IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 9 December 2024, accepted 14 December 2024, date of publication 17 December 2024, date of current version 26 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3519296

RESEARCH ARTICLE

A Novel Cascaded Approach for Classification of Tuberculosis Using Cough Audio in Real-Time Environment

HAROON MAHMOOD[1], MANAL IFTIKHAR[2], AAMIR WALI[2], ARSHAD ALI[2], AND MARYAM GULZAR[3]

[1]College of Engineering, Al Ain University, Al Ain, United Arab Emirates
[2]School of Computing, National University of Computer and Emerging Sciences, Lahore 54770, Pakistan
[3]SE Department, LUT University, 53850 Lappeenranta, Finland

Corresponding authors: Maryam Gulzar (maryam.gulzar@lut.fi) and Arshad Ali (arshad.ali1@nu.edu.pk)

This work was supported in part by the National University of Computer and Emerging Sciences, Lahore; and in part by LUT University, Lappeenranta, Finland.

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board (IRB) Committee of Mayo Hospital, Lahore.

ABSTRACT Tuberculosis (TB) is an infectious disease primarily impacting the lungs. It spreads through the air when an infected person coughs, sneezes, or talks. Diagnosing TB involves clinical examinations and specialized tests performed by medical professionals. Coughing is a common symptom. The diagnosis of TB involves clinical examinations and specialized tests. However, studies have shown that medical doctors can distinguish between cough sounds associated with different respiratory conditions. Therefore, using artificial intelligence to analyze cough recordings of patients to diagnose TB is a promising research direction. In this study, we propose a customized cascaded approach for diagnosing TB using cough audio. This approach involves a series of models arranged in a sequence, where the output of one model serves as the input for the next. In the first phase, we distinguish between bursts in audio signal as noise or cough. In the second phase, we classify cough as TB or non-TB. Non-TB cough includes both voluntary and non-TB reflex cough. For this study we collected a dataset consisting of cough audio recordings from TB and non-TB patients at Mayo Hospital in Lahore, Pakistan. The recordings were obtained using the AI4LYF DCT application, a fully automated phone-based system, with no manual annotation. We apply statistical classifiers based on spectral and time domain features, both with and without clinical metadata. Through a stratified grouped cross-validation approach, our results show that using cough sounds along with demographic and clinical factors yielded an accuracy of 97% when the random forest was used. Similarly, for all other classifiers, the accuracies were $\geq$ 90% when demographic and clinical data was included (from $\leq 80\hat{)}$. Our findings suggest that our model based on patient data and cough auido could support community health workers and health programs in identifying TB cases more effectively and cost-efficiently.

INDEX TERMS Tuberculosis, cough, audio recordings, machine learning, diagnosis.

I. INTRODUCTION

Tuberculosis (TB) is a chronic infectious disease caused by the bacterium Mycobacterium tuberculosis (MTB). It is one of the top 10 causes of death worldwide, second only to HIV/AIDS, and in 2020, it caused an estimated 1.5 million deaths [1]. TB is a major public health problem, particularly in low- and middle-income countries, where it disproportionately affects vulnerable populations such as people living in poverty, those with HIV/AIDS, and children under five years old [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko.

According to the World Health Organization (WHO), in 2020, there were an estimated 10 million new TB cases globally, of which 9.9 million were cases of pulmonary TB (PTB) and 0.1 million were cases of extrapulmonary TB (EPTB). PTB is the most common form of TB, affecting the lungs, while EPTB affects other parts of the body, such as the lymph nodes, bones, and joints [3]. There is a vaccine available for TB, called the Bacillus Calmette-Guérin (BCG) vaccine, but it does not always provide complete protection against TB [3].

TB is transmitted from person to person through the air when an infected person coughs, sneezes or speaks. When an infected person expels droplets containing MTB bacteria, these droplets can be inhaled by others, who may then become infected. The risk of infection depends on several factors, including the proximity to the infected person, the duration of exposure, and the immune status of the individual [4].

Not everyone who is exposed to MTB bacteria will develop active TB disease. In fact, most people who are infected with MTB bacteria will never develop symptoms or become infectious. However, people with weakened immune systems, such as those with HIV/AIDS, malnutrition, or diabetes, are at increased risk of developing active TB disease [5]. The symptoms of TB can vary depending on the location of the infection. However, common symptoms of PTB include a persistent cough that lasts for more than three weeks, chest pain, coughing up blood or phlegm, weight loss, fatigue, fever, and night sweats [6].

The diagnosis of TB can be challenging, as the symptoms can be similar to those of other respiratory conditions. The diagnosis of PTB typically relies on chest X-rays and sputum smear microscopy. However, these methods have limitations, such as low sensitivity and specificity, subjective interpretation, and the requirement for laboratory infrastructure [7]. In rural areas especially, the unavailability of advanced equipment and expert healthcare providers leads to delayed or missed diagnoses. This delay increases the risk of TB transmission and the development of more severe disease forms, including drug-resistant TB. EPTB can be more difficult to diagnose, as it may not present with pulmonary symptoms. The diagnosis of EPTB often requires more invasive procedures, such as biopsies or cultures of affected tissues [8]. Studies have also shown that medical doctors can distinguish between cough sounds associated with different respiratory conditions. This seems like a promising research direction where artificial intelligence can be used to analyze cough recordings to diagnose TB.

