs as TB or non-TB. Mel-frequency cepstral coefficients (MFCCs) were extracted as audio features, and two machine learning algorithms artificial neural networks (ANN) and support vector machines (SVM) were applied for classification. The performance of these models was evaluated using accuracy and F1-score metrics. The ANN model achieved an accuracy of 92.3% and an F1-score of 87.7%, while the SVM model attained an accuracy of 89.49% and an F1-score of 82.2%. These results indicate that ANN outperformed SVM in handling the high-dimensional and non-linear nature of cough sound data, demonstrating its effectiveness for TB classification tasks [7]. Similarly, Suda *et al.* conducted a machine learning-based method for early TB detection using cough audio recordings collected in real-world conditions. The study used a dataset consisting of 724,694 cough samples from 1,105 adult participants (aged 18 and older) across seven high TB-burden countries. Coughs were recorded using the Hyfe Research smartphone app. Spectral features such as MFCCs and mel-spectrograms were extracted and used to train models including 2D CNNs, achieving an AUROC of 88%, meeting WHO triage tool criteria [54].

Another study investigated TB diagnosis using 1,358 forced cough recordings collected in a clinical setting. Acoustic features such as MFCCs, log-filter bank energies, zero-crossing rate (ZCR), and kurtosis were extracted. Machine learning classifiers including logistic Regression (LR), k-Nearest neighbors (KNN), SVM, multilayer merceptron (MLP), and CNN were evaluated using nested cross-validation. Logistic regression (LR) achieved the best performance with 93% sensitivity, 95% specificity, and an AUC of 0.94 when combined with sequential forward selection (SFS) for feature selection. The study emphasized the practicality of interpretable, low-complexity models like LR for deployment in resource-limited settings [9].

Kafentzis et al. applied statistical machine learning models for TB prediction using real-world cough audio recordings in combination with clinical metadata. The statistical classifiers were applied with and without clinical metadata using spectral and temporal domain features. Their dataset included 9772 cough recordings with accompanying demographic and clinical information. Their model achieved an AUC of 0.70 ± 0.05 using only cough audio, which improved to 0.81 ± 0.05 with the inclusion of clinical metadata. This study highlights the value of integrating clinical information with audio features to enhance TB screening [55].

Yellapu and colleagues created a model to screen and prioritize individuals at risk of pulmonary TB using distinct cough sounds and clinical data. Their study included 567 participants (278 TB-positive, 289 TB-negative). Two models were used: a CNN based on mel spectrograms and a fully-fusible CNN (FFANN) using both time- and frequency-domain features along with clinical information. The final prediction was made via a merged layer

combining both outputs. This model achieved 90.36% sensitivity, 84.67% specificity, and 86.82% overall accuracy [56].

Audiovisual multimodal cough data analysis for TB detection explores the use of deep learning models on cough audio recordings to develop a non-invasive tool for TB screening. The study used over 502,000 cough samples from 1,105 individuals across seven countries, integrating both audio features (MFCCs, mel-spectrograms) and clinical metadata. Among various models tested, a 1D CNN trained on MFCC features achieved 91% accuracy, surpassing the WHO benchmark for TB screening tools [57].

Table 3.1 summarizes the additional studies reviewed, highlighting the methods, datasets, features, results, and gaps observed.

Table 3.1: Literature reviewed on TB detection using cough sound analysis

Authors, Ref	**Samples**	**Method**	**Features**	**Results**	**Gaps**
Frost et al, [58], 2022	74 patients, 1,564 coughs	Logistic regression, BiLSTM, BiLSTM-Att and BiLSTM-Att + SFS	Mel-spectrograms, LFB, MFCCs	BiLSTM-Att + SFS (Sens: 0.778, Spec: 0.813, Acc: 0.800 and AUC: 0.85)	No clinical metadata used, limited dataset size and lacks geographic and demographic diversity and Feature engineering required
Pahar et al, [59], 2022	354 TB+ and 881 TB- coughs from 48 patients	LR, SVM, KNN, MLP, NLP embeddings + CNN and LSTM	MFCCs, ZCR, kurtosis	AUC: 0.81 (LSTM), 0.72 (CNN)	Limited dataset size and lacks geographic and demographic diversity, no clinical metadata used, and Feature engineering required
Xu et al. [60], 2024	70 TB+ patients (230 coughs) and 74 TB- patients (226 coughs)	Bi-LSTM, Bi-GRU	MFCC, ZCR, short-time energy, root mean square, and chroma_cens	Bi-LSTM (with feature fusion) Acc: 96.33%, Sens: 98.13%, Spec: 94.99%	Dataset is relatively small and lacks geographic and demographic diversity, no clinical metadata used, and Feature engineering required

Authors, Ref	Samples	Method	Features	Results	Gaps
Mahmood et al. [61], 2024	145 participants (435 samples) with clinical data	SVC (SVM), LR, Random Forest, MLP, KNN	Spectral and time domain (MFCC, ZCR, etc) metadata (age, smoking history, and TB exposure)	Acc: 97% (Random Forest) using cough with demographic and clinical factors	Feature engineering required, dataset is relatively small and lacks geographic and demographic diversity
Bao et al. [62], 2023	1,000 cough segments	ResNet50, GoogLeNet, Bi-LSTM (ensemble)	Not specified	Acc: 98.05%, Sens: 98.82%, Spec: 97.49%	Dataset is relatively small, no clinical metadata used
Knight, [63], 2024	74 patients, 1,564 coughs	ResNet18, LR, BiLSTM SkipNet	LFB, MFB, MFCC	ResNet-18 with FSS AUC: 77.49%	Small dataset and lacks geographic and demographic diversity, moderate performance, no clinical metadata used, Feature engineering required
Botha et al. [64], 2018	38 subjects (746 coughs) and 5 clinical data (MUAC, temperature, BMI, etc)	Logistic regression	MFCC, Log spectral energy	Acc: 82%, Sens: 95%, Spec: 72%, AUC: 0.95	Small and low geographic and demographic diversity dataset, audio record in controlled environment, feature engineering required
Rajasekar et al. [65], 2024	9,772 coughs, 1,105 individuals from CODA TB DREAM Challenge Dataset	CapsuleNet + FCNN, CNN, VGG16, ResNet50	Spectral images, HOG features	CapsuleNet + FCNN achieves Acc: 97%, Sens: 98%, Spec: 96%, F1: 0.97	Needs audio to image conversion, feature engineering required

Authors, Ref	Samples	Method	Features	Results	Gaps
Xu et al. [66], 2023	345 participants, 1,323 coughs	DMRNet (ResNet-based CNN with PSA/MHSA)	Mel spectrograms	Acc: 94.32%, Sens: 97.73%, Spec: 99.43%	Controlled setting, moderate dataset and lacks geographic and demographic diversity, no clinical metadata used, feature engineering required
Yuan et al. [67], 2023	345 samples	CapsuleNet + Feedforward NN + CNN	MFCC, Mel spectrograms	AUC: 0.9375 F1: 91.74%	Small dataset and lacks geographic and demographic diversity, feature engineering required

The reviewed studies collectively highlight the potential of cough sound analysis for TB diagnosis and contribute significantly to the evolution of machine learning and deep learning approaches. However, they also reveal different research gaps.

- **Small datasets:** Many studies used a limited number of samples or patients, reducing the generalizability of the models.
- **Lack of geographic and demographic diversity of data:** Data often came from similar population groups, which limits model applicability to broader populations.
- **Limited use of clinical metadata:** many studies did not incorporate relevant patient information (e.g., age, symptoms, smoking history), which could improve predictive performance.
- **Need for feature engineering:** Some studies depend on manually extracted features instead of learning representations automatically, which may not capture all relevant patterns. Additionally, feature engineering needs deep audio signal processing and domain knowledge, unlike a model that uses raw waveforms directly.
- **Moderate model performance:** Certain studies reported relatively lower accuracy or AUC compared to others, potentially due to the above limitations.

The current study aims to address these gaps by developing a robust, generalizable, and clinically adaptable TB prescreening model based on cough sound and patient-specific clinical information.

Chapter 4

Materials and Methods

The proposed method in the current study uses a two-stage CNN architecture to predict TB from cough audio and clinical information. As illustrated in Figure 4.1, the overall methodology involves collecting cough sound data from TB-positive and TB-negative individuals, preprocessing and splitting the data, and initially training a CNN model to learn audio-based features associated with TB status. In the second stage, clinical information is integrated with the learned audio features through dense layers to enhance prediction accuracy. The model is evaluated on a separate test dataset and prepared for real-world deployment.

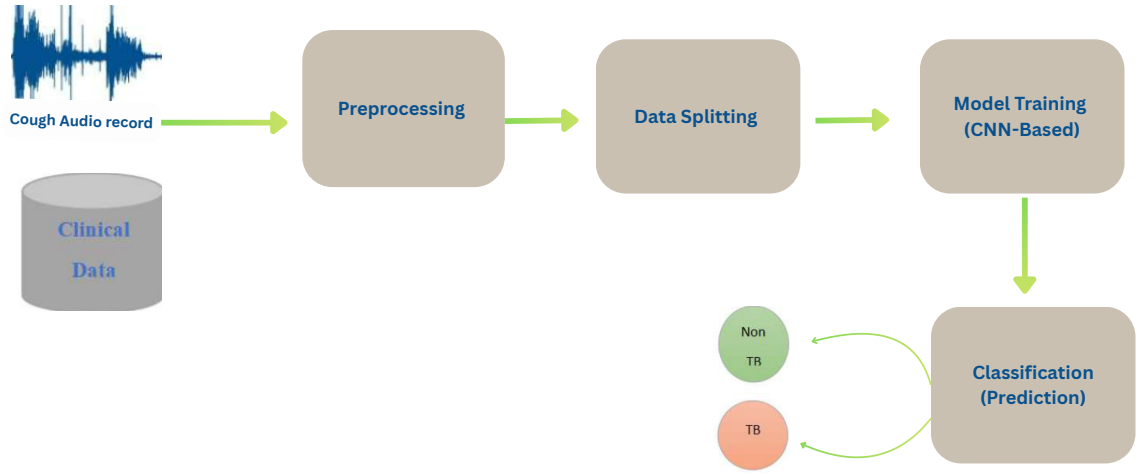


Figure 4.1: Block diagram of the proposed method.

4.1 Data Collection

4.1.1 Data Sources

The dataset used in the current study originates from the CODA TB DREAM challenge. The dataset consists of cough sound recordings and corresponding patient-specific clinical metadata collected from 1,105 individuals across seven countries: India, Philippines, South Africa, Uganda, Vietnam, Tanzania, and Madagascar. While the entire dataset was developed to support the CODA TB challenge for algorithm development in TB diagnosis, only the training portion has been made publicly available for post-challenge research use. This study exclusively utilizes the accessible total of 29,768 cough audio samples from training dataset. The held-out validation and test sets remain confidential and were not used in this work [13].

Access to the dataset is governed by the Synapse commons platform and is subject to a formal request process designed to ensure ethical and responsible data usage. Researchers must first create a Synapse account and become both certified and validated users. This involves passing a certification quiz, linking a public ORCID profile, submitting the Synapse pledge, and providing recent identity

attestation (such as a notarized letter or official institutional documentation). Once validated, users must submit institutional data use statement (IDU) describing their research objectives and analysis plan, which is shared publicly. The validation requirements are supported by reference files listed in **Appendix C** [68].

Cough sounds were collected by using Hyfe research mobile application. The cough recordings were sampled at a frequency of 44.1 kHz and included both solicited (instructed) and spontaneous coughs. Structured clinical metadata was collected through participant surveys and diagnostic procedures. This metadata included demographic information, symptoms, vital signs, and TB diagnostic results based on sputum collection for TB testing [13, 55, 68]. Table 4.1 summarizes the demographic characteristics of the participants included in the dataset used for this study. A full list of the available demographic, clinical, and microbiologic variables is provided in **Appendix A** (Table A.1) [13].

Table 4.1: Participant demographics for the dataset (N=1105)

Variable	**Category**	**Dataset (N=1105)**
Sex	Female	517 (46.8%)
	Male	588 (53.2%)
Age (years)	Median [Min, Max]	39.0 [18.0, 85.0]
Height (cm)	Median [Min, Max]	162.0 [138.0, 198.0]
Weight (kg)	Median [Min, Max]	55.0 [30.0, 158.8]
Duration of cough (days)	Median [Min, Max]	30.0 [4.0, 365.0]
TB Status	Negative (0)	808 (73.1%)
	Positive (1)	297 (26.9%)

Additional Real-World Dataset Collected Locally from Ethiopia :

Table 4.2: Participant demographics for the local dataset (N=12)

Variable	**Category**	**Dataset (N=12)**
Sex	Female	3 (25.0%)
	Male	9 (75.0%)
Age (years)	Median [Min, Max]	32 [18, 61]
Height (cm)	Median [Min, Max]	165 [152, 182]
Weight (kg)	Median [Min, Max]	55 [38, 70]
Duration of cough (days)	Median [Min, Max]	21 [10, 90]
TB Status	Negative (0)	3 (25.0%)
	Positive (1)	9 (75.0%)

In addition to the CODA TB DREAM Challenge dataset, this study includes a dataset collected from three Ethiopian health facilities: ALERT Hospital, St.Peter Hospital, and Dilfre Health Center, to evaluate the post-deployment model's performance in a real-world setting. Cough audio recordings

were collected from suspected TB patients using a mobile application, *cough detector and recorder*. The cough recordings from this App recorder were sampled at a frequency of 16 kHz. Clinical metadata was collected in parallel using a structured questionnaire that mirrored the variables present in the original CODA dataset, including demographic information, symptom history, and vital signs. In total, data from 12 patients (a total of 12 recordings) were collected and included in the post-deployment evaluation. This deployment-level data collection was conducted in collaboration with local hospital/health center staffs and followed ethical data handling protocols.Table 4.2 summarizes the demographic characteristics of the participants included in the local dataset.

This dataset was used solely for testing the deployed model to assess its generalization capability outside the original training dataset (observational purpose). The questionnaire, collection protocol, and support letter used for data collection are presented in **Appendix B**.

4.2 Data Preprocessing

Preprocessing is a critical step in the success of machine learning models, especially when working with heterogeneous real-world collected health data such as clinical information and audio data. Before model training, both structured clinical data and cough audio recordings were scaled and encoded to ensure consistency and compatibility with neural network input requirements. This step helps create a strong foundation for learning meaningful patterns related to TB detection [69, 70]. The key preprocessing steps are outlined below:

1. **Clinical Data Preprocessing**
   The following preprocessing steps were applied to the clinical information after merging the metadata files, including categorical encoding, missing value imputation, and scaling.

   (a) **Metadata Merging**
   The metadata is a collection of three different comma-separated values (CSV) files containing both clinical and audio information. These files were combined using a common participant identifier. First, two CSV metadata files were created by merging the clinical metadata with audio metadata (solicited and longitudinal) separately. These two merged datasets were then concatenated into a single combined data table. Each row in the resulting CSV file includes a filename reference that corresponds to a separate audio file stored in a ZIP archive.

   (b) **Categorical Encoding**
   This transformation is used to convert string categories into numerical values to make data suitable for analysis or training machine learning models. In the current study, categorical columns such as sex, fever, hemoptysis, etc., were encoded using LabelEncoder from the sklearn preprocessing module. Before encoding, missing categorical values were filled using the mode (i.e., the most frequently occurring value) of each respective column. To prevent data leakage, encoders were fitted exclusively on the training data and then applied those same encoders to transform the validation and test data.

   (c) **Numerical Imputation**
   Numerical Imputation is the process of filling in missing values in numerical data. For

numerical features such as age, weight, heart rate, etc., missing values were filled using the mean of the corresponding feature, calculated from the training set. These values were then applied to the validation and test sets to ensure statistical consistency while preventing data leakage. This approach ensures that missing values are filled with statistically consistent values, decreasing the impact of missing data on model training.