In this study, we propose and develop a cough analysis system for the detection of TB using machine learning techniques like random forest, K-nearest neighbors, etc. By learning patterns from cough sounds and other acoustic features, this model can differentiate between TB and other respiratory conditions. Such systems offer advantages over current methods, including non-invasiveness, objectivity, and potential use as point-of-care tools, particularly in areas with limited laboratory infrastructure. Such a system holds great promise for improving TB diagnosis and control, particularly in resource-limited settings. They can facilitate early detection and treatment, monitor latent TB infections to prevent reactivation, and potentially detect drug-resistant TB, guiding treatment decisions and preventing its spread.

Due to the unavailability of cough audio recordings of TB patients, the collection of such recordings was a challenging task. With the support of health personnel at Mayo Hospital, Lahore, we gathered data from a diverse group of TB patients, taking into account environmental factors like noise, age, gender, smoking history, and vocal characteristics, all in a real-time environment. A mobile application was developed just for collecting cough data from both patients and healthy individuals.

A. PROBLEM AND MOTIVATION

TB is a serious infectious disease that primarily affects the lungs. It fall under top 10 causes of death globally, and in 2021, it was estimated to have caused 1.5 million deaths, mostly in low and middle income countries.

The current methods for diagnosing TB have several limitations, including:

- Low sensitivity and specificity: Chest X-rays and sputum smear microscopy, the two most common methods for diagnosing TB, have low sensitivity, meaning that they may miss some cases of TB. They also have low specificity, meaning that they may falsely identify some people as having TB when they do not.
- Subjective interpretation: The interpretation of chest X-rays and sputum smear microscopy can be subjective, meaning that different doctors may interpret the same results differently. This can lead to inconsistent diagnoses.
- Requirement for laboratory infrastructure: Chest X-rays and sputum smear microscopy require access to laboratory infrastructure, which is not always available in resource-limited settings.

These limitations contribute to delayed diagnosis and treatment of TB, which can lead to increased morbidity and mortality. Delayed diagnosis can allow TB to spread to others and can make the disease more difficult to treat. Additionally, delayed diagnosis can lead to the development of drug-resistant TB, which is more difficult and expensive to treat than drug-susceptible TB.

B. ORGANIZATION OF THE PAPER

The rest of the paper is organized as follows. Section II provides a concise literature review of the few studies on TB detection using CT scans etc., since there was no significant study on TB diagnosis using cough. Section III presents the details of the dataset. Section IV describes the methodology. Finally, sections V and VI provide the results and discussion, respectively. Section VII concludes this work and provides future directions.

II. LITERATURE REVIEW

Sound signal processing supporting disease recognition is not just limited to cough but also has broader applications. For example, sound processing was recently applied to perform recognition of Obstructive Sleep Apnea Hypopnea Syndrome (OSAHS) and to classify/recognize some heart diseases by digital phonocardiograms (PCG) processing [24], [25], [29], [30]. Within TB detection, there are numerous studies on chest X-rays and very few studies on cough to detect whether a patient has a TB or not [9], [11], [12] and [13]. Xu et al. [9] used a feature fusion approach consisting of a combination of 5 traditional features to enhance the accuracy of classifying cough sounds as either indicative of TB or not. Once the features were extracted, Bidirectional Long Short-Term Memory (Bi-LSTM) and Bidirectional Gated Recurrent Unit (Bi-GRU) networks were used for classification. The study experimented with different combinations of features. The best result was obtained when all five features were used.

Ramkumar et al [11] proposed a new system that facilitates the analysis of various medical image segmentation techniques. They used the MR U-net system to segregate between images of TB and non-TB patients. They used lung CT scan medical images from various online resources as a data set. Their system achieved an accuracy of 92.50%. The proposed system still needs images that require at least a one-time hospital visit.

In [12], Pahar et al. describe an approach based on a dataset of 1358 forced cough recordings taken in a developing-world clinic from 16 patients with confirmed active pulmonary TB and 35 patients suffering from respiratory symptoms indicative of TB but confirmed to be TB free. The dataset is not publicly available. Using layered cross-validation, five machine learning classifiers namely logistic regression (LR), support vector machines, k-nearest neighbor, multi-layer perceptrons, and convolutional neural networks were trained and assessed. The highest accuracy of 94% was obtained with linear regression. The data used for TB positives were not of confirmed cases. Additionally, using an N95 mask while recording cough could impact the quality of the recorded audio. Furthermore, the low performance of advanced classifiers like CNNs highlights the limitations of the dataset's small size.

Hyfe's approach is used in [13] for diagnosing tuberculosis (TB) using publicly available training data from the CODA DREAM Challenge. The study considers both demographic and clinical metadata. Results are encouraging, especially for models trained on both audio and tabular data. This study compares machine learning techniques on a diverse and large dataset of cough audio recordings. The findings suggest that cough sounds may be sufficient for prioritizing TB diagnostic evaluation in patients. The study also identifies areas for potential improvement and further research in audio analysis for TB prediction. The inclusion of demographic and clinical factors improved the performance from AUC value of 0.70 to 0.81.such as [14], [15], and [16], and many more exist but lack consideration of cough recordings. A study by [17] used the same CODA DREAM Challenge to detect TB. The study utilized a dataset collected from 2,143 adults residing in seven countries infected with TB. Unfortunately, the study low specificity of only 70%. The use of 0.5-second cough recordings limited the model's ability to capture complete patterns.