(d) **Clinical Data Scaling**

Scaling is an important step to take before training a model to ensure that the dataset is within the same scale, since datasets mostly have varying scales, which can affect the model's performance. Therefore, to ensure that all features contribute equally to the model, common scaling methods like normalization and standardization are used, including min-max normalization, Z-score normalization, and log transformation to scale data [71]. Among these methods, standardization, which is also known as Z-score normalization, was applied in the current study using the `StandardScaler` method to maintain numerical stability and bring all clinical data to a common scale. The scaler was fitted exclusively on the training data and then applied to the validation and test sets to ensure consistent scaling without introducing information leakage. This transformation adjusted each feature to have a mean of zero and a standard deviation of one, as described by the Z-score formula in Equation 4.1 [70]:

$$z = \frac{x - \mu}{\sigma} \tag{4.1}$$

where:

- x is the original feature value,
- μ is the mean of the feature,
- σ is the standard deviation of the feature,
- z is the standardized (normalized) value.

2. **Audio Preprocessing**

Cough audio files were provided in three ZIP archives (`solicated_coughs.zip`, `longitudinal_1.zip`, and `longitudinal_2.zip`), were processed individually. Each file was loaded using the `soundfile` library. The original audio durations ranged from 0.340 to 0.500 seconds, with an average of approximately 0.499 seconds.

(a) **Audio Clipping and Padding:** To ensure consistency, all audio samples were either trimmed or zero-padded to a fixed length of 0.5 seconds, which corresponds to 22,050 samples at a 44.1 kHz sampling rate. This sampling rate preserves high-frequency content (up to 22.05 kHz, the Nyquist frequency) and provides a time resolution of approximately 22.68 μs, which helps to capture rapid acoustic events such as cough bursts.

(b) **Memory Optimization:** Due to the high dimensionality of raw audio data, all arrays were explicitly cast to `float32` format. This reduced memory usage and improved processing efficiency during model training. After conversion, the training audio data occupied approximately 1.8 GB of memory.

(c) **Audio Standardization:** Each audio waveform was standardized using the Z-score method by subtracting the mean and dividing by the standard deviation. This resulted in input tensors of shape `(22050, 1)`, suitable for one-dimensional CNN model. The waveform was standardized using the Z-score method, as defined in Equation 4.1 [70, 72, 73].

3. **Data Splitting**

   To evaluate model generalization while preventing data leakage, the combined metadata was first split into training (90%) and test (10%) subsets using stratified sampling based on the target label (`tb_status`). The training set was then further divided into training (80%) and validation (20%) subsets, again using stratified sampling to maintain the class distribution. This resulted in **21,432 training samples**, **5,359 validation samples**, and **2,977 test samples**. Stratified sampling at both stages ensured that the proportion of TB-positive and TB-negative cases remained consistent across all subsets [74].

4.3 Model Selection

To benchmark the effectiveness of *CNNbasedNet* in TB detection, its performance was compared against custom-made baseline models—*LightResNet* and *LightEfficient*—inspired by the simplified designs of ResNet and EfficientNet, respectively. LightResNet includes skip connections to support more stable and efficient training, while LightEfficient includes a lightweight attention block to enhance feature representation. All models process standardized 1D audio and clinical data input, enabling multimodal learning. All models were trained using early stopping and learning rate scheduling to improve generalization. They were evaluated using clinically relevant performance metrics, including accuracy, precision, recall, F1-score, AUC, and average inference time per sample.

Additionally, MetforNet, a top-performing model from the CODA TB DREAM Challenge, was trained on a subset of the publicly accessible CODA dataset, consisting of 9,772 cough recordings from 1,105 participants. In contrast, the developed CNNbasedNet model was trained on a larger portion of the same publicly accessible dataset, totaling 29,768 cough recordings. While MetforNet integrates a complex architecture combining five CNN blocks, BiGRUs, and attention mechanisms, CNNbasedNet model develops a distinct, simplified CNN-based architecture with three CNN blocks, global average pooling, and dense layers by reducing CNN blocks, removing BiGRUs, and attention layers, which were used by MetforNet. Therefore, this significantly reduces training time and computational load. Due to the high complexity of MetforNet (requiring approximately 8 days to train), retraining it on the larger dataset was not feasible. Thus, while the model comparison provides valuable insights into architectural efficiency and scalability, differences in training dataset size should be taken into account when interpreting the results.

4.4 The Developed Model

This section describes the developed deep learning model, which employs a two-stage architecture based on CNNs. It adopts a multimodal approach that combines unstructured audio signals with

structured clinical metadata, allowing the model to capture more informative features and complementary information that may be missed when depending on a single data source. The 1D convolutional layers extract temporal patterns from the raw cough audio, which are relevant for identifying pathological conditions. In parallel, the clinical input pathway processes structured patient-specific metadata through fully connected layers to generate meaningful feature embeddings.

To enhance model performance and generalization, batch normalization is used after each major layer block to stabilize learning, while the Leaky ReLU activation function prevents the issue of dying neurons, ensuring continuous gradient flow. Additionally, the integration of dropout layers throughout the architecture provides regularization, reducing the risk of overfitting problems [48]. The final fusion of audio and clinical features enables the model to generate well-informed predictions by leveraging both acoustic and contextual clinical information. Overall, the architecture is both lightweight and efficient, making it well-suited for deployment in resource-constrained environments, such as mobile health platforms or remote diagnostic tools. The overall architecture of the proposed two-stage model is illustrated in Figure 4.2, showing the parallel processing of audio and clinical inputs, the fusion mechanism, and the final prediction path.

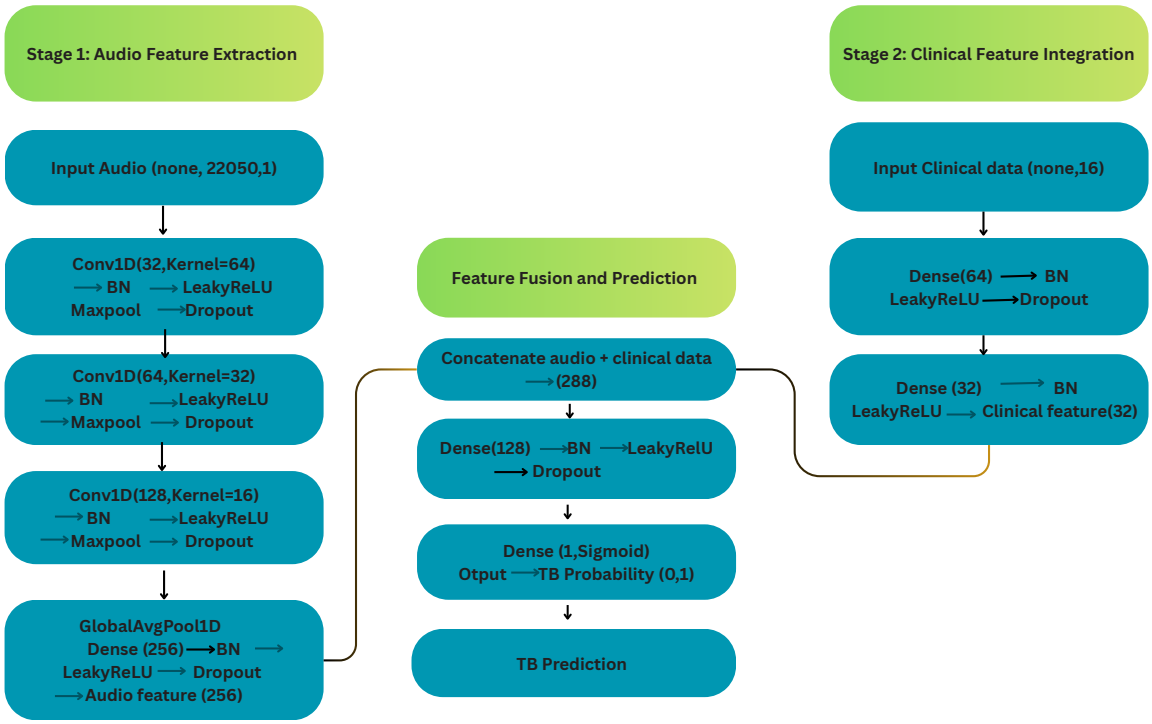


Figure 4.2: Block diagram of the proposed two-stage TB prediction model architecture.

The model accepts two types of input:

- **Cough audio recording:** 1-dimensional time-series data representing 0.5 seconds of mono audio, sampled at 44.1 kHz.
- **Structured clinical features:** A vector of **16** patient-specific attributes, including demographic information, symptom data, and vital signs.

The developed model is conceptually inspired by MetforNet, the top-performing model from the

CODA TB DREAM challenge. Both CNNbasedNet and MetforNet follow a multimodal design that combines convolutional processing of cough audio with clinical data integration for TB prediction. Each model uses 1D convolutional layers for audio feature extraction and merges the learned audio representations with clinical inputs through dense layers. However, CNNbasedNet was designed with a simplified and distinct architecture compared to MetforNet121, reducing the number of CNN blocks and removing recurrent and attention layers to lower complexity and training time.

4.4.1 Developed CNN Model Architecture

Stage 1: Base Model (CNNbasedNet121)

In the first stage, base model (CNNbasedNet121) takes input tensors of shape `(None, 22050, 1)`, representing 22,050 number of audio samples in 0.5-second (mono audio recordings) sampled at 44.1 kHz. Its architecture consists of three sequential convolutional (Conv1D) blocks, each followed by batch normalization, LeakyReLU activation, MaxPooling, and Dropout for robust feature extraction. A GlobalAveragePooling layer and a Dense(256) layer are used to generate a 256-dimensional audio feature vector, followed by a sigmoid output for binary classification. The model is trained using the Adam optimizer (learning rate = 0.001, clipnorm = 1.0), with binary cross-entropy loss, and evaluated using accuracy and AUC metrics.

- **Block 1:**
  - `Conv1D (32 filters, kernel size=64)`
  - `BatchNormalization`
  - `LeakyReLU`
  - `MaxPooling1D (pool size=8)`
  - `Dropout (rate=0.2)`
- **Block 2:**
  - `Conv1D (64 filters, kernel size=32)`
  - `BatchNormalization → LeakyReLU → MaxPooling → Dropout (rate=0.2)`
- **Block 3:**
  - `Conv1D (128 filters, kernel size=16)`
  - `BatchNormalization → LeakyReLU → MaxPooling → Dropout (rate=0.2)`
- **GlobalAveragePooling1D**
- **Dense Layers:**
  - `Dense (256 units)` → `BatchNormalization → LeakyReLU → Dropout (rate=0.5)`
  - `Dense (1 unit, activation=sigmoid)`

Stage 2: Enhanced Model (CNNbasedNet)

In the second stage, enhanced model (CNNbasedNet) combines the audio features from the pre-trained base model with 16-dimensional clinical features. The clinical input of shape `(None, 16)` passes through two Dense layers (64 and 32 units) with BatchNorm, LeakyReLU, and Dropout. Both branches are concatenated to form a 288-dimensional feature vector, followed by a Dense (128) layer and a final sigmoid output. It is optimized using Adam with a reduced learning rate (0.0005) to allow fine-tuning.

- `Dense (64 units)` → `BatchNormalization → LeakyReLU → Dropout (rate=0.3)`
- `Dense (32 units)` → `BatchNormalization → LeakyReLU`

The outputs from the audio feature extractor `(None, 256)` (after the Dense (256) layer in Stage 1) and clinical feature module `(None, 32)` are concatenated into `(None, 288)`. This vector is then processed as follows:

- `Dense (128 units)` → `BatchNormalization → LeakyReLU → Dropout (rate=0.5)`
- `Dense (1 unit, activation=sigmoid)`

Output Layer

The sigmoid output layer produces a probability value indicating the likelihood of TB.

Parameter Summary

The parameter summary of the proposed model, including both the base and enhanced stages, is presented in Table 4.3.

Table 4.3: Model parameter summary (combined base and enhanced stages)

Component	**Value**
Trainable Parameters	273,601
Non-Trainable Parameters	1,408
Activation Functions	LeakyReLU, Sigmoid
Regularization	Dropout (rates = 0.2, 0.3, 0.5)
Normalization	Batch Normalization after major layers

Training Configuration

Both the base and enhanced models were trained using Keras' model.fit() function, which automates the training process by handling data input, loss computation, and weight updates over multiple epochs. Training was performed with a batch size of 16. Regularization and training control techniques such as early stopping, learning rate scheduler, and model checkpoint to prevent overfitting, improve convergence, and preserve best results, respectively. Table 4.4 outlines the key training hyperparameters used during model development, including optimization settings, regularization, and early stopping criteria.

Table 4.4: Training hyperparameters

Parameter	**Value**
Optimizer	Adam
Learning Rate	0.0005 (enhanced), 0.001 (base)
Loss Function	Binary Crossentropy
Batch Size	16
Epochs	50 (with early stopping)
Dropout Rate	0.2 (Conv blocks), 0.3 (clinical), 0.5 (fusion)
Early Stopping Patience	10 epochs

The detailed layer-wise architecture of the proposed two-stage model, including output shapes and parameter counts, is summarized in Table 4.5.

Table 4.5: Layer-wise architecture summary of the proposed model

Layer	Output Shape	Parameters
Input (Audio)	(None, 22050, 1)	0
Conv1D (32, kernel=64)	(None, 11025, 32)	2,080
BatchNormalization	(None, 11025, 32)	128
LeakyReLU + MaxPooling1D + Dropout	(None, 1378, 32)	0
Conv1D (64, kernel=32)	(None, 689, 64)	65,600
BatchNormalization	(None, 689, 64)	256
LeakyReLU + MaxPooling1D + Dropout	(None, 86, 64)	0
Conv1D (128, kernel=16)	(None, 43, 128)	131,200
BatchNormalization	(None, 43, 128)	512
LeakyReLU + MaxPooling1D + Dropout	(None, 5, 128)	0
GlobalAveragePooling1D	(None, 128)	0
Dropout (audio)	(None, 128)	0
Dense (256, audio)	(None, 256)	33,024
BatchNormalization + LeakyReLU	(None, 256)	1,024
Input (Clinical, 16 features)	(None, 16)	0
Dense (64, clinical)	(None, 64)	1,088
BatchNormalization + LeakyReLU + Dropout	(None, 64)	256
Dense (32, clinical)	(None, 32)	2,080
BatchNormalization + LeakyReLU	(None, 32)	128
Concatenate (audio + clinical)	(None, 288)	0
Dense (128, fusion)	(None, 128)	36,992
BatchNormalization + LeakyReLU + Dropout	(None, 128)	512
Dense (1, sigmoid output)	(None, 1)	129
Total Parameters		**275,009**
Trainable Parameters		**273,601**
Non-trainable Parameters		**1,408**

Regularization and Training Control Techniques:

- **EarlyStopping:** Stops training when validation loss stops improving to prevent overfitting. A patience of 10 epochs is used.
- **Learning Rate Scheduler (ReduceLROnPlateau):** reduces the learning rate when the model's performance stops improving, allowing finer weight updates and better convergence. Learning rate was reduced by half if the validation loss didn't improve for 5 epochs.
- **ModelCheckpoint:** Automatically saves the best model based on validation performance during training.
- **Dropout:** Dropout was used in different layers of the model to prevent overfitting by randomly deactivating some neurons during training, helping the model learn more general patterns.

Loss function: A Binary Cross-Entropy is used for binary classification. It measures how well the predicted probability matches the true label (0 or 1).

Optimizer: An optimizer is an algorithm that helps the model minimize the loss by adjusting its weights during training based on the error it makes, to achieve better predictions. Both the base and enhanced models were trained using the Adam optimizer, with learning rates of 0.001 and 0.0005, respectively, and gradient clipping (clip norm = 1.0) to prevent exploding gradients and ensure stable training.

Class Weights: Class Weighting was applied to address class imbalance in the training data by assigning a higher weight to underrepresented class samples. This ensured the model learned fairly from both classes during training and improves the overall classification performance. The weights were automatically calculated using `compute_class_weight` from `sklearn`.

4.5 Evaluation Metrics

To evaluate the performance of the proposed model in binary classification of TB, a range of standard evaluation metrics were used. These metrics are particularly relevant for medical diagnosis tasks, where both false negatives and false positives can have serious consequences.

Accuracy: Accuracy represents the proportion of correct predictions (both true positives and true negatives) among all predictions (Equation 4.2). [75]

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

Precision: Precision measures the proportion of true positive predictions among all predicted positive cases [75], as shown in Equation 4.3:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.3}$$

Recall (Sensitivity): Recall evaluates the model's ability to correctly identify actual TB-positive cases [75], as shown in Equation 4.4:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.4}$$

Specificity Specificity quantifies the model's ability to correctly classify TB-negative cases [75], as shown in Equation 4.5:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4.5}$$

F1-Score The F1-score is the harmonic mean of precision and recall, providing a balance between the two [76] [77], as shown in Equation 4.6:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.6}$$

AUC-ROC A ROC curve helps visualize how well the model can separate TB-positive cases from TB-negative independent of any specific threshold. This is particularly valuable in medical diagnostics, where class imbalance or threshold sensitivity may affect decision-making. The Area Under the Curve (AUC-ROC) provides a single scalar summary of the ROC curve. The ROC reflects the model's ability to distinguish between classes across various threshold settings, which is collective measure of sensitivity (TP/(TP+FN)) and specificity (TN/(TN+FP)) over a wide range of all possible

threshold values [78], it is mostly used performance evaluations metric for machine learning due to its consistency when dealing with imbalanced datasets [79] [80].

Confusion Matrix A confusion matrix are the major mean to evaluate errors in classification and visualize comprehensive model prediction performance in terms of [81]:

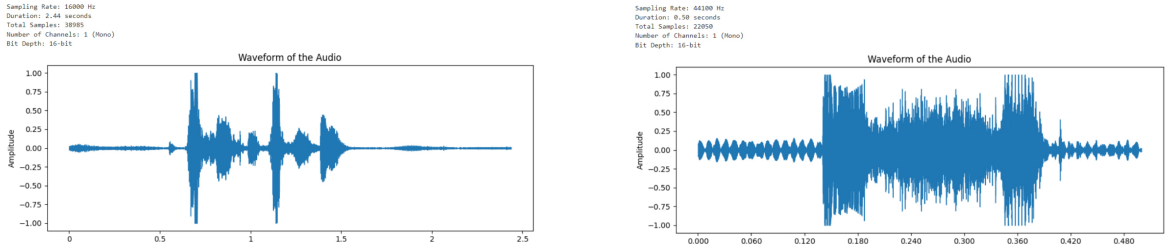
- **TP** – True Positives
- **FP** – False Positives
- **TN** – True Negatives
- **FN** – False Negatives

4.6 User Interface Deployment

Front-end deep learning web apps development and deployment have been introduced to make the model more accessible and usability to end-users using TB prediction pre-trained CNN model. A graphical user interface (GUI) was developed using Streamlit—an open-source Python framework designed for building interactive machine learning web applications with minimal effort [82]. Google Cloud Run platform was used to deploy and host the Streamlit-based TB prescreening web interface, ensuring public accessibility.

For further assessing real-world applicability, the deployed model was tested using a separate dataset collected from Ethiopian healthcare facilities. Patient cough audio was recorded using a mobile application called Cough Detector and Recorder, and clinical data was collected through a structured questionnaire.

The GUI preprocessing steps for both audio and clinical data are consistent with those used during training. Additionally, the GUI resamples input audio from 16 kHz to 44.1 kHz to match the model's expected sampling rate and applies an energy-based sliding window to extract the loudest 0.5-second segment, improving robustness to variable input quality.Figure 4.3 shows the transformation involving resampling and extraction of the loudest 0.5-second segment. These preprocessed audio and clinical inputs were fed into the Streamlit interface to generate predictions. The deployment results confirmed the model's practical effectiveness in real-world LMIC settings, particularly in Ethiopia.



(a) Audio (16 kHz, 2.44s) (b) Audio (44.1 kHz, 0.5s)

Figure 4.3: Waveform of audio sampled at 16 kHz and audio sampled to 44.1 kHz and a 0.5-second segment.

4.7 Materials

Hardware used includes an AMD Ryzen 5 5625U laptop with 8GB RAM for model training and data analysis. Multiple mobile devices with a cough detector and recorder app are used for cough audio recording in WAV format. Software components include Python 3.11.11 for implementation, streamlit for developing the web interface, and Google Cloud Run for application deployment.

4.8 Code Implementation Details

The model architecture, data preprocessing steps, and training scripts used in this study have been fully implemented and documented. The complete source code is publicly available on GitHub: GitHub Repository.

Chapter 5

Results and Discussion

5.1 Data Visualization

Data visualization is crucial in the development of machine learning models. It provides intuitive, graphical representations of feature distributions, relationships, and patterns, which are essential for understanding the underlying structure of the dataset. Data visualizations help identify outliers, anomalies, and class imbalances that may not be clear from statistical summaries alone. For this study, 16 clinical features were explored: `sex`, `age`, `height`, `weight`, `reported_cough_dur`, `tb_prior`, `tb_prior_Pul`, `tb_prior_Extrapul`, `tb_prior_Unknown`, `hemoptysis`, `heart_rate`, `temperature`, `weight_loss`, `smoke_lweek`, `fever`, `night_sweats`. The four features, namely, `tb_prior`, `tb_prior_Pul`, `tb_prior_Extrapul`, and `tb_prior_Unknown`, indicate known previous TB cases, known previous pulmonary TB cases, known previous extra pulmonary TB cases, and unknown cases, respectively while `smoke_lweek` indicates a 1 week smoking history.

A. Descriptive Statistics

Basic statistics including mean, standard deviation, maximum, and minimum values were calculated to describe the central tendency and variability of each data before and after scaling. This analysis provided a foundational understanding of the data's distribution, range, and variance.

B. Feature Distribution Analysis:

Distribution of Clinical Features Across Training and Validation Sets.

To ensure data consistency and support model generalization, the distribution of selected clinical features was analyzed across training and validation sets using histograms and boxplots. Figure 5.1 presents histograms and box plots for each clinical feature, comparing their distributions across the training and validation datasets. These visualizations help identify distributional differences and ensure the consistency of data splits.

Histogram: Presented in Figure 5.1, the histograms (top two rows) show how each clinical feature is distributed in the training set (top row) and validation set (middle row). The X-axis shows the range of values for each feature, and the Y-axis (Count) shows the number of patients who fall into each value range (bin). Each bar represents a group of patients with similar values (each vertical bar represents a bin), and the height of the bar shows how common that range is. A tall bar means many patients had values in that range; a short bar means fewer patients did. The smooth blue line shows the overall shape of the distribution. Features like `heart_rate` and `weight` have similar distribution shapes across both sets, confirming the data split is balanced.

Boxplot: The boxplots (bottom row in Figure 5.1) summarize each clinical feature by showing the median (horizontal center line), spread (box = middle 50% of values), whiskers (range of typical

values), and outliers (dots). The X-axis shows the dataset group (Train vs Validation), and the Y-axis represents the actual values of each feature. Features like `weight` and `heart_rate` show notable outliers, but maintain similar medians and distributions across the sets. The whiskers represent the range of typical values, while a tiny or no box indicates that most patients had the same value. The middle 50% of the data refers to the values that are not too low and not too high, but sit in the central range of the data distribution. Overall, the distributions are consistent, confirming the data split is fair and supports reliable model training.

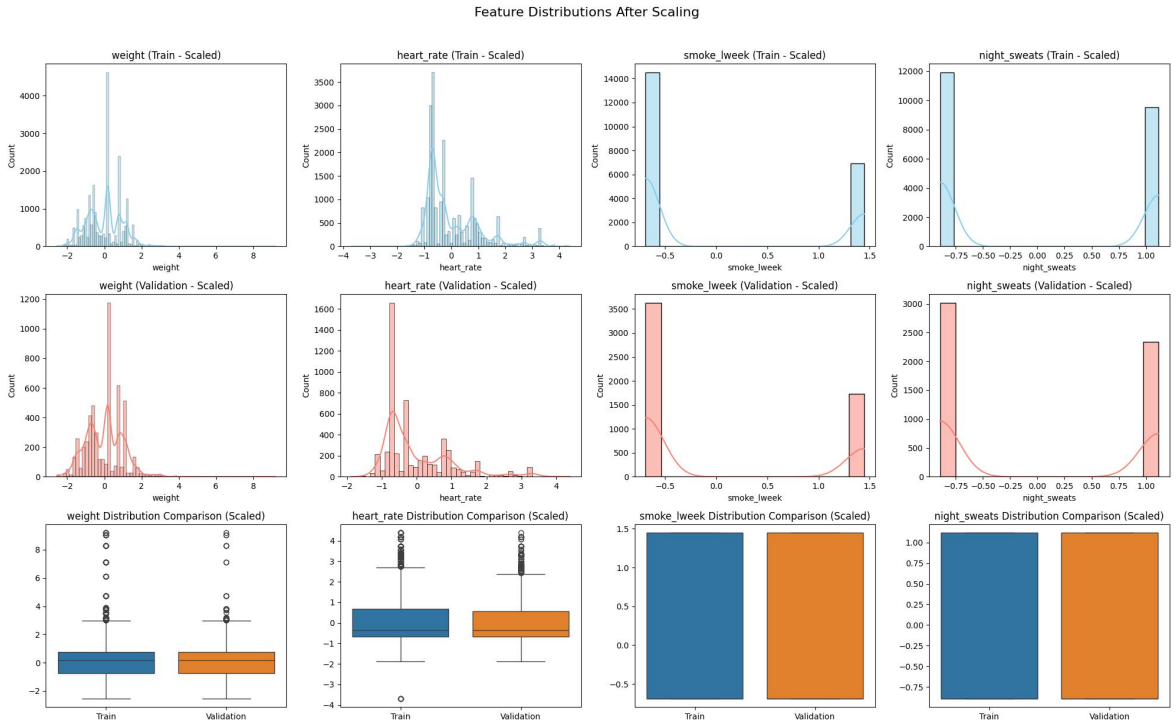


Figure 5.1: Histogram and box plots comparing the distribution of clinical features across training and validation datasets.

C. Feature Correlation Matrix

The correlation matrix shows how clinical features are related to each other, and a correlation heatmap was generated (see Figure 5.2). This matrix visualizes Pearson correlation coefficients (r), which measure the linear association between feature pairs. The intensity of red (strong positive correlation) and blue (strong negative correlation) shading in the matrix reflects the strength and direction of each correlation. Coefficients range from -1 (perfect negative correlation) to $+1$ (perfect positive correlation), with values near 0 indicating weak or no linear relationship.

Strong positive correlations were observed between `tb_prior` and `tb_prior_Pul`, as well as between `smoke_1week` and `weight_loss`, indicating that they provide highly similar information. Symptom-related features, including `fever`, `night_sweats`, and `temperature`, were moderately correlated, suggesting these features often co-occur in TB cases and may strengthen the model when combined. Most other feature pairs showed weak or no correlation, which is beneficial for model training, as it reduces redundancy and helps ensure that each feature contributes unique information.

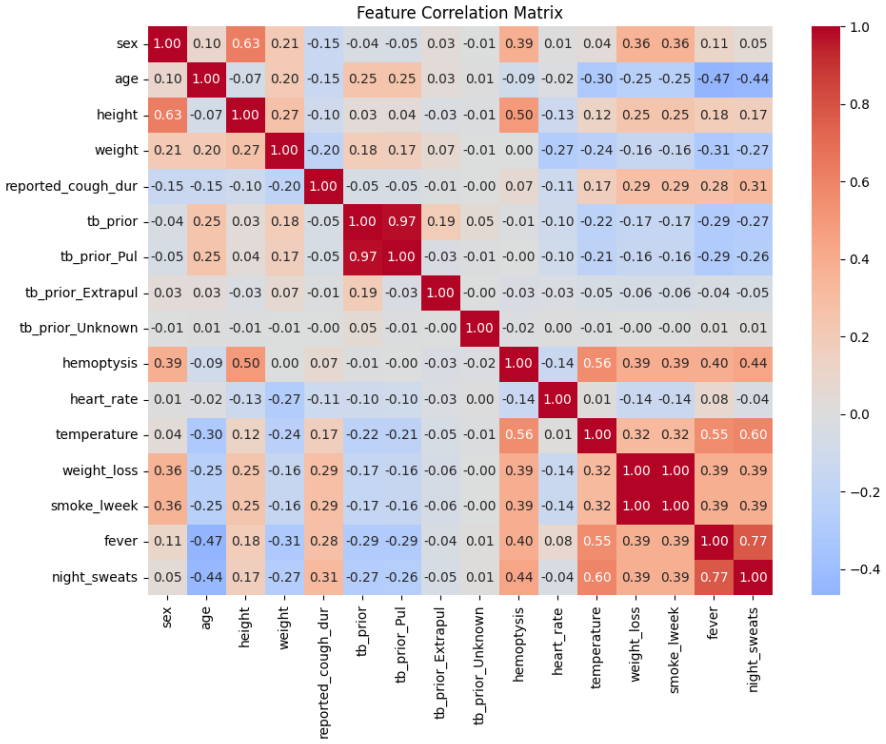


Figure 5.2: Correlation matrix of clinical features.

5.2 Feature Importance Analysis using SHAP

To understand how each clinical feature influenced the model's predictions, SHAP (SHapley Additive exPlanations) was applied as part of the Explainable AI, a popular method that explains how machine learning models make decisions. Since the model uses both audio and clinical data, a custom setup was created to focus only on the clinical data. The audio input was kept fixed, and the clinical features were changed to see how each one affected the model's output. SHAP's `KernelExplainer`, which works for any type of model, was used to calculate how important each feature was.

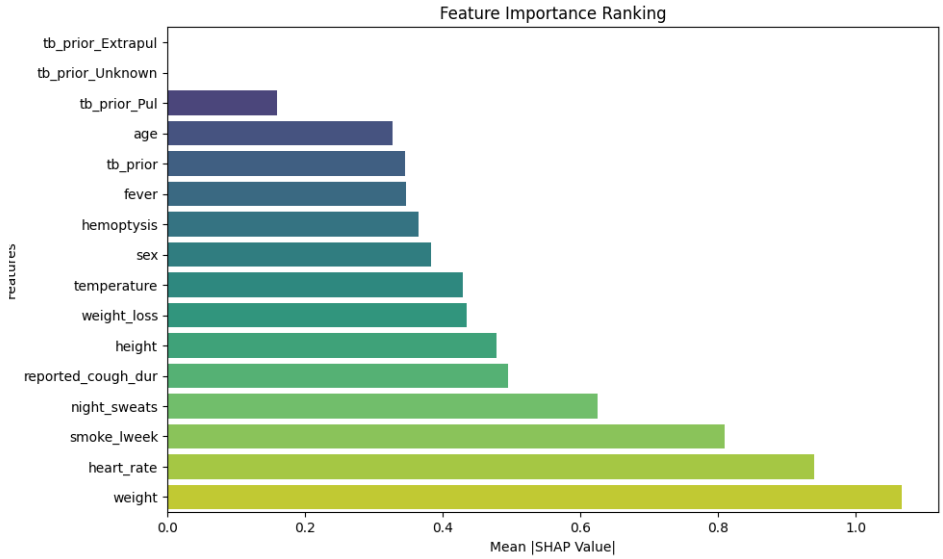


Figure 5.3: Bar chart showing SHAP-based feature importance ranking.

The results were shown as a bar chart (Figure 5.3) that ranked the features based on how much they influenced the model's predictions. The most important features were `weight` (1.0670), `heart_rate`

(0.9402), `smoke_lweek` (0.8092), and `night_sweats` (0.6257). Other features with moderate impact included `reported_cough_dur` (0.4945), `height` (0.4789), `weight_loss` (0.4350), and `temperature` (0.4290). Features like `sex` (0.3831), `hemoptysis` (0.3654), `fever` (0.3469), `tb_prior` (0.3452), and `age` (0.3266) also contributed, but to a lesser degree. In contrast, `tb_prior_Pul` (0.1589), `tb_prior_Unknown` (0.0000), and `tb_prior_Extrapul` (0.0000) had very little or no effect.