A recent study on respiratory disease screening shows that cough classifiers can be promising, though there are concerns about biases in model training and dataset quality [18]. This study used a dataset comprising 33,000 passive coughs and 1600 forced coughs collected in controlled settings. The researchers utilized scalogram images of the passive cough sounds as input to train a ResNet-18-based classifier. Realtime detection was the main challenge for this study which remained unsolved and data was collected and processed in a controlled environment. The main limitations of this paper include the lack of a specified device dedicated to data collection, the absence of a publicly accessible dataset, and a limited extraction of cough audio features, which restricts the depth of information available for analysis.

According to [9], COVID-19 pandemic has shown that cough sounds can be useful for detecting diseases like tuberculosis (TB) with the help of smart technology. TB remains a serious health issue, especially in poorer areas, with around 10 million cases globally in 2021 and rising by 4.5% each year. Common TB tests, like the PPD test, have lower accuracy, with only 77% sensitivity, while better tests like IGRA and Xpert MTB/RIF are often too costly for communities with fewer resources. To address this, researchers are exploring the use of cough sounds as a low-cost screening option. Two types of models, Bi-LSTM and Bi-GRU were used to examine common sound features (like energy levels, pitch, and frequency) and combined these with detailed sound patterns from spectrograms. Spectrograms, which are image representations of sound, have been widely used for analyzing speech waves [19] due to their easier processing capabilities. However, relying solely on spectrograms for detecting tuberculosis (TB) coughs is not feasible. Bi-LSTM alone identified TB patients with 93.99% accuracy, but when combined with spectrogram features, it reached 96.33% accuracy, making it much more effective than traditional PPD tests.

Table 1 provides a summary of existing studies.

III. DATASET

A. IRB APPROVAL

The research proposal, including the data collection methodology, was submitted to the Institutional Review Board (IRB) for approval. This step ensured that the study adhered to ethical standards, particularly in terms of patient consent, confidentiality, and data security. The IRB approval was crucial to proceed with data collection in a manner that respects patients' rights and well-being.

TABLE 1. Literature review of tuberculosis pre-screening.

Ref	Authors	Dataset	Methodology	Results	Notes
[10]	Akanksha Soni et al.	Digital lung medical images	Digital lung medical images	96%	Requires CT scan images
[11]	Ramkumar. M.O et al.	MR U-net for TB image segmentation	Lung CT scan medical images	92.50%	Requires images from hospital visit
[12]	Madhurananda Pahar et al.	Machine learning classifiers for forced cough recordings	1358 cough recordings	94%	Dataset not publicly available
[13]	Kafentzis, George P et al.	Deep learning and audio processing for TB diagnosis using cough sounds	CODA DREAM Challenge dataset	AUC of approximately 0.70 - 0.81	Audio-based, includes metadata
[18]	Sharma, Manuja and Nduba,	Deep learning and audio processing for TB diagnosis using cough sounds	Custome Dataset of Audios and metadata	AUC of approximately 0.87 - 0.88	Audio-based, includes metadata

B. DATA COLLECTION SETUP

For data collection, an Android application named AI4LYF, was designed with user-friendly interfaces, making it easy for hospital staff to operate with minimal training. It was specifically developed for recording cough data. This application was installed on devices attached to tripods, ensuring stability and consistency in the recording process. This setup was strategically placed in designated areas within the participating hospitals, ensuring ease of access for patients and minimal disruption to hospital operations. It included features in starting, stopping, and saving recordings automatically, along with tagging each recording with relevant patient metadata while maintaining patient confidentiality and data security.

As part of the study, the participants were requested to cough, and the cough sounds were recorded using the AI4LYF DCT App. The app guides the participants with a countdown (3-2-1) and prompts them to cough, capturing approximately half a second of the "explosive" sounds within a five-second timeframe. This process is repeated 3 times. It is important to note that the number of solicited coughs may vary for each participant depending on how many times they coughed during each five-second recording interval. Moreover, the act of producing a solicited cough may trigger additional coughing, so the cough files in this dataset comprise a combination of solicited and spontaneous coughs. The sampling frequency of all recordings is 48000 Hz but the signals are sub-sampled for each experiment.

C. DATA COLLECTION GUIDELINES

A set of guidelines was developed for the data collection process. These guidelines covered aspects such as the duration of cough sound recording, patient positioning, and the handling of the audio data. Staff at the medical facilities were trained on these guidelines to ensure consistency and reliability in the data collected across different sites. Some of the guidelines were included:

- Greet patients and inform them about the purpose and process of data collection.
- Sanitize mobile devices before and after sample collection.
- Prioritize personal safety and maintain a hygienic environment, avoiding touching your face.
- Maintain a safe distance from the patient when they are coughing and wait before entering the area.
- Wear gloves and a mask at all times, disposing of gloves properly after handling the phone for data collection.
- Instruct patients not to cover their faces while coughing.
- Ask patients to sanitize their hands after the recording is complete.
- Guide patients to record three cough and breath samples separately.
- Provide instructions on proper coughing and taking deep breaths.
- Before recording, ask the patient to cough once to assess their coughing pattern and offer instructions for improvement if necessary.
- Communicate with patients and advise them to take deep breaths before coughing for a clear recording.
- Maintain a distance of 20cm from the microphone and designate a fixed location for the mobile device during recording.
- Ensure that assistants charge the mobile devices to 100.
- Continue cough and breath recordings even without an internet connection, uploading patient records when the internet is restored.
- In case of issues during recording, take a screenshot and save it.
- Maintain a recording room with minimal background noise.