Why this matters? This analysis shows that the model is not just making accurate predictions, but also using the right medical information to do so. In healthcare, it's important that doctors can trust and understand AI decisions. SHAP makes this possible by clearly showing which symptoms the model relied on. This helps build confidence, supports faster decision-making, and encourages collaboration between healthcare professionals and AI tools. Note that due to the multi-input structure of the model, SHAP analysis was limited to the structured clinical features. The raw audio signal was held constant during the computation to isolate the contribution of clinical information. Although the audio input plays a vital role in classification, quantifying its influence via SHAP remains challenging with current tools.

5.3 Model Training Results

5.3.1 Base and Enhanced Model Results

A two-stage deep learning architecture for TB detection was employed. The training outcomes for both models are summarized below:

- **Base Model (`CNNbasedNet121`)**
  Trained solely on raw cough audio, achieved a training accuracy of 0.9494 and an AUC of 0.9896. These results indicate that cough audio alone can offer valuable predictive features for TB classification, though it may not capture the complete clinical picture.

- **Enhanced Model (`CNNbasedNet`)**
  This model integrated both audio features and structured clinical metadata. The addition of clinical information resulted in improved performance, achieving a training accuracy of 0.9865 and an AUC of 0.9993. This reflects the strength of multimodal learning in capturing more comprehensive TB-related patterns.

Table 5.1: Training performance metrics of the base and enhanced models.

Metric	**Base Model (Audio Only)**	**Enhanced Model (Audio + Clinical)**
Accuracy	0.9494	**0.9865**
AUC (Area Under Curve)	0.9896	**0.9993**

As Table 5.1 shows, combining cough sound features with clinical metadata improves model performance. The improvement in AUC indicates enhanced discriminative power, while the increase in accuracy reflects better classification confidence.

5.3.2 Learning Progress of the Model

Figure 5.4 illustrates the model's learning progress across 20 training epochs. The left plot shows the accuracy trends: training accuracy improves steadily and consistently, while validation accuracy fluctuates slightly but shows an overall upward trend, indicating learning progress. The right plot displays the loss curves: both training and validation loss decrease, with training loss showing a smoother descent. Although validation loss exhibits some variability—common in deep learning—it does not indicate severe overfitting. Together, these plots confirm that the model is effectively learning from the data and generalizing reasonably well.

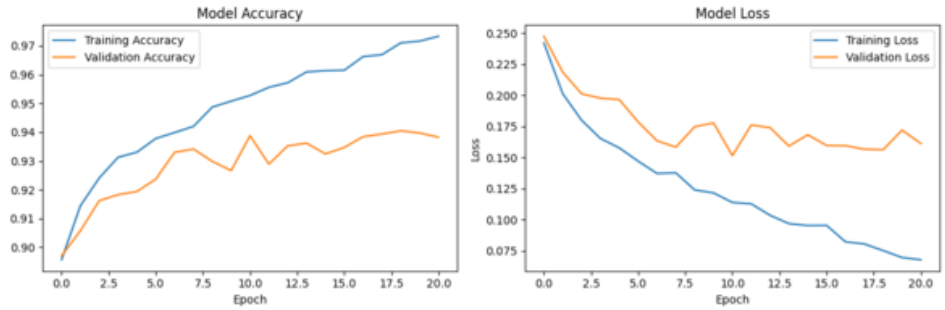


Figure 5.4: Training and validation accuracy and loss over 20 epochs.

5.4 Testing of the developed CNN based model

5.4.1 Confusion Matrix

The confusion matrix is a standard evaluation tool used to visualize the performance of a classification model. It presents the distribution of true and false predictions.

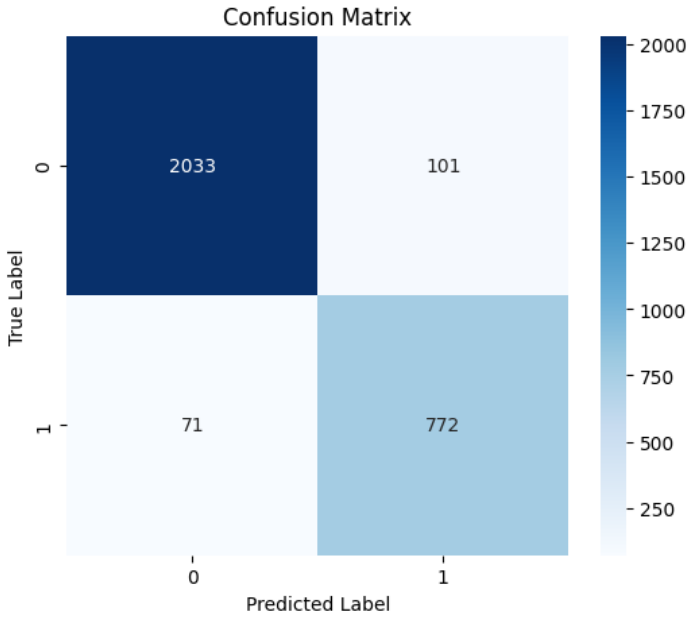


Figure 5.5: Confusion matrix representing classification outcomes on the test dataset.

Figure 5.5 illustrates the model's classification performance on the test dataset. The top-left cell (True Negatives = 2033) represents non-TB cases correctly identified as negative. The bottom-right

cell (True Positives = 772) shows TB-positive cases accurately predicted. The off-diagonal entries indicate classification errors: 101 false positives (non-TB cases incorrectly classified as TB) and 71 false negatives (missed TB cases). The low number of false negatives (compared to the number of false positives) is particularly important in a medical context, suggesting the model is effective at identifying actual TB cases, which is critical for early diagnosis and treatment.

5.4.2 Receiver Operating Characteristic (ROC) Analysis

Figure 5.6 illustrates the ROC curve, which plots the True Positive Rate against the False Positive Rate across a range of classification thresholds. The proposed model achieved an AUC score of 0.9859, indicating excellent discriminative performance. The ROC curve's shape, bending sharply toward the top-left corner, further confirms the model's strong sensitivity and specificity across thresholds. This level of performance suggests the model is highly reliable for TB screening applications.

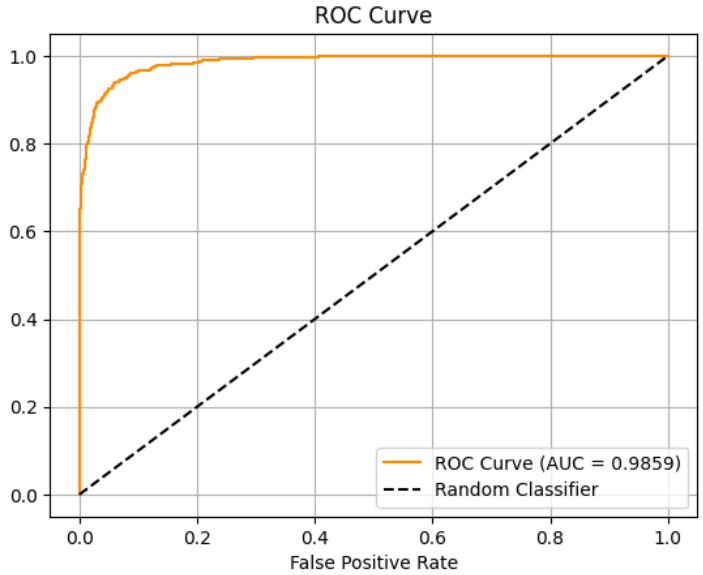


Figure 5.6: Receiver Operating Characteristic (ROC) curve with AUC score.

5.4.3 Precision-Recall (PR) curve

Figure 5.7 illustrates the Precision-Recall (PR) curve. The PR curve highlights the trade-off between precision and recall across different decision thresholds.

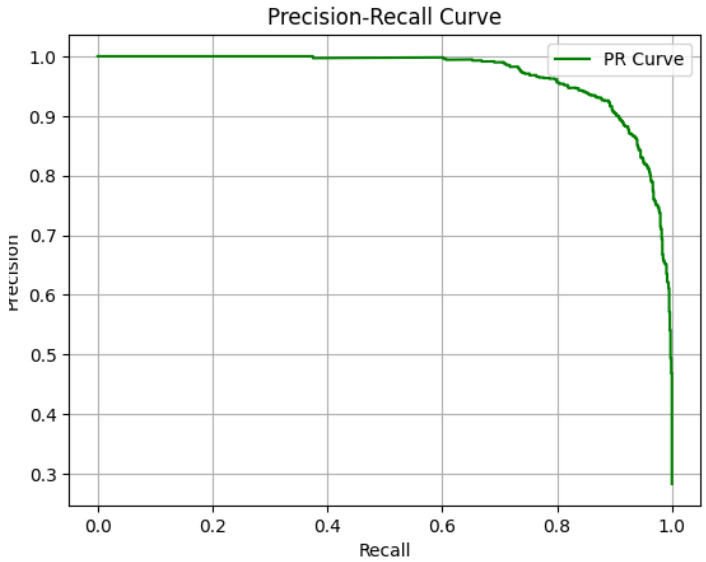


Figure 5.7: Precision-Recall (PR) curve illustrating the trade-off between precision and recall.

The curve remains near the top-right corner for a large portion of the threshold range, indicating that the model maintains both high precision and high recall across various decision boundaries. This is especially important in TB detection, where minimizing false negatives is critical—without significantly increasing false positives. The high performance shown in the PR curve is also reflected in the model's F1 score of 0.8998, which balances both precision and recall, confirming the model's robustness in correctly identifying TB-positive cases while minimizing misclassification.

5.4.4 Calibration Curve

Figure 5.8 shows the calibration curve. The calibration curve determines how close the model's predicted risk of TB is to the actual number of positive cases. The x-axis represents the mean predicted probability, while the y-axis indicates the actual fraction of positive cases. The dashed diagonal line represents perfect calibration, where predicted probabilities match actual outcomes exactly. The model's calibration curve (in purple) closely follows these actual outcomes, showing it is well-calibrated and its confidence scores are reliable.

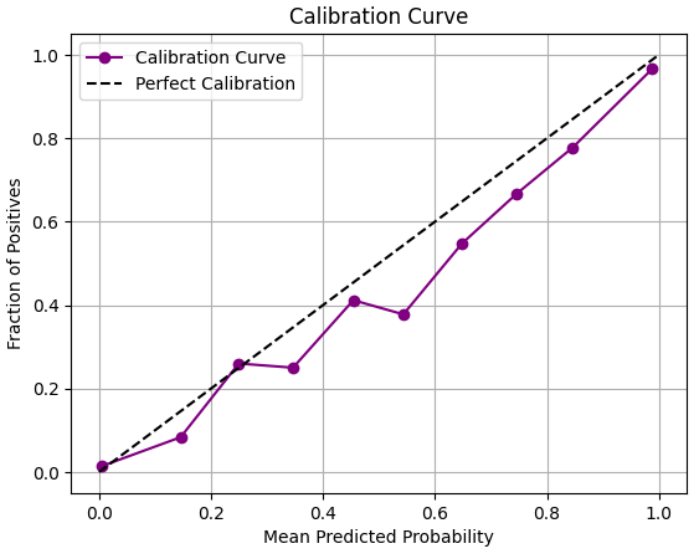


Figure 5.8: Calibration curve comparing predicted probabilities to actual outcomes.

5.4.5 Summary of Core Performance Metrics.

Table 5.2: Core performance metrics on the test dataset.

Metric	Value
Accuracy	0.9422
AUC (Area Under Curve)	0.9859
F1 Score	0.8998
Precision	0.8843
Recall (Sensitivity)	0.9158

Table 5.2 presents the key evaluation metrics of the proposed model on the test dataset. The model achieved an accuracy of 94.22%, which measures the overall correctness of the model by calculating the proportion of correctly predicted samples out of the total number of samples. An AUC of 0.9859 indicates excellent discriminative capability between TB-positive and TB-negative cases. The F1 score of 0.8998 demonstrates a strong balance between precision and recall—critical in minimizing both false positives and false negatives in a medical screening context. The precision of 0.8843 reflects the proportion of true TB cases among those predicted as positive, while the recall of 0.9158 confirms the model's ability to correctly identify the majority of actual TB-positive cases.

5.5 Comparative Evaluation of CNNbasedNet and Baseline Models

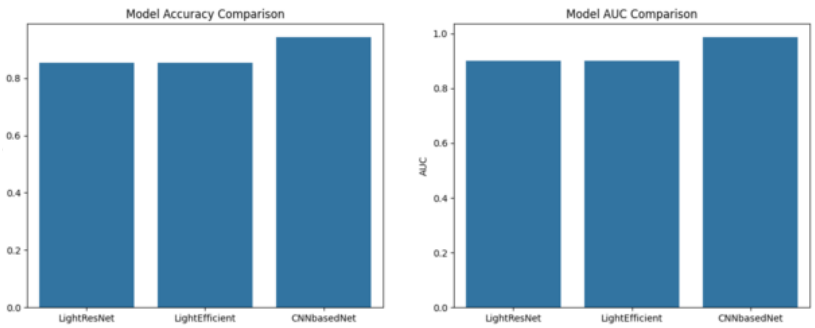


Figure 5.9: Comparison of model accuracy, AUC

To evaluate the effectiveness of the proposed CNNbasedNet model, a comparative analysis was conducted against two lightweight baseline architectures: LightResNet and LightEfficient. Evaluation metrics included accuracy, precision, recall, F1 score, AUC, and inference time. Figure 5.9 shows that CNNbasedNet outperforms both baselines in accuracy and AUC. It achieves an accuracy of 94.22% and an AUC of 0.9859, compared to 85.29% and 0.9005 for LightResNet, and 85.22% and 0.9002 for LightEfficient.

5.5.1 Core Performance Metrics.

Figure 5.10 confirms that CNNbasedNet achieves the best overall performance across all key metrics. It shows higher precision (0.8843), recall (0.9158), and F1 score (0.8998), making it the most balanced model in both detection and confidence.

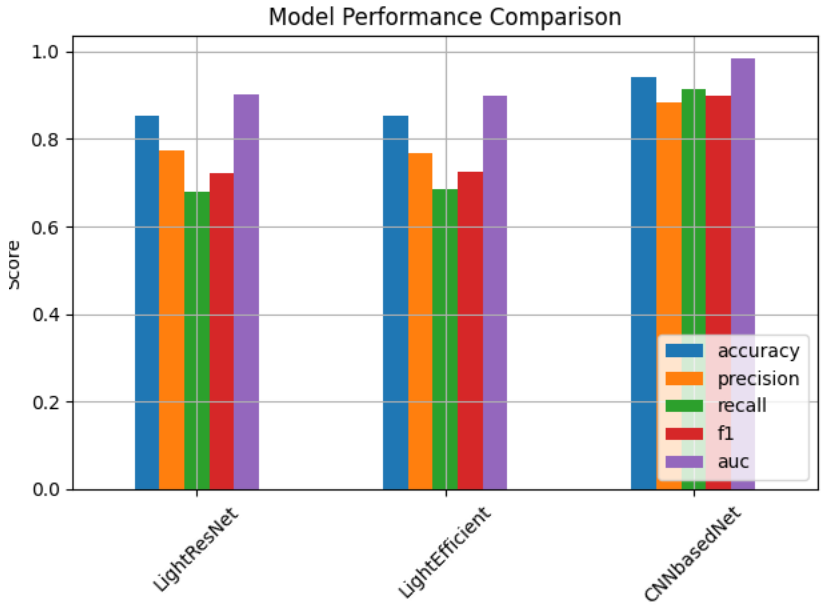


Figure 5.10: Bar chart comparing core performance metrics (accuracy, precision, recall, F1, AUC).

Table 5.3: Benchmark performance comparison of all models.