D. SAFETY MEASURES

To ensure the safety and comfort of both patients and healthcare staff during the data collection process, we implemented several measures. Below is a summary of these measures, focusing on the development and implementation of a specialized Android application for cough data recording:

- Developed a specialized Android application designed for cough data recording.
- The application streamlined the recording process, allowing for efficient, contactless data capture.
- Installed the application on devices attached to tripods, positioned at a safe distance from the patients.
- This setup minimized the need for close physical interaction, reducing the risk of infection transmission - a crucial consideration in a TB-prevalent environment.
- The tripod provided stability and consistency in recording, ensuring that the cough sounds were captured accurately without necessitating manual handling of the recording device by healthcare workers or patients.

E. TYPES OF PATIENTS

The clinical research included the evaluation of all individuals who tested positive after conducting X-rays and an Acid-Fast Bacillus (AFB) smear test. Those who had a new or worsening cough that persisted for a minimum of two weeks were selected to participate in the study.

Patients suffering from tuberculosis can be categorized into distinct types based on the stage of their disease and the timing of their diagnosis. Understanding these categories is crucial for tailoring appropriate medical interventions and for the accurate collection and analysis of data.

- New: Patients diagnosed with TB for the first time, either by a doctor or through tests, who come to the data collection premises for sampling. Some may have taken medicine before coming to the office.
- Follow-up: Patients whose data has been collected once and return to the hospital for a second checkup and sampling. This process can repeat, with visits incrementally numbered accordingly.
- Suspect: Patients who visit the hospital for the first time after experiencing disease symptoms. They haven't seen a doctor or taken any medicines. Suspects can also be individuals who accompany potential TB patients.
- Presumptuous: Individuals diagnosed by a doctor as potential TB patients based on symptoms but have not yet tested positive by medical tests. They assume themselves to be healthy.

Additionally, the data was also collected from non-TB patients. This consisted of two categories. Firstly, healthy patients, i.e., individuals diagnosed as negative by a doctor or medical tests. They have no known medical conditions or symptoms. This data served as a control group, essential for training the machine learning model to distinguish between coughs associated with TB and those from healthy individuals. Secondly, patients who had a cough and were diagnosed with a condition that was not TB.

Patients unwilling to participate or had difficulty providing cough samples were excluded from the study.

F. DATA STATS

The collected data included hundreds of hours of cough sound recordings from both TB patients and healthy individuals, along with associated patient metadata. This comprehensive dataset provided the robust foundation for training and testing machine learning algorithms, ensuring the reliability and validity of the findings. In this study, there were 145 participants (435 samples), consisting of 123 (369 samples) males and 22 (66 samples) females. The participants' ages ranged from 40 to 70 years, with an average age of 50 years. We collected three 6-second cough samples from each patient, aiming to capture a variety of cough sounds. We also tried to include patients across different ages and demographics to make the dataset as representative as possible. Based on our analysis, TB is more common in men in Pakistan, likely due to their higher exposure to the public while working. Additionally, TB is more frequent in people aged 40 and above, which we considered in our dataset.

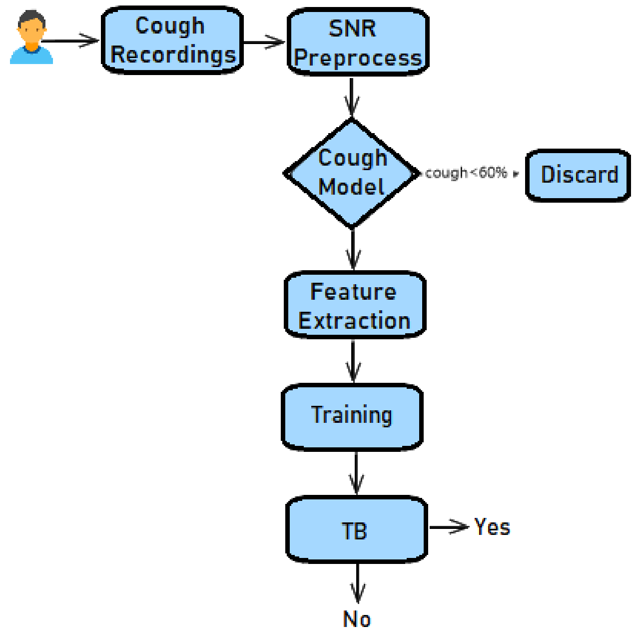


FIGURE 1. **Proposed methodology diagram.**

IV. PROPOSED METHODOLOGY

Figure 1 shows a high-level overview of the suggested methodology. The raw audio signal goes through a series of preprocessing steps to distinguish cough audio from non-cough audio. Once the cough model categorizes a recording as having a cough with a reasonable confidence value ($\geq 60\%$), it is passed to the feature extractor where a series of features are extracted. Finally, these features are used to train multiple classifiers that are evaluated using test samples. To increase the size of training data, data augmentation approaches for audio were also utilized.

A. DATA PREPROCESSING

In our study on detecting tuberculosis (TB) from cough sounds, we observed that machine learning models struggle with very large or extended audio recordings. These files often contain significant amounts of background noise and silence, which can confuse the models and reduce their accuracy. To address this issue, we implemented a two-stage data preparation approach, enhancing both the quality of the input and the model's focus on relevant data.

In the first stage, we preprocess the audio using Signal-to-Noise Ratio (SNR) techniques to remove extraneous background sounds, enhancing the clarity of potential cough sounds. This step significantly improves the data quality by amplifying relevant signals over noise.

The second stage employs a specialized deep learning model we developed, called the "Cough vs. Non-Cough" model. This model effectively isolates cough sounds by filtering out non-cough sounds and silence, ensuring that only meaningful cough segments are retained. By working with this refined dataset, our machine learning models can more accurately identify which coughs are likely associated with tuberculosis.

This dual-step preprocessing approach—SNR enhancement followed by cough segmentation—ensures that our tuberculosis detection models operate with cleaner, more focused data. As a result, our models achieve improved accuracy and reliability in distinguishing TB-related coughs from non-TB coughs, minimizing potential errors in the detection process.

1) SIGNAL-TO-NOISE RATIO

In the preprocessing stage, the raw cough audio data is segmented into discrete chunks using a sliding window technique. Typically audio samples were 6s long. A window of 2s with a 50% overlap was applied, ensuring each chunk captures a manageable duration of the audio for detailed analysis. Each chunk is evaluated for its signal-to-noise ratio (SNR) [20] to quantify the clarity of the audio signal relative to the background noise.

To isolate meaningful audio segments, the calculated SNR values were compared against a predetermined threshold. This threshold was established through empirical testing and guided by relevant literature. Chunks exceeding this SNR threshold are classified as 'bursts.' A burst is defined as any chunk with a sufficiently high SNR value, indicating it may be a cough or other noise. Any chunk identified by SNR as burst was stored in a separate file. This segmentation enables the isolation of potentially meaningful audio segments for further analysis. Importantly, while all bursts are high-SNR chunks, not all bursts represent cough signals, necessitating an additional classification step.

To address the limitation of SNR in differentiating between cough and non-cough sounds, a two-step approach was implemented. Initially, the SNR thresholding identifies chunks with significant acoustic energy. Subsequently, these chunks are passed through a 'Cough vs. Non-Cough' deep learning model. This model was trained using annotated cough and noise data and leverages advanced audio features extracted using the Librosa library. It classifies each burst as either a cough or noise. One of the author and a specialist manually annotated the data by listening to each burst.

2) COUGH VS NON-COUGH MODEL

In addition to SNR, we also propose a Cough vs Non-Cough stage where audio cough generated by SNR preprocessing was analyzed to identify noise bursts. This phase utilizes the Librosa package, a comprehensive and versatile toolkit designed for audio signal processing. Librosa's ability to provide a rich feature space is integral to our research, as it captures vital aspects of the cough audio signals. Its array of functionalities enables the extraction of various audio features crucial for a detailed analysis of cough sounds.

These features include aspects like frequency, intensity, and temporal characteristics of the cough, each offering unique insights critical for tuberculosis detection. Librosa's application in our study not only simplifies the complex task of audio analysis but also ensures a thorough and nuanced examination of cough sounds.

B. FEATURE EXTRACTION

In this study, we employed a series of audio features of librosa [21] to analyze cough signals. While existing literature offers various feature sets tailored for audio, music, and speech signal processing, these features might not be directly suitable for cough signals due to their non-stationary nature. Cough signals rapidly change over time, necessitating the use of short segments, commonly referred to as frames, which exhibit approximately stationary characteristics. Stationarity implies that temporal and spectral characteristics within a frame remain relatively constant. The frame length should be long enough to ensure reliable feature estimation while remaining short enough to represent the entire frame accurately. In our case, we adopted a standard frame length of 20 - 50 ms. For instance, a 50-ms window slid over the cough waveform with a step of 25 ms, resulting in a 50% overlap between consecutive frames. Additionally, to improve the spectral properties, each frame was windowed using a Hamming window, which is a time-domain function designed to mitigate spectral leakage.

1) MFCC

In our approach we used MFCC, a signal processing technique specifically designed (or used) in speech analysis, for tuberculosis detection. It transforms audio signal frames into a 12-dimensional feature vector, capturing relevant acoustic patterns. We detected MFCC's compact representation can be used to enhance accuracy in non-invasive tuberculosis screening, aiding early disease detection and healthcare advancements. MFCCs were used in this study as a feature extraction method rather than as a tuberculosis-specific detection technique. MFCC can be calculated by using:

$$\mathrm{MFCC} = \mathrm{DCT}\left(\log\left(\sum_{k=1}^{N} |X(k)|^2 H_m(k)\right)\right) \tag{1}$$

where $H_m(k)$ represents the Mel-filter for the m-th filter, $X(k)$ is the FFT output of the audio signal and m is the index of the Mel filter. The size of the audio window is 2 seconds. The size of the FFT was 32,768 samples with 40 mel filters. The number of MFCC coefficients was 12.

2) RMS ENERGY

The RMS energy, a commonly used feature in the field of audio signal processing, is calculated as the square root of the mean of the squared signal values within short-time intervals. It provides a crucial measure of the signal's loudness and finds widespread application in tasks such as audio segmentation and classification. Consequently, the RMS energy proves to be a valuable asset in analyzing acoustic events, including coughs and other audio-related phenomena. RMS can be calculated by using:

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x[n-1]^2} \tag{2}$$

where x[n] is the amplitude of the audio sample and N is the number of samples.