Model	Accuracy	Precision	Recall	F1 Score	AUC	Inference Time (s)
LightResNet	0.8530	0.7736	0.6798	0.7237	0.9005	0.00165
LightEfficient	0.8522	0.7673	0.6864	0.7246	0.9002	0.00088
CNNbasedNet	**0.9422**	**0.8843**	**0.9158**	**0.8998**	**0.9859**	**0.00153**

As shown in Table 5.3, CNNbasedNet outperforms the other models in all metrics, achieving the highest accuracy, precision, recall, F1 score, and AUC.

5.5.2 Inference Time

Inference time is the time required by a trained model to process an output from a given input, and it is a measure of computational efficiency, particularly for real-time or resource-constrained applications

[83].

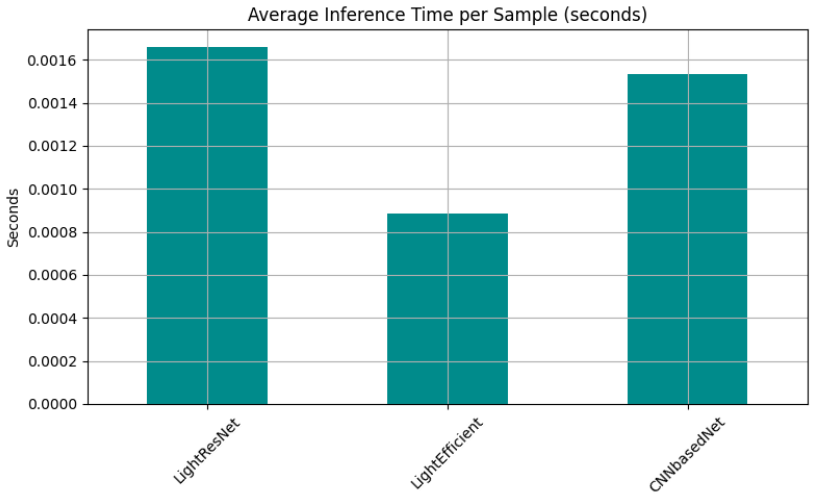


Figure 5.11: Average inference time per sample (in seconds) across all models.

Figure 5.11 shows, all models maintain low inference latency, with LightEfficient being the fastest (0.88ms), followed closely by CNNbasedNet (1.53ms) and LightResNet (1.65 ms) respectively. This confirms that performance gains do not significantly compromise speed.

5.5.3 Comparative Evaluation with MetforNet (CODA TB DREAM Challenge)

MetforNet, the top-performing model from the CODA TB DREAM challenge, was trained on a subset of the publicly accessible CODA dataset, consisting of 9,772 cough recordings from 1,105 participants. In contrast, the developed CNNbasedNet model was trained on a larger portion of the same publicly accessible dataset. While MetforNet integrates a complex architecture, CNNbasedNet model develops a distinct, simplified CNN-based architecture reducing the number of CNN blocks and removing recurrent and attention layers to lower complexity and training time. Thus, while the model comparison provides valuable insights into architectural efficiency and scalability, differences in training dataset size should be taken into account when interpreting the results. A comparison of key performance metrics between CNNbasedNet and MetforNet is provided in Table 5.4.

Table 5.4: Performance comparison between CNNbasedNet and MetforNet

Model	**Accuracy**	**F1 Score**	**AUC**
MetforNet	0.8488	0.7178	0.9092
CNNbasedNet	**0.9422**	**0.8998**	**0.9859**

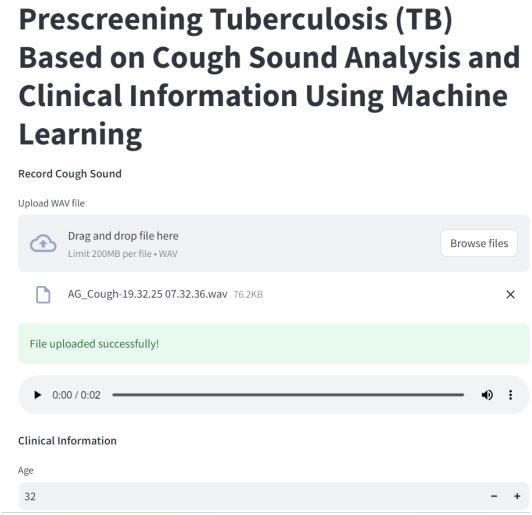
Overall, CNNbasedNet demonstrates superior predictive accuracy, reliability, and generalization compared to the baseline models, without compromising inference efficiency. These results support its suitability for real-time TB screening in practical, resource-constrained healthcare settings. The comparative evaluation results highlight the clear advantages of CNNbasedNet over the baseline models with significantly higher accuracy, precision, recall, F1 score, and AUC. Its balanced performance indicates that it effectively minimizes both false positives and false negatives, which is critical in

clinical diagnostics.

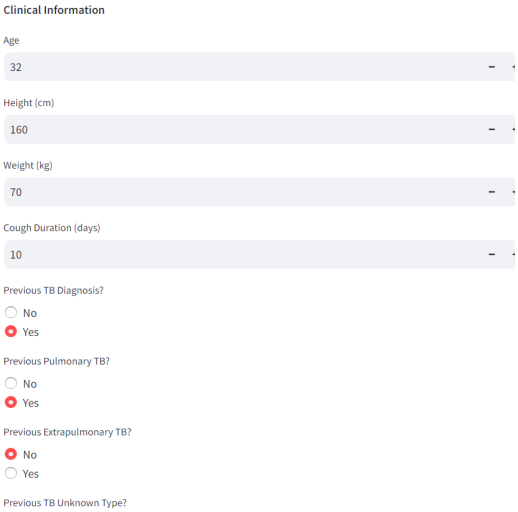
Although LightResNet and LightEfficient provide competitive speed and lightweight architectures, their lower recall and F1 scores suggest limitations in reliably identifying TB-positive cases. In contrast, CNNbasedNet combines strong discriminative power, achieving higher performance in identifying positive instances case.

Importantly, CNNbasedNet achieves these gains without compromising inference efficiency, maintaining low latency comparable to the lighter models. This makes it well-suited for deployment in real-time or resource-constrained environments, where accuracy and response time are both crucial. Overall, the results support CNNbasedNet as a robust and practical solution for automated TB screening from cough sounds and clinical metadata.

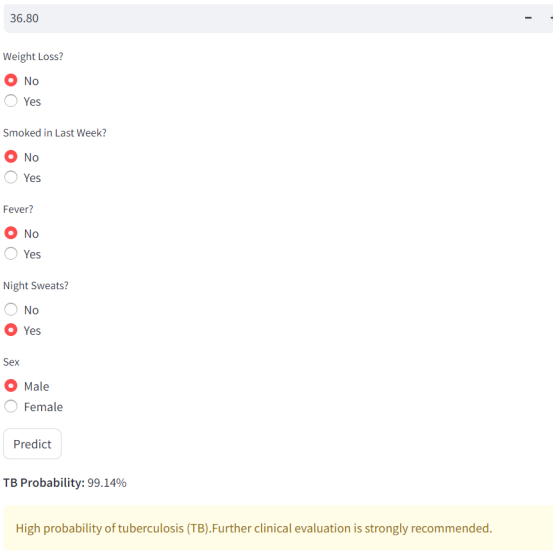
5.6 GUI



(a) Upload audio file



(b) Filling clinical data



(c) Prediction result

Figure 5.12: Web-based prediction interface used during deployment. Upload a cough audio and fill out clinical data to receive model outputs with probabilities and predictions.

The interface allows users to upload a cough audio recording, enter clinical information, and receive a prediction result in real time. A snapshot of the deployed prediction interface is shown in Figure 5.12, where users upload cough recordings and receive classification results through a web-based GUI. The key components of the interface are summarized below:

- **Library Imports:** Essential libraries such as `streamlit`, `numpy`, and `soundfile` are imported.
- **Model Loading:** The pre-trained CNN model is loaded for use in the app.
- **Preprocessing Functions:** Custom functions are defined to preprocess both audio and clinical

input data before feeding them into the model.

- **Main Function:** A `main()` function is used to control the flow of the Streamlit application.
- **GUI Elements:** Streamlit widgets are used to create interactive components. Users can upload a cough recording, fill out clinical data fields, and click a ”Predict” button.
- **Prediction Logic:** Upon clicking ”Predict”, the app preprocesses the inputs, passes them to the model, and computes a prediction probability.
- **Result Interpretation:** The model output is interpreted as either TB-positive or TB-negative based on the probability score. A threshold of 0.5 is used as the decision boundary. Scores closer to 1 indicate TB-positive, while those closer to 0 indicate TB-negative.
- **App Execution:** The application is executed by calling the `main()` function.

Prediction Output Interpretation The model outputs a probability score between 0 and 1. This score is interpreted based on a threshold of 0.5, as summarized in Table 5.5.

Table 5.5: Probability Score Interpretation

TB Negative	Cut-off (Threshold)	TB Positive
Probability score $\rightarrow 0$	0.5	Probability score $\rightarrow 1$

5.7 Deployment Results on Ethiopian Dataset

The model was tested during deployment on a small real-world sample of 12 patients. The confusion matrix results are as follows:

- True Positives (TP) = 7
- True Negatives (TN) = 3
- False Positives (FP) = 0
- False Negatives (FN) = 2

Based on these results, the post-deployment model evaluation results demonstrated :

- **Accuracy:** 83.33%
- **Sensitivity (Recall):** 77.78%

These metrics indicate promising post-deployment performance. However, the given sample size is small (12 patients), making the findings highly sensitive to individual outcomes and not statistically conclusive. Also, this limits the reliability of the metrics. Observed differences are expected and may be influenced by sampling variability. These results should be viewed as an early observational insight rather than a definitive evaluation. To properly validate the model’s performance and generalizability, future testing on a larger and more diverse cohort is strongly recommended.

Chapter 6

Conclusion, Limitations and Future Works

6.1 Conclusion

This study presented a deep learning approach for tuberculosis (TB) pre-screening using cough sound and structured clinical information. The proposed solution consists of a two-stage architecture: a base model (CNNbasedNet121) for audio feature extraction and an enhanced Model (CNNbasedNet) that integrates clinical information with the learned audio features. The final model was evaluated against two lightweight baselines — LightResNet and LightEfficient.

The results show that CNN model (CNNbasedNet) outperforms the baseline models in all major performance metrics, including accuracy, precision, recall, F1 score, and AUC. Furthermore, CNNbasedNet outperformed MetforNet121 model in terms of accuracy, F1 score, and AUC. Specifically, the model achieved an accuracy of 94.22%, a recall of 91.58%, and an AUC of 0.9859, indicating a high confidence in identifying TB-positive cases while maintaining strong overall reliability. Importantly, these gains were accomplished without a significant increase in inference time, confirming the model's suitability for real-time deployment. Visualizations such as the ROC curve, precision-recall curve, calibration plot, and confusion matrix further validated the model's robustness, interpretability, and clinical relevance. SHAP-based feature importance analysis further enhanced interpretability by identifying the specific clinical features that the model relied on during prediction. This transparency helps build clinician trust, enables faster decision-making, and enhances the model's practical value.

According to the WHO 2024 Target Product Profile (TPP) for community-based TB triage tools, an effective model should achieve $\geq 90\%$ sensitivity and $\geq 70\%$ specificity. CNNbasedNet met and exceeded these thresholds, achieving 91.58% sensitivity and 95.27% specificity, demonstrating its potential as a reliable triage tool for early TB detection in low-resource settings.

Overall, CNNbasedNet offers a promising solution for scalable, non-invasive, low-cost, rapid, and easily accessible TB screening, particularly in low-resource settings where timely diagnosis can have a substantial public health impact. Its integration with a simple web interface and real-time prediction capability supports broader accessibility and community-based use. While the model shows strong performance, its current limitations include lack of adequate local data for further training, variability in audio quality, and the absence of a built-in cough recording for the web-app interface. Future work will focus on addressing these limitations.

6.2 Contribution

This study presents a CNN-based deep learning approach for TB prescreening that meets the WHO Target Product Profile (TPP) criteria for community-based triage tests. It incorporates explainability using SHAP to show which clinical features the model generally relied on in making predictions, enhancing transparency and building trust among healthcare professionals. The system also enables

real-time prediction through a web-based application and includes post-deployment evaluation using a small set of Ethiopian data.

6.3 Limitations

While the proposed model demonstrates strong performance in TB detection, several limitations should be acknowledged:

- **Limited Representation of Local Data:** Although the model was trained on a diverse multi-national dataset from seven countries, it did not include any data from Ethiopia which is the intended deployment context. While a small number of Ethiopian samples were collected and used for post-deployment evaluation, their limited size constrains the ability to fully assess or adapt the model to local population characteristics, potentially impacting its generalizability and reliability in this setting.

- **Audio Quality Variability:** The model relies heavily on the quality of cough recordings. Background noise, recording devices, and user compliance during recording could introduce variability that affects performance in real-world settings.

- **Local Dataset Sampling Rate:** The local dataset was collected at a 16 kHz sampling rate, which may limit capturing higher-frequency cough sounds and potentially affect model performance.

- **Lack of Integrated Cough Recording Interface:** The current deployment (via Streamlit GUI) does not include a built-in cough audio recorder. Instead, it relies on users uploading pre-recorded audio files, which may reduce usability and limit real-time data collection or diagnosis.

- **Limited Post-Deployment Validation Data:** User data collected after deployment was limited in size and diversity. This constrains the ability to robustly evaluate the model's real-world performance and generalizability.

6.4 Future Work

Future work could address these limitations by incorporating representative Ethiopian data during training to improve local relevance and reduce population bias. Expanding the post-deployment dataset in size and diversity is also necessary to enable more reliable real-world evaluation. Integrating a built-in cough recording interface would enhance usability and support real-time data collection, as well as testing on various hardware configurations in field settings. Additionally, future work will aim to extend the model to differentiate TB from other coexisting respiratory diseases.

References

[1] World Health Organization, "Global tuberculosis report 2024," World Health Organization, Geneva, Switzerland, Tech. Rep., 2024, accessed April 18, 2025. [Online]. Available: https://iris.who.int/bitstream/handle/10665/379339/9789240101531-eng.pdf

[2] ——, "1.1 tb incidence," Available at: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023/tb-disease-burden/1-1-tb-incidence, 2023, accessed: Dec. 02, 2023.

[3] ——, "World Life Expectancy Report – Tuberculosis in Ethiopia," https://www.worldlifeexpectancy.com/ethiopia-tuberculosis, 2020, accessed April 18, 2025.

[4] ——, "Global Tuberculosis Report 2022," https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2022, 2022, accessed: Aug. 24, 2023.

[5] A. J. Zimmer, C. Ugarte-Gil, R. Pathri, P. Dewan, D. Jaganath, A. Cattamanchi, M. Pai, and S. Grandjean Lapierre, "Making cough count in tuberculosis care," *Communications medicine*, vol. 2, no. 1, p. 83, 2022.

[6] G. H. R. Botha, "Lung health diagnosis through cough sound analysis," Ph.D. dissertation, Stellenbosch: Stellenbosch University, 2017.

[7] A. F. Jember, Y. M. Ayano, and T. G. Debelee, "Robust cough analysis system for diagnosis of tuberculosis using artificial neural network," in *Pan African Conference on Artificial Intelligence*. Springer, 2022, pp. 3–26.

[8] J. Jonathan and A. A. Barakabitze, "Ml technologies for diagnosing and treatment of tuberculosis: a survey," *Health and Technology*, vol. 13, no. 1, pp. 17–33, 2023.

[9] M. Pahar, M. Klopper, B. Reeve, R. Warren, G. Theron, and T. Niesler, "Automatic cough classification for tuberculosis screening in a real-world environment," *Physiological Measurement*, vol. 42, no. 10, p. 105014, 2021.

[10] K. S. Alqudaihi, N. Aslam, I. U. Khan, A. M. Almuhaideb, S. J. Alsunaidi, N. M. A. R. Ibrahim, F. A. Alhaidari, F. S. Shaikh, Y. M. Alsenbel, D. M. Alalharith *et al.*, "Cough sound detection and diagnosis using artificial intelligence techniques: challenges and opportunities," *Ieee Access*, vol. 9, pp. 102 327–102 344, 2021.