3) SPECTRAL CENTROID

The spectral centroid, in conjunction with other audio features, enables to develop robust classification models for cough sound recognition and pattern identification. The spectral centroid is a measure of the "center of mass" of the spectrum and describes the perceived brightness of a sound. This Feature can be calculated by:

$$\text{Spectral Centroid} = \frac{\sum_{k=1}^{N} f(k)\cdot|X(k)|}{\sum_{k=1}^{N}|X(k)|} \tag{3}$$

where $f(k)$ is the frequency of the k-th FFT bin and $X(k)$ is the FFT magnitude of the signal at k-th bin.

4) SPECTRAL BANDWIDTH

The spectral bandwidth is a crucial acoustic feature used in audio signal processing, offering valuable insights into the frequency distribution of an audio signal. It measures the width of the frequency content and is calculated from the power spectral density. In our research work, we explore the significance of spectral bandwidth for a specific audio analysis task (e.g., cough detection, speech recognition, or music genre classification). We demonstrate how this feature can enhance the accuracy and robustness of our proposed audio processing algorithm or model. By comparing it with other acoustic features, we highlight its unique characteristics and complementary nature. Our findings contribute to the broader field of audio signal processing, providing practical implications for speech, music, and audio-based machine learning applications.

$$\text{Bandwidth} = \sqrt{\frac{\sum_{k=1}^{N}(f(k)-\text{Spectral Centroid})^2\cdot|X(k)|}{\sum_{k=1}^{N}|X(k)|}} \tag{4}$$

5) SPECTRAL ROLL OFF

We explore the application of spectral roll-off as one of the key features in our audio analysis methodology. We investigate its effectiveness in classifying and recognizing various sound events in real-world scenarios. The spectral roll-off feature, combined with other audio descriptors, contributes to the development of a robust and efficient audio processing system that can be leveraged in diverse fields, ranging from environmental monitoring to automatic speech recognition. Spectral Roll-off f_r is defined as:

$$\sum_{k=1}^{k_r}|X(k)| = R\cdot\sum_{k=1}^{N}|X(k)| \tag{5}$$

where $|X(k)|$ is the magnitude of the k-th frequency bin, N is the total number of frequency bins, R is the threshold value set to 0.95, k_r is the bin index where the cumulative energy reaches R of the total energy and f_r is the frequency corresponding to k_r.

6) ZERO CROSSING RATE (ZCR)

For a given audio signal segment, ZCR is computed by counting the number of times the signal crosses the zero amplitude line within the specified time window. Higher ZCR values indicate that the signal undergoes frequent changes in its waveform, while lower ZCR values suggest smoother or more periodic patterns. In research, ZCR is often used in conjunction with other audio features to characterize various audio phenomena, such as identifying speech and non-speech regions in audio streams, distinguishing between different musical instruments, detecting onsets and offsets in music, and even detecting specific sound events in environmental audio recordings.

$$ZCR = \frac{1}{N-1}\sum_{n=1}^{N-1}|\text{sign}(x[n]) - \text{sign}(x[n-1])| \tag{6}$$

7) MEL SPECTROGRAM

In this work, we employ the Mel spectrogram to preprocess audio data before feeding it to machine learning models. This spectral representation aids in addressing the challenges posed by the dynamic nature of audio signals, allowing us to achieve higher accuracy and robustness in various audio-related tasks. Through our experiments, we demonstrate the effectiveness of the Mel spectrogram in enhancing the performance of our proposed audio analysis and recognition algorithms, leading to valuable insights and contributions to the field of audio signal processing and machine learning applications. Visualization of Mel Spectrogram of Audio Sample is provided in Figure 2.

8) CHROMA SHORT-TIME FOURIER TRANSFORM

Chroma Short-Time Fourier Transform is a feature extraction technique used in audio signal processing and music information retrieval. It converts audio signals into a chroma gram, which represents pitch content in a compact form.

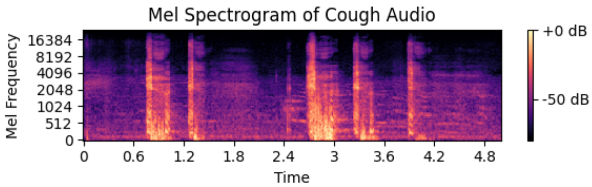


FIGURE 2. **Visualisation of mel spectrogram of audio sample.**

The algorithm employs the Short-Time Fourier Transform to analyze the spectral content at different time points. The resulting chroma gram is a 12-dimensional vector, representing the 12 chroma pitch classes. The energy in each chroma bin c for frame n is calculated by:

$$\text{Chroma STFT} = \sum_{k \in \mathrm{B(c)}} |X(n, k)| \tag{7}$$

where $B(c)$ is the set of frequency bins k that map to chroma bin c.

9) CHROMA CQT

Chroma CQT is a powerful feature extraction method widely used in audio signal processing and music information retrieval. It is based on the Chromagram representation, which encodes the twelve distinct pitch classes in the musical octave. Unlike traditional chroma methods that rely on the discrete Fourier transform (DFT), Chroma CQT employs the constant-Q transform (CQT) to compute the chroma features. The CQT, being a logarithmically spaced frequency transform, provides better resolution in the low-frequency range and is well-suited for capturing tonal information in audio signals. By computing the CQT on an audio signal and mapping the resulting magnitude spectra to the twelve chroma bins, Chroma CQT offers an effective and robust representation of the harmonic content in music and speech. This feature extraction technique finds wide applications in various tasks, including chord recognition, music genre classification, and content-based music retrieval, due to its ability to characterize the tonal characteristics of audio signals and its resistance to key changes and timbral variations. This feature can be computed by using:

$$\text{Chroma CQT} = \sum_{k=1}^{M} |X(k)|\, Q_m(f(k)) \tag{8}$$

where $Q_m(f)$ is the frequency response of the m-th filter in a Constant-Q Transform.