[11] M. Cohen-McFarlane, R. Goubran, and F. Knoefel, "Comparison of silence removal methods for the identification of audio cough events," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 1263–1268.

[12] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," in *2017 15th international conference on ICT and knowledge engineering (ICT&KE)*. IEEE, 2017, pp. 1–6.

[13] S. Huddart, V. Yadav, S. K. Sieberts, L. Omberg, M. Raberahona, R. Rakotoarivelo, I. N. Lyimo, O. Lweno, D. J. Christopher, N. V. Nhung *et al.*, "Solicited cough sound analysis for tuberculosis triage testing: the coda tb dream challenge dataset," *MedRxiv*, 2024.

[14] World Health Organization, "World Health Organization (WHO)," https://www.who.int/, 2023, accessed: Oct. 31, 2023.

[15] Centers for Disease Control and Prevention, "Clinical and laboratory diagnosis — tb — cdc," 2023, accessed: 2025-04-23. [Online]. Available: https://www.cdc.gov/tb/hcp/testing-diagnosis/clinical-and-laboratory-diagnosis.html

[16] ——, "Clinical and laboratory diagnosis — tb — cdc," 2023, accessed: 2025-04-23. [Online]. Available: https://www.cdc.gov/tb/hcp/testing-diagnosis/clinical-and-laboratory-diagnosis.html

[17] S. H. Lee, "Tuberculosis infection and latent tuberculosis," *Tuberculosis and respiratory diseases*, vol. 79, no. 4, p. 201, 2016.

[18] S. Kiazyk and T. Ball, "Latent tuberculosis infection: An overview," *Canada Communicable Disease Report*, vol. 43, no. 3-4, p. 62, 2017.

[19] E. Nuermberger, W. R. Bishai, and J. H. Grosset, "Latent tuberculosis infection," in *Seminars in Respiratory and critical care Medicine*, vol. 25. Copyright© 2004 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New . . . , 2004, pp. 317–336.

[20] H. Mohajan, "Tuberculosis is a fatal disease among some developing countries of the world," *Tuberculosis and respiratory diseases*, vol. 0, no. 0, p. 18, 2014.

[21] World Health Organization, "Tuberculosis," 2023, accessed: 2025-04-23. [Online]. Available: https://www.who.int/health-topics/tuberculosis#tab=tab_2

[22] C. Wejse, P. Gustafson, J. Nielsen, V. F. Gomes, P. Aaby, P. L. Andersen, and M. Sodemann, "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*, vol. 40, no. 2, pp. 111–120, 2008.

[23] World Health Organization, "Culture and drug susceptibility testing," 2022, accessed: 2025-04-23. [Online]. Available: https://tbksp.who.int/en/node/734

[24] R. Vongthilath-Moeung, A. Poncet, G. Renzi, J. Schrenzel, and J.-P. Janssens, "Time to detection of growth for mycobacterium tuberculosis in a low incidence area," *Frontiers in cellular and infection microbiology*, vol. 11, p. 704169, 2021.

[25] P. J. Brennan and H. Nikaido, "The envelope of mycobacteria," *Annual review of biochemistry*, vol. 64, no. 1, pp. 29–63, 1995.

[26] A. Gupta, A. Kaul, A. G. Tsolaki, U. Kishore, and S. Bhakta, "Mycobacterium tuberculosis: Immune evasion, latency and reactivation," *Immunobiology*, vol. 217, no. 3, pp. 363–374, 2012.

[27] Centers for Disease Control and Prevention, "How tb spreads," 2023, accessed: 2025-04-23. [Online]. Available: https://www.cdc.gov/tb/causes/index.html

[28] World Health Organization, "Tuberculosis: Transmission," 2023, accessed: 2025-04-23. [Online]. Available: https://www.who.int/health-topics/tuberculosis#tab=tab_2

[29] R. Long and K. Schwartzman, "Pathogenesis and transmission of tuberculosis," *Canadian tuberculosis standards*, p. 25, 2014.

[30] M. U. Shiloh, "Mechanisms of mycobacterial transmission: how does mycobacterium tuberculosis enter and escape from the human host," pp. 1503–1506, 2016.

[31] M. Gengenbacher and S. H. Kaufmann, "Mycobacterium tuberculosis: success through dormancy," *FEMS microbiology reviews*, vol. 36, no. 3, pp. 514–532, 2012.

[32] Centers for Disease Control and Prevention, "Clinical and laboratory diagnosis — tb — cdc," 2023, accessed: 2025-04-23. [Online]. Available: https://www.cdc.gov/tb/hcp/testing-diagnosis/clinical-and-laboratory-diagnosis.html

[33] D. J. Beste, M. Espasa, B. Bonde, A. M. Kierzek, G. R. Stewart, and J. McFadden, "The genetic requirements for fast and slow growth in mycobacteria," *PloS one*, vol. 4, no. 4, p. e5349, 2009.

[34] a. R. S. R.Reyna, F. M. Smithuis, "Imaging findings in tb," 2014, accessed: 2025-04-24. [Online]. Available: https://radiologyassistant.nl/chest/tb/tuberculosis

[35] E. Skoura, A. Zumla, and J. Bomanji, "Imaging in iuberculosis," *International journal of infectious diseases*, vol. 32, pp. 87–93, 2015.

[36] Centers for Disease Control and Prevention, "Mycobacterium tuberculosis bacteria (colorized)," 2002, accessed: 2025-04-23. [Online]. Available: https://phil.cdc.gov/Details.aspx?pid=4427

[37] M. P. La Manna, B. Tamburini, V. Orlando, G. D. Badami, P. Di Carlo, A. Cascio, M. Singh, F. Dieli, and N. Caccamo, "Liodetect® tb-st: Evaluation of novel blood test for a rapid diagnosis of active pulmonary and extra-pulmonary tuberculosis in igra confirmed patients," *Tuberculosis*, vol. 130, p. 102119, 2021.

[38] G. Gualano, P. Mencarini, F. N. Lauria, F. Palmieri, S. Mfinanga, P. Mwaba, J. Chakaya, A. Zumla, and G. Ippolito, "Tuberculin skin test–outdated or still useful for latent tb infection screening?" *International Journal of Infectious Diseases*, vol. 80, pp. S20–S22, 2019.

[39] M. D. Rakhmawatie, T. Wibawa, P. Lisdiyanti, W. R. Pratiwi *et al.*, "Evaluation of crystal violet decolorization assay and resazurin microplate assay for antimycobacterial screening," *Heliyon*, vol. 5, no. 8, 2019.

[40] J. S. Zifodya, J. S. Kreniske, I. Schiller, M. Kohli, N. Dendukuri, S. G. Schumacher, E. A. Ochodo, F. Haraka, A. A. Zwerling, M. Pai *et al.*, "Xpert ultra versus xpert mtb/rif for pulmonary tuberculosis and rifampicin resistance in adults with presumptive pulmonary tuberculosis," *Cochrane Database of Systematic Reviews*, no. 2, 2021.

[41] M. Pai, "Tuberculosis: The story after the primer," *Nature Reviews Disease Primers*, vol. 6, no. 1, p. 29, 2020.

[42] R. Singh, S. P. Dwivedi, U. S. Gaharwar, R. Meena, P. Rajamani, and T. Prasad, "Recent updates on drug resistance in mycobacterium tuberculosis," *Journal of applied microbiology*, vol. 128, no. 6, pp. 1547–1567, 2020.

[43] N. Vodnala, P. R. Lankireddy, and P. Yarlagadda, "Characterization of cough sounds using statistical analysis," *arXiv preprint arXiv:2308.03019*, 2023.

[44] S. Hegde, S. Sreeram, I. L. Alter, C. Shor, T. A. Valdez, K. D. Meister, and A. Rameau, "Cough sounds in screening and diagnostics: a scoping review," *The Laryngoscope*, vol. 134, no. 3, pp. 1023–1031, 2024.

[45] S. W. Smith, F. Point, F. Point, S. Linearity, S. Fidelity, C. Decompositions, P. Notation, P. Nuisances, F. Basics, B.-P. High-Pass *et al.*, "The scientist and engineer's guide to digital signal processing by steven w. smith, ph. d." *Proceedings of the IRE*, 1997.

[46] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental concepts of convolutional neural network," *Recent trends and advances in artificial intelligence and Internet of Things*, pp. 519–567, 2020.

[47] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S.-Y. Chang, and T. Sainath, "Deep learning for audio signal processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 2, pp. 206–219, 2019.

[48] K. K. Lella and A. Pja, "Automatic covid-19 disease diagnosis using 1d convolutional neural network and augmentation with human respiratory sound based on parameters: Cough, breath, and voice," *AIMS public health*, vol. 8, no. 2, p. 240, 2021.

[49] S. Mastromichalakis, "Alrelu: A different approach on leaky relu activation function to improve neural networks performance," *arXiv preprint arXiv:2012.07564*, 2020.

[50] G. Chen, P. Chen, Y. Shi, C.-Y. Hsieh, B. Liao, and S. Zhang, "Rethinking the usage of batch normalization and dropout in the training of deep neural networks," *arXiv preprint arXiv:1905.05928*, 2019.

[51] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions," *Journal of big Data*, vol. 8, pp. 1–74, 2021.

[52] M. Vakalopoulou, S. Christodoulidis, N. Burgos, O. Colliot, and V. Lepetit, "Deep learning: Basics and convolutional neural networks (cnns)," *Machine learning for brain disorders*, pp. 77–115, 2023.

[53] J. Ayeni, "Convolutional neural network (cnn): the architecture and applications," *Appl. J. Phys. Sci*, vol. 4, pp. 42–50, 2022.

[54] C. Suda, "Early detection of tuberculosis with machine learning cough audio analysis: Towards more accessible global triaging usage," *arXiv preprint arXiv:2310.17675*, 2023.

[55] G. P. Kafentzis, S. Tetsing, J. Brew, L. Jover, M. Galvosas, C. Chaccour, and P. M. Small, "Predicting tuberculosis from real-world cough audio recordings and metadata," *arXiv preprint arXiv:2307.04842*, 2023.

[56] G. D. Yellapu, G. Rudraraju, N. R. Sripada, B. Mamidgi, C. Jalukuru, P. Firmal, V. Yechuri, S. Varanasi, V. S. Peddireddi, D. M. Bhimarasetty *et al.*, "Development and clinical validation of swaasa ai platform for screening and prioritization of pulmonary tb," *Scientific Reports*, vol. 13, no. 1, p. 4740, 2023.

[57] J. Yadav, A. S. Varde, H. Liu, G. Antoniou, and L. Xie, "Audiovisual multimodal cough data analysis for tuberculosis detection," in *2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2024, pp. 1–8.

[58] G. Frost, G. Theron, and T. Niesler, "Tb or not tb? acoustic cough analysis for tuberculosis alassification," *arXiv preprint arXiv:2209.00934*, 2022.

[59] M. Pahar, G. Theron, and T. Niesler, "Automatic tuberculosis detection in cough patterns using nlp-style cough embeddings," in *2022 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE, 2022, pp. 1–6.

[60] W. Xu, X. Bao, X. Lou, X. Liu, Y. Chen, X. Zhao, C. Zhang, C. Pan, W. Liu, and F. Liu, "Feature fusion method for pulmonary tuberculosis patient detection based on cough sound," *Plos one*, vol. 19, no. 5, p. e0302651, 2024.

[61] H. Mahmood, M. Iftikhar, A. Wali, A. Ali, and M. Gulzar, "A novel cascaded approach for classification of tuberculosis using cough audio in real-time environment," *IEEE Access*, 2024.

[62] X. Bao, W. Xu, C. Pan, W. Liu, X. Lou, Y. Chen, X. Zhao, C. Zhang, and F. Liu, "Tuberculosis detection based on cough sounds: a multi-model voting mechanism," in *2023 16th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2023, pp. 1–6.

[63] M. S. Knight, "The role of features in predictive deep learning models for auditory tuberculosis classification," Ph.D. dissertation, Stellenbosch University, 2024.

[64] G. Botha, G. Theron, R. Warren, M. Klopper, K. Dheda, P. Van Helden, and T. Niesler, "Detection of tuberculosis by automatic cough sound analysis," *Physiological measurement*, vol. 39, no. 4, p. 045005, 2018.

[65] S. J. S. Rajasekar, A. R. Balaraman, D. V. Balaraman, S. Mohamed Ali, K. Narasimhan, N. Krishnasamy, and V. Perumal, "Detection of tuberculosis using cough audio analysis: a deep learning approach with capsule networks," *Discover Artificial Intelligence*, vol. 4, no. 1, p. 77, 2024.

[66] W. Xu, H. Yuan, X. Lou, Y. Chen, and F. Liu, "Dmrnet based tuberculosis screening with cough sound," *IEEE Access*, vol. 12, pp. 3960–3968, 2023.

[67] H. Yuan, X. Lou, Y. Chen, Z. Meng, J. Bai, F. Liu, and W. Xu, "Tuberculosis screening with cough sounds using the deep learning models," in *2023 4th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*. IEEE, 2023, pp. 444–448.

[68] S. Huddart, V. Yadav, S. K. Sieberts, L. Omberg, M. Raberahona, R. Rakotoarivelo, I. N. Lyimo, O. Lweno, D. J. Christopher, N. V. Nhung *et al.*, "A dataset of solicited cough sound for tuberculosis triage testing," *Scientific Data*, vol. 11, no. 1, p. 1149, 2024.

[69] J. C. Ferrao, M. D. Oliveira, F. Janela, and H. M. Martins, "Preprocessing structured clinical data for predictive modeling and decision support," *Applied clinical informatics*, vol. 7, no. 04, pp. 1135–1153, 2016.

[70] H. S. Obaid, S. A. Dheyab, and S. S. Sabry, "The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning," in *2019 9th annual information technology, electromechanical engineering and microelectronics conference (iemecon)*. IEEE, 2019, pp. 279–283.

[71] H. D. Salaudeen, S. Anuyah, and A. M. Folawewo, "Advancing tuberculosis prediction: Integrating ai, cnn, and matlab for enhanced predictive modelling," *International Journal of Computer Applications Technology and Research*, 2024.

[72] S. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462*, 2015.

[73] H. Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of min-max normalization and z-score normalization in the k-nearest neighbor (knn) algorithm to test the accuracy of types of breast cancer," *International Journal of Informatics and Information Systems*, vol. 4, no. 1, pp. 13–20, 2021.

[74] J. Sadaiyandi, P. Arumugam, A. K. Sangaiah, and C. Zhang, "Stratified sampling-based deep learning approach to increase prediction accuracy of unbalanced dataset," *Electronics*, vol. 12, no. 21, p. 4423, 2023.

[75] D. Berrar *et al.*, "Performance measures for binary classification." 2019.

[76] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Computer science on-line conference*. Springer, 2023, pp. 15–25.

[77] E. Hussain, M. Hasan, S. Z. Hassan, T. H. Azmi, M. A. Rahman, and M. Z. Parvez, "Deep learning based binary classification for alzheimer's disease detection using brain mri images,"

in *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2020, pp. 1115–1120.

[78] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[79] C. Halimu, A. Kasem, and S. S. Newaz, "Empirical comparison of area under roc curve (auc) and mathew correlation coefficient (mcc) for evaluating machine learning algorithms on imbalanced datasets for binary classification," in *Proceedings of the 3rd international conference on machine learning and soft computing*, 2019, pp. 1–6.

[80] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.

[81] E. Beauxis-Aussalet and L. Hardman, "Visualization of confusion matrix for non-expert users," in *IEEE Conference on Visual Analytics Science and Technology (VAST)-Poster Proceedings*. sn, 2014, pp. 1–2.

[82] H.-A. Goh, C.-K. Ho, and F. S. Abas, "Front-end deep learning web apps development and deployment: a review," *Applied intelligence*, vol. 53, no. 12, pp. 15 923–15 945, 2023.

[83] V. J. Hellendoorn, C. Bird, E. T. Barr, and M. Allamanis, "Deep learning type inference," in *Proceedings of the 2018 26th acm joint meeting on european software engineering conference and symposium on the foundations of software engineering*. ACM, 2018, pp. 152–162.

Appendix A

Additional Tables

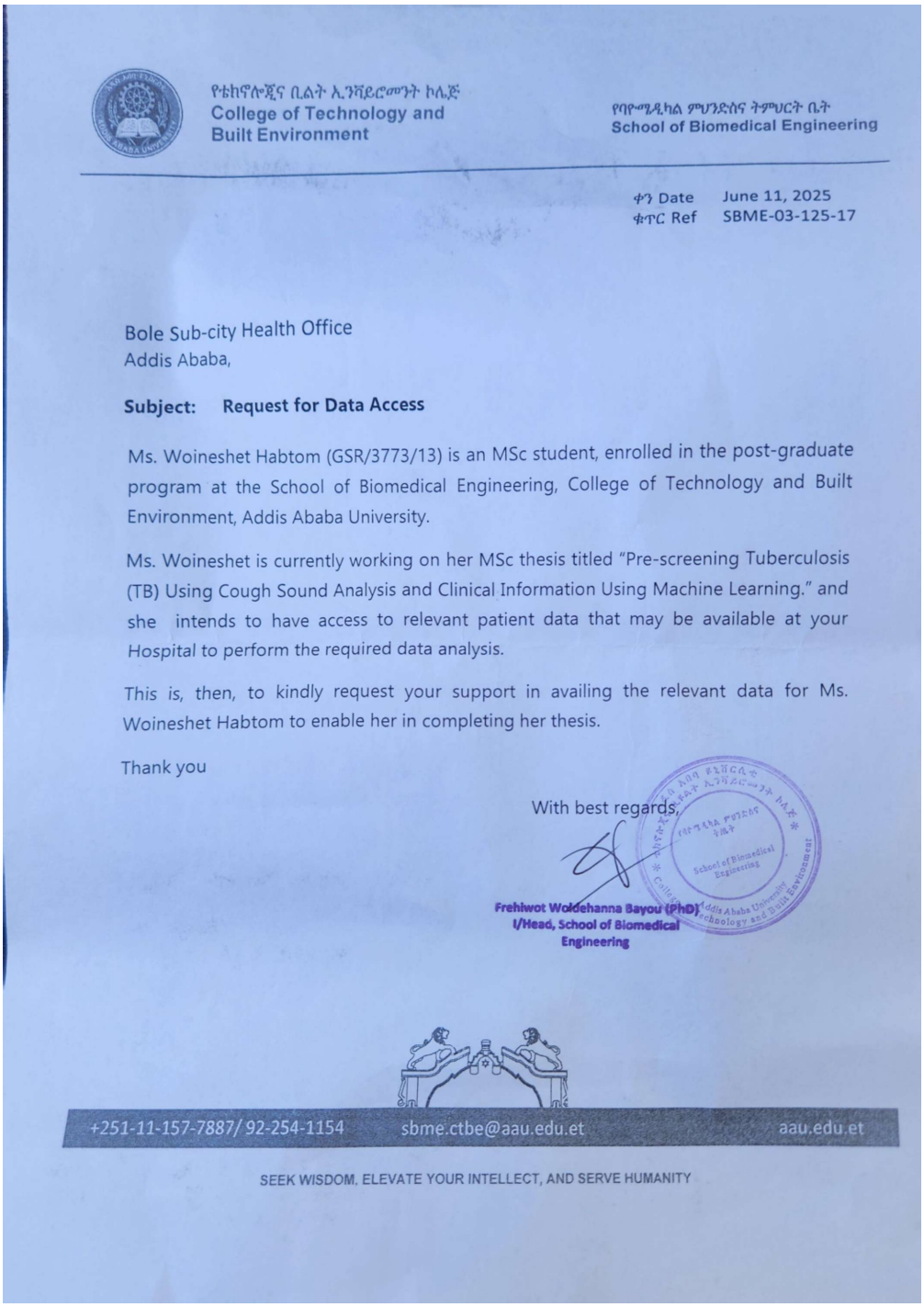
Table A.1: Available Demographic, Clinical, and Microbiologic Variables [13]

Variable	**Values**	**Format/Definition**
Country	Philippines, Vietnam, South Africa, Uganda, India, Madagascar, Tanzania	Country where the participant was enrolled
Sex	Male, Female	Sex at birth reported by participant
Age	Numeric	Age calculated as date of collection minus date of birth (if known). If unknown, reported age at time of collection.
Height	Numeric	Height in centimeters
Weight	Numeric	Weight in kilograms
Reported duration of cough	Numeric	Self-reported duration in days.
Prior TB	Yes, No	Self-reported history of TB diagnosis
Prior TB type: Pulmonary	Checked, Unchecked	Selected “Pulmonary TB” at baseline
Prior TB type: Extrapulmonary	Checked, Unchecked	Selected “Extrapulmonary TB” at baseline
Prior TB type: Unknown	Checked, Unchecked	Selected “Unknown” at baseline
Hemoptysis	Yes, No	Self-reported coughing up blood in past 30 days
Heart rate	Numeric	Beats per minute at baseline
Temperature	Numeric	Temperature in Celsius at baseline
Weight loss	Yes, No	Self-reported weight loss in past 30 days
Smoked in last week	Yes, No	Use of tobacco or vaping products in past 7 days
Fever	Yes, No	Self-reported fever in past 30 days
Night sweats	Yes, No	Self-reported night sweats in past 30 days
Microbiologic reference standard*	TB Positive, TB Negative, Indeterminate	TB diagnosis based on sputum and culture results
Sputum Xpert reference standard*	Positive, Negative, Indeterminate	TB diagnosis based on sputum results alone
Xpert combined semi-quant	Trace, Very Low, Low, Medium, High	Semiquantitative result from Xpert Ultra (if test is positive)

Appendix B

Supplementary Materials

B.1 Support Letter



የቴክኖሎጂና ቢልት ኢንቫይሮመንት ኮሌጅ
College of Technology and Built Environment

የባዮሜዲካል ምህንድስና ትምህርት ቤት
School of Biomedical Engineering

ቀን Date June 11, 2025
ቁጥር Ref SBME-03-125-17

Bole Sub-city Health Office
Addis Ababa,

Subject: Request for Data Access

Ms. Woineshet Habtom (GSR/3773/13) is an MSc student, enrolled in the post-graduate program at the School of Biomedical Engineering, College of Technology and Built Environment, Addis Ababa University.

Ms. Woineshet is currently working on her MSc thesis titled "Pre-screening Tuberculosis (TB) Using Cough Sound Analysis and Clinical Information Using Machine Learning." and she intends to have access to relevant patient data that may be available at your Hospital to perform the required data analysis.

This is, then, to kindly request your support in availing the relevant data for Ms. Woineshet Habtom to enable her in completing her thesis.

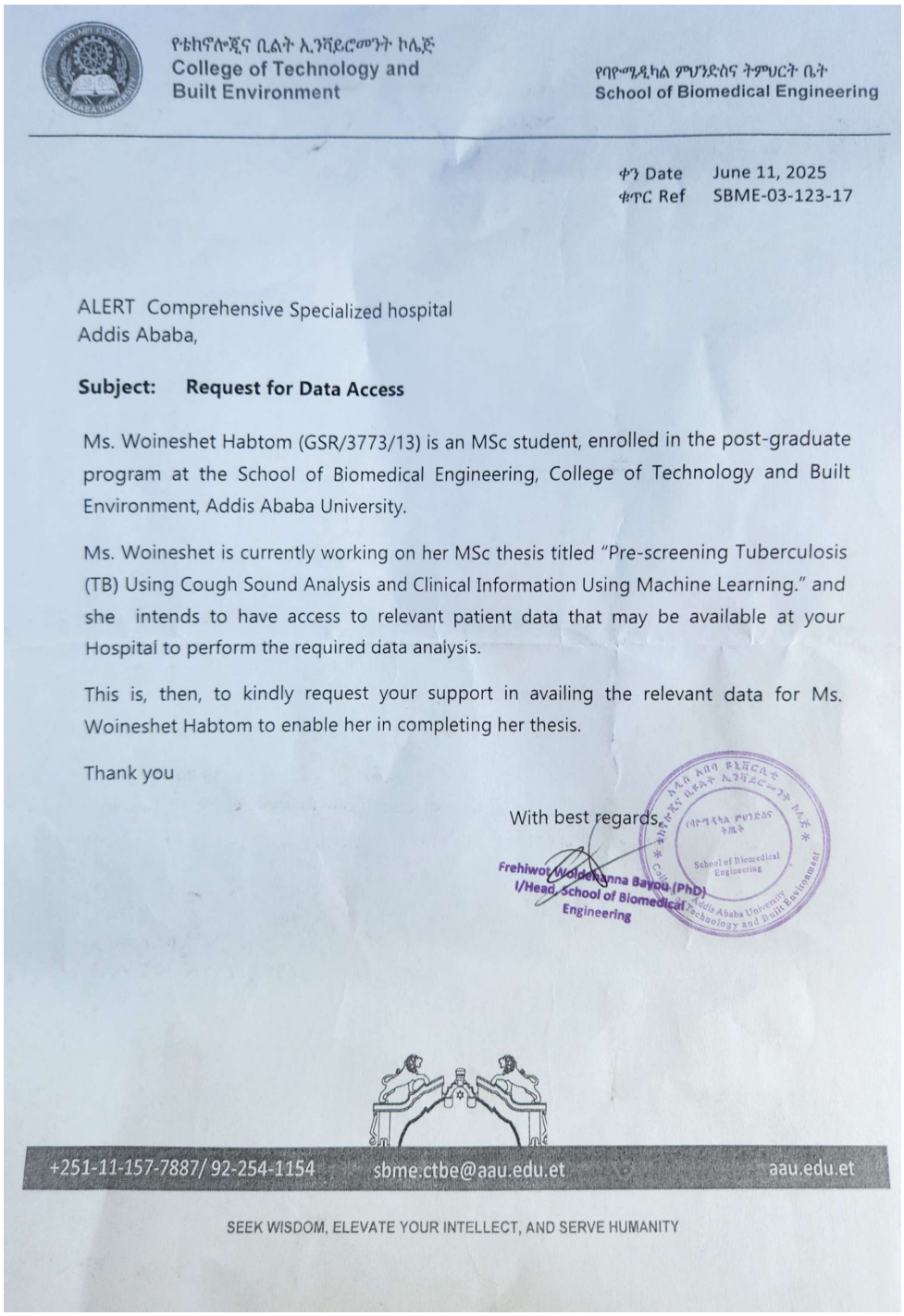
Thank you

With best regards,

Frehiwot Woldehanna Bayou (PhD)
I/Head, School of Biomedical Engineering

+251-11-157-7887/ 92-254-1154 sbme.ctbe@aau.edu.et aau.edu.et

SEEK WISDOM, ELEVATE YOUR INTELLECT, AND SERVE HUMANITY



የቴክኖሎጂና ቢልት ኢንቫይሮመንት ኮሌጅ
College of Technology and Built Environment

የባዮሜዲካል ምህንድስና ትምህርት ቤት
School of Biomedical Engineering

ቀን Date June 11, 2025
ቁጥር Ref SBME-03-123-17

ALERT Comprehensive Specialized hospital
Addis Ababa,

Subject: Request for Data Access

Ms. Woineshet Habtom (GSR/3773/13) is an MSc student, enrolled in the post-graduate program at the School of Biomedical Engineering, College of Technology and Built Environment, Addis Ababa University.

Ms. Woineshet is currently working on her MSc thesis titled "Pre-screening Tuberculosis (TB) Using Cough Sound Analysis and Clinical Information Using Machine Learning." and she intends to have access to relevant patient data that may be available at your Hospital to perform the required data analysis.

This is, then, to kindly request your support in availing the relevant data for Ms. Woineshet Habtom to enable her in completing her thesis.

Thank you

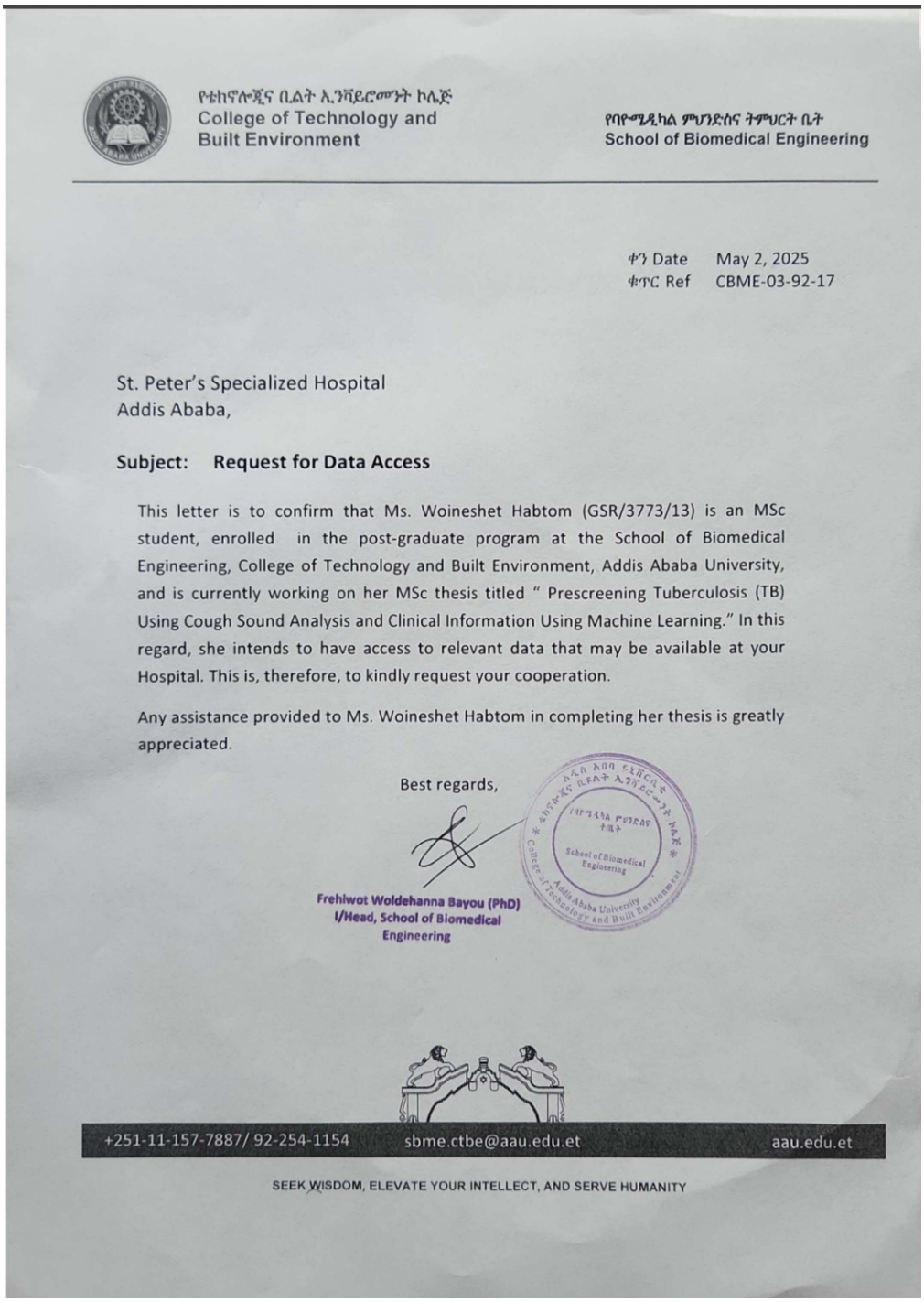
With best regards,

Frehiwot Woldekanna Bayou (PhD)
I/Head, School of Biomedical Engineering

School of Biomedical Engineering
Addis Ababa University College of Technology and Built Environment

+251-11-157-7887/ 92-254-1154 sbme.ctbe@aau.edu.et aau.edu.et

SEEK WISDOM, ELEVATE YOUR INTELLECT, AND SERVE HUMANITY



የቴክኖሎጂና ቢልት ኢንቫይሮመንት ኮሌጅ
College of Technology and Built Environment

የባዮሜዲካል ምህንድስና ትምህርት ቤት
School of Biomedical Engineering

ቀን Date May 2, 2025
ቁጥር Ref CBME-03-92-17

St. Peter's Specialized Hospital
Addis Ababa,

Subject: Request for Data Access

This letter is to confirm that Ms. Woineshet Habtom (GSR/3773/13) is an MSc student, enrolled in the post-graduate program at the School of Biomedical Engineering, College of Technology and Built Environment, Addis Ababa University, and is currently working on her MSc thesis titled " Prescreening Tuberculosis (TB) Using Cough Sound Analysis and Clinical Information Using Machine Learning." In this regard, she intends to have access to relevant data that may be available at your Hospital. This is, therefore, to kindly request your cooperation.