10) CHROMA CENS

Chroma Constant-Q Normalized Energy Subbands (Chroma CENS) is an advanced feature extraction technique used in audio signal processing and music analysis. It improves upon the standard chroma representation by incorporating constant-Q scaling and normalization, making it robust to variations in loudness and background noise. Chroma CENS has proven to be highly effective in tasks like pitch estimation, key detection, music similarity analysis, and chord recognition. Its accurate tonal information and resistance to noise make it valuable for audio analysis and music retrieval research. Chroma CENS is calculated by mapping audio to a 12-dimensional chroma vector, normalizing energy within each band, and smoothing over time.

11) CONTRAST

Next, we explore the concept of contrast, a fundamental element that plays a crucial role in understanding various phenomena. Contrast involves the differentiation between different entities, elements, or attributes within a given context. By highlighting variations, disparities, or distinctions, contrast helps reveal patterns, emphasize significance, and enrich our comprehension of the subject matter. Throughout our investigation, we delve into the diverse applications of contrast, analyzing its impact on visual perception, language comprehension, and decision-making processes. By shedding light on this pivotal aspect, our research aims to contribute valuable insights that can inform and enhance various fields of study, ultimately advancing our understanding of the world around us. Spectral Contrast can be calculated by:

$$\text{Spectral Contrast} = \frac{1}{B} \sum_{b=1}^{B} (\text{Peak}_b - \text{Valley}_b) \tag{9}$$

where B is the number of spectral bands. $Peak_b$ and $Valley_b$ are peak and valley amplitudes in the b-th spectral band.

12) TONETTZ

Tonnetz is a valuable concept in music theory and computational music analysis, which facilitates the understanding of musical harmony and chord relationships. Developed by Euler in the 18th century and later refined by Riemann, the Tonnetz represents musical pitches and harmonies as a geometric lattice or network. In this lattice, each node corresponds to a specific musical pitch, and the edges connect related pitches based on specific harmonic transformations. By visually mapping harmonic progressions onto the Tonnetz, music analysts and researchers gain profound insights into the underlying structure and chord progressions of compositions. Additionally, in computational music analysis, Tonnetz serves as a foundational tool for building sophisticated algorithms that detect key changes, chord transitions, and harmonic complexity in musical pieces. Its versatile applications across diverse musical genres make it a crucial resource for exploring the intricacies of music and unraveling its harmonic mysteries. Tonnetz is derived from harmonic relationships and often represented as six dimensions in a tonal space related to pitch. It's computed using the pitch class profile (PCP) for tonality analysis.

13) FEATURE SET

Once all the features had been extracted, they were combined to form the final feature set. In total, the feature space consisted of 206 features.

C. DATA AUGMENTATION

Data augmentation is a technique used to artificially increase the size and diversity of a dataset by transforming or modifying existing data samples. This technique has been used for all kinds of data images [26], [33], including medical images [22], [27], text [23], [28], audio or speech, and videos [31], [32]. Since the cough audio data is limited, this can be particularly beneficial for this study. During recording, it was observed that since cough recordings were taken in different places, they had different background noises. The most common background noise in healthy recordings was white noise. Whereas, in TB recordings, the most common noises were birds chirping and fans.

To make more data, white noise was added from healthy recordings of TB coughs. Similarly, bird chirping and fan noises were added from TB recordings to healthy coughs. By mixing these noises with the original recordings, more samples were generated which made the dataset larger and more diverse.

D. TRAIN AND TEST SPLIT

In our study, we used data from 97 patients (291 samples, 580 chunks) to train the model and data from approximately 48 patients ((144 samples, 290 chunks) for testing. This split allowed us to validate the performance of the model on unseen data effectively. Additionally, to further assess the model's practicality and accuracy, we conducted real-time testing in local medical facilities, providing insight into its applicability in real-world settings.

E. CLASSIFICATION MODELS

A cascaded approach was used to identify cough as TB. First, a model was used to distinguish a burst as noise or cough. Then another model was used to classify the cough as TB. For the second stage, five different classifiers were experimented with. These models were placed in series with the cough vs. noise model one at a time. Some of these classifies are discussed next.

1) SUPPORT VECTOR CLASSIFICATION (SVC)

Support Vector Classification (SVC), a specialized form of Support Vector Machines (SVM), is particularly beneficial in distinguishing tuberculosis (TB) from non-TB cases using cough characteristics. Its ability to handle high-dimensional data makes it adept at analyzing complex cough features, such as duration, frequency, and sound patterns. We picked SVC because it works well even when we don't have a lot of data, which is common in medical studies. We also tested other machine learning models, but Random Forest and SVC gave us comparatively good results for cough classification.