Any assistance provided to Ms. Woineshet Habtom in completing her thesis is greatly appreciated.

Best regards,

Frehiwot Woldehanna Bayou (PhD)
I/Head, School of Biomedical Engineering

+251-11-157-7887/ 92-254-1154 sbme.ctbe@aau.edu.et aau.edu.et

SEEK WISDOM, ELEVATE YOUR INTELLECT, AND SERVE HUMANITY

B.2 Clinical Data Collection form

TB Cough Data Collection Protocol

1. Data Collection Protocol

This protocol outlines how to collect cough sound recordings and clinical metadata from patients for TB prescreening using a mobile app. Please follow these instructions to ensure consistency and qualit across all recordings.

Who to Record

- ✓ Adults (18+) with a new or worsening cough lasting ≥ 2 weeks
- ✓ Provided informed consent

What You Need

- ✓ Android phone with the 'Cough Detector and Recorder' app installed
- ✓ Quiet room with minimal background noise
- ✓ Tripod or stable surface for phone

How to Record

- ✓ Open the app and prepare to record
- ✓ Ask the patient to stand 60–90 cm from the phone
- ✓ Record 5 clear, voluntary coughs
- ✓ Use a 3-2-1 countdown before each cough

Audio File

- ✓ Save the audio file as: PatientID.wav
- ✓ Only keep recordings with clearly audible coughs

Clinical Information to Record

- ✓ Patient ID
- ✓ Age, Sex, Height, Weight
- ✓ Cough Duration (in days)
- ✓ Prior TB (general, pulmonary, extrapulmonary, unknown)
- ✓ Hemoptysis (blood in cough)
- ✓ Heart rate, Temperature
- ✓ Weight loss, Fever, Night sweats
- ✓ Smoking status (last 7 days)

Submit Data

- ➢ Ensure the sameness of patient ID for recorded audio file, clinical metadata and diagnosis status confirmed by the hospital for specific patient
- ➢ Do not include names or direct identifiers
- ➢ Email the audio file and the matching clinical form to: grapehabtom2@gmail.com

Clinical Metadata and Cough Audio Collection Form

1. Patient specifc clinical metadata and cough audio record

Patient ID: _____ Date: _____

Audio Filename (WAV): _____

Feature	Value	Circle or Enter
Sex	Male / Female	☐ Male ☐ Female
Age (in years)	Numeric	_____
Height (cm)	Numeric	_____
Weight (kg)	Numeric	_____
Reported cough duration (days)	Numeric	_____
Prior TB	Yes / No	☐ Yes ☐ No
Prior Pulmonary TB	Yes / No	☐ Yes ☐ No
Prior Extrapulmonary TB	Yes / No	☐ Yes ☐ No
Prior Unknown TB	Yes / No	☐ Yes ☐ No
Hemoptysis (blood in cough)	Yes / No	☐ Yes ☐ No
Heart rate (bpm)	Numeric	_____
Temperature (°C)	Numeric	_____
Weight loss (last 30 days)	Yes / No	☐ Yes ☐ No
Smoked in last 7 days	Yes / No	☐ Yes ☐ No
Fever (last 30 days)	Yes / No	☐ Yes ☐ No
Night sweats (last 30 days)	Yes / No	☐ Yes ☐ No

2. Diagnosis status of the patient by the Hospital

☐ TB Positive ☐ TB Negative ☐ Suspected / Under Investigation ☐ Other: _____

Notes(optional):_____

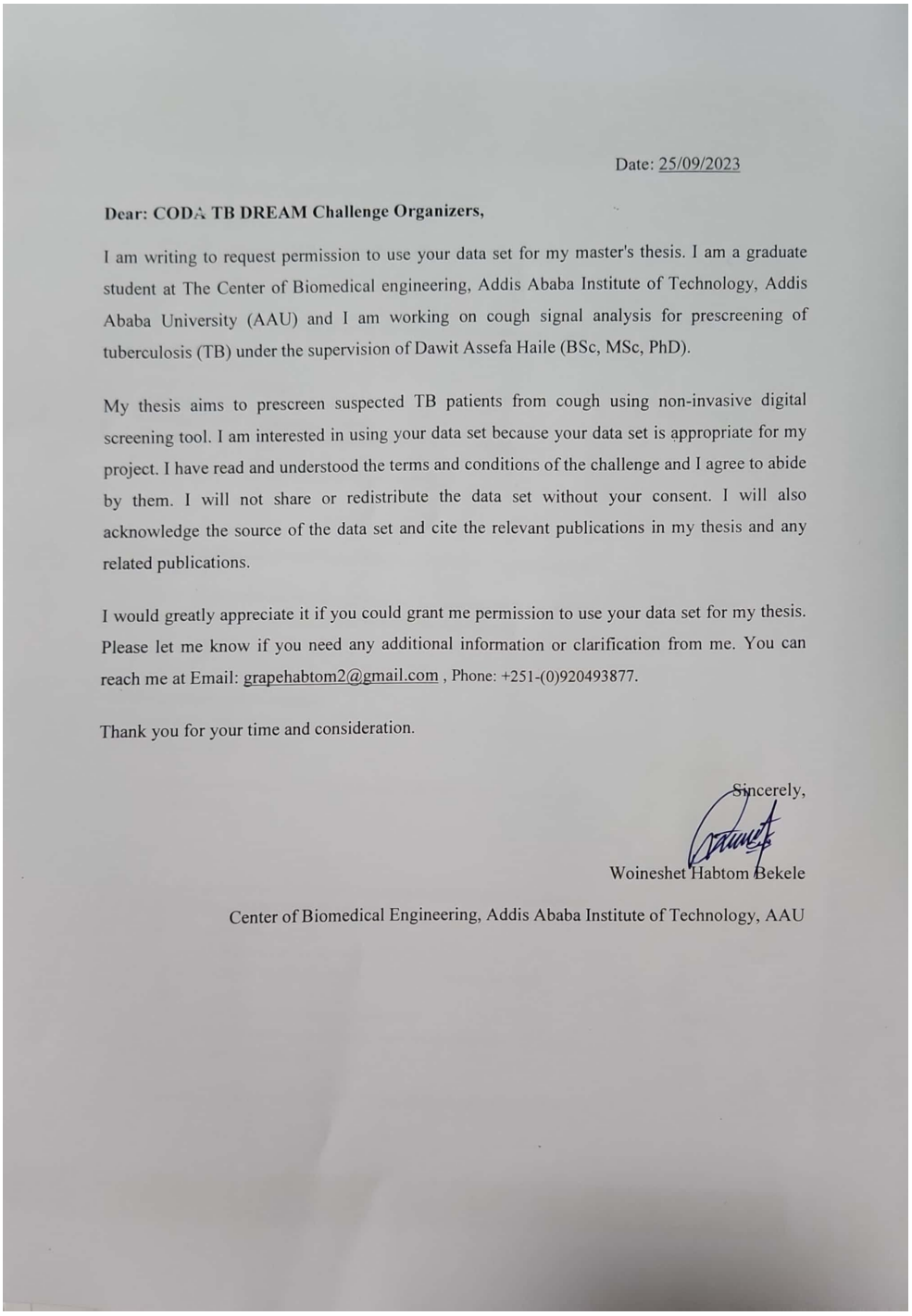
_____

Staff Name: _____

Appendix C

Coda TB challenge dataset access

Validation documents



Date: 25/09/2023

Dear: CODA TB DREAM Challenge Organizers,

I am writing to request permission to use your data set for my master's thesis. I am a graduate student at The Center of Biomedical engineering, Addis Ababa Institute of Technology, Addis Ababa University (AAU) and I am working on cough signal analysis for prescreening of tuberculosis (TB) under the supervision of Dawit Assefa Haile (BSc, MSc, PhD).

My thesis aims to prescreen suspected TB patients from cough using non-invasive digital screening tool. I am interested in using your data set because your data set is appropriate for my project. I have read and understood the terms and conditions of the challenge and I agree to abide by them. I will not share or redistribute the data set without your consent. I will also acknowledge the source of the data set and cite the relevant publications in my thesis and any related publications.

I would greatly appreciate it if you could grant me permission to use your data set for my thesis. Please let me know if you need any additional information or clarification from me. You can reach me at Email: grapehabtom2@gmail.com , Phone: +251-(0)920493877.

Thank you for your time and consideration.

Sincerely,

Woineshet Habtom Bekele

Center of Biomedical Engineering, Addis Ababa Institute of Technology, AAU

Addis Ababa University
አዲስ አበባ ዩኒቨርሲቲ
SEEK WISDOM, ELEVATE YOUR INTELLECT AND SERVE HUMANITY!

Addis Ababa Institute of Technology
Addis Ababa, Ethiopia
25/09/2023

To Whom It May Concern

Subject: Identity Attestation for Woineshet Habtom Bekele

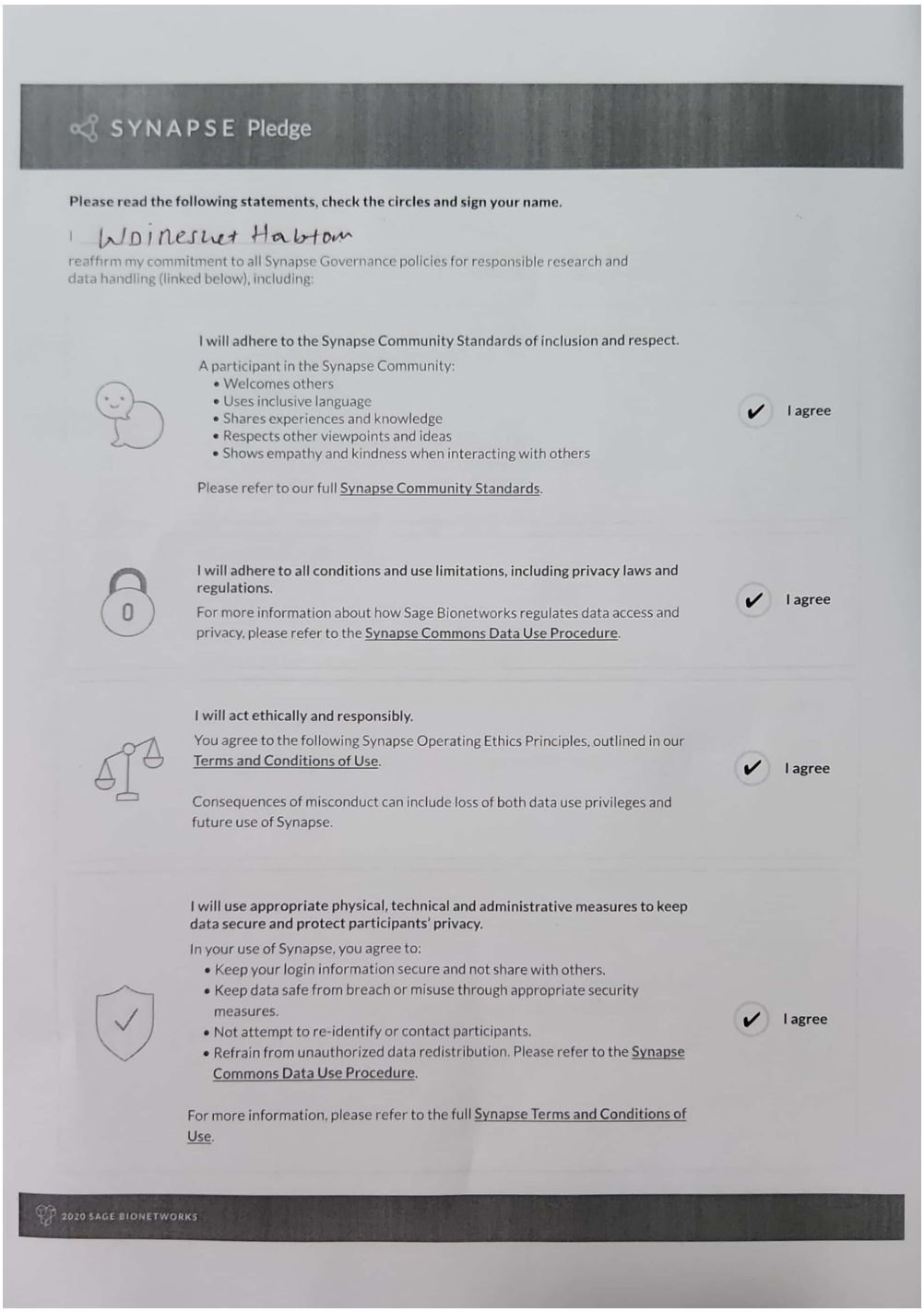
I, Dr. Dawit Assefa Haile, in my capacity as a signing official at the Center of Biomedical Engineering, Addis Ababa Institute of Technology, Addis Ababa University (AAU), hereby provide this letter to attest the identity and affiliation of Woineshet Habtom Bekele.

I can personally verify and confirm her identity as an official representative of our esteemed Center and her main Advisor, I have had the privilege of knowing Woineshet. During this time, she has consistently demonstrated her dedication, integrity, and commitment to her academic pursuits and responsibilities.

I am pleased to state that Woineshet is an esteemed member of the Biomedical Engineering Center and actively engaged in her studies and research endeavors, displaying a strong passion for her chosen field. Therefore, I confidently endorse Woineshet and verify that she is indeed a valued member of the Center of Biomedical Engineering here at the Addis Ababa Institute of Technology, AAU. If you require any further information or have specific queries, please do not hesitate to contact me.

Sincerely,

Signature
Dawit Assefa Haile (BSc, MSc, PhD)
Assist. Professor, Center of Biomedical Engineering
Chair, Biomedical Informatics
Addis Ababa Institute of Technology, AAU
Email: dawit.assefa@aau.edu.et, Phone: +251-(0)922759688



SYNAPSE Pledge

Please read the following statements, check the circles and sign your name.

I Wdinesiet Habtom

reaffirm my commitment to all Synapse Governance policies for responsible research and data handling (linked below), including:

I will adhere to the Synapse Community Standards of inclusion and respect.

A participant in the Synapse Community:

- Welcomes others
- Uses inclusive language
- Shares experiences and knowledge
- Respects other viewpoints and ideas
- Shows empathy and kindness when interacting with others

Please refer to our full Synapse Community Standards.

✔ I agree

I will adhere to all conditions and use limitations, including privacy laws and regulations.

For more information about how Sage Bionetworks regulates data access and privacy, please refer to the Synapse Commons Data Use Procedure.

✔ I agree

I will act ethically and responsibly.

You agree to the following Synapse Operating Ethics Principles, outlined in our Terms and Conditions of Use.

Consequences of misconduct can include loss of both data use privileges and future use of Synapse.

✔ I agree

I will use appropriate physical, technical and administrative measures to keep data secure and protect participants' privacy.

In your use of Synapse, you agree to:

- Keep your login information secure and not share with others.
- Keep data safe from breach or misuse through appropriate security measures.
- Not attempt to re-identify or contact participants.
- Refrain from unauthorized data redistribution. Please refer to the Synapse Commons Data Use Procedure.

For more information, please refer to the full Synapse Terms and Conditions of Use.

✔ I agree