SVC is especially effective in medical research scenarios where datasets are often limited in size, as it maintains high performance even with smaller sample sizes. One of the key strengths of SVC is its robustness against overfitting, a critical factor in medical classification tasks where patient data can vary significantly. This is achieved through the use of various kernel functions, allowing the algorithm to find the most effective way to separate TB from non-TB cases in a multidimensional space. Moreover, SVC's principle of margin maximization ensures a clear and distinct decision boundary, enhancing the reliability of the classification. Its versatility in handling different data types and integration with other diagnostic features further empowers researchers and clinicians to make more accurate diagnoses.

In practice, implementing SVC would involve collecting a comprehensive dataset of cough characteristics from TB and non-TB patients, followed by rigorous data preprocessing, before applying the SVC algorithm for classification, potentially improving the accuracy and efficiency of TB diagnosis. It can be derived using the following function:

$$f(x) = \text{sign}(w \cdot x + b)$$

2) LOGISTIC REGRESSION

In the quest to distinguish tuberculosis (TB) from non-TB conditions using cough characteristics, logistic regression (LR) emerges as a powerful statistical tool. This method leverages the binary nature of the TB vs. non-TB classification, making it ideal for handling categorical outcomes. By analyzing various cough attributes such as duration, frequency, and nature (productive or dry), LR can efficiently calculate the probability of a patient having TB. For instance, longer duration and higher frequency of cough might be assigned greater weights in the LR model, indicating a higher likelihood of TB. This approach not only enhances the diagnostic accuracy but also allows for the incorporation of other relevant variables like patient age, smoking history, and previous TB exposure, offering a comprehensive risk assessment. The simplicity and interpretability of LR models make them particularly useful in clinical settings, where they can aid healthcare professionals in making informed decisions about further testing and treatment for suspected TB cases. By applying LR to cough characteristics, we can potentially streamline the TB diagnostic process, improving early detection and thereby contributing to more effective TB management and control.

3) RANDOM FOREST

Random Forest, an advanced machine-learning technique, holds significant potential in classifying tuberculosis (TB) versus non-TB cases using cough characteristics. By integrating numerous decision trees to make a more accurate and robust model, Random Forest can analyze complex patterns in cough data, such as frequency, duration, and sound features. This ensemble approach enhances predictive accuracy and reduces the risk of overfitting, common in singular decision tree models. In the context of TB diagnosis, Random Forest can efficiently process and differentiate subtle variations in cough sounds between TB and non-TB patients. It can handle a large dataset comprising various cough attributes, demographic data, and clinical symptoms, providing a comprehensive analysis. This capability is particularly beneficial in screening large populations, where it's crucial to accurately identify TB cases for early intervention while minimizing false positives. Additionally, Random Forest's feature importance ranking helps in identifying the most predictive cough characteristics, which can guide clinicians in refining diagnostic criteria. Overall, the application of Random Forest in cough-based TB classification represents a promising advance in leveraging machine learning for more precise and efficient disease detection. Random Forest was chosen because it uses lots of small decision trees together, which helps it be more accurate and spot patterns in cough sounds, like the length and frequency of a cough. This is helpful in telling TB and non-TB cases apart. Random Forest also works well with larger amounts of data and lowers the chance of errors.

4) MULTI-LAYER PERCEPTRON

A multi-layer perceptron consists of a series of layers: 1 input, 1 or more hidden and 1 output layer. Each layer consists of nodes called perceptrons that mimic a neuron in the human brain. Hence the name MLP. In this study, we employed a two hidden layer MLP with 16 nodes each. The hidden layers consisted of relu activation function and the Adam's optimizer was used.

F. KNN

Finally, we also used the K-Nearest Neighbors model.The K-Nearest Neighbors (KNN) Classifier is a useful tool for detecting tuberculosis (TB) by analyzing cough sounds. This method works by comparing the cough sound of a person with a collection of known cough sounds from people who either have TB or don't. KNN then finds the most similar cough sounds to make a prediction. TB coughs often have distinct characteristics, like their duration, intensity, and unique sound patterns, which are different from coughs caused by other illnesses. By identifying these patterns, KNN can help in detecting TB. This approach doesn't require heavy computations and is easy to use, which makes it especially helpful in areas with limited medical resources. Since KNN can quickly analyze cough sounds, it provides a fast, non-invasive way to help identify possible TB cases. This can reduce the need for more expensive and time-consuming tests, like chest X-rays or sputum tests. With KNN, healthcare workers can identify TB earlier, allowing for faster treatment and helping to prevent the spread of the disease, particularly in areas where TB is widespread.

V. RESULTS

This section provides results obtained by using our chosen classifiers namely SVC, linear regression, multilayer perceptron, K-Nearest Neighbor and random forest. The confusion matrices presented reflect the results from the testing dataset only, which consisted of a subset of our full recordings. Recall that we collected 3 recordings each from 145 patients. Then these recordings were split into chunks resulting in 890 samples. We utilized around 290 of these audio chunks specifically for classifier testing to evaluate model accuracy. Each chunk represents a distinct segment of cough sounds, typically a few seconds in duration.

A. SUPPORT VECTOR

The use of a SVC plays a vital role in detecting tuberculosis through the analysis of cough audio. The results obtained are quite impressive, showing a 95% accuracy rate on the test dataset. The confusion matrix accompanying the results effectively demonstrates the model's ability to distinguish between cases of TB with 143 negatives and 132 true positives. However, there were misclassifications observed, including 2 false positives and 13 false negatives.

To provide insight into these findings, a detailed classification report is presented, which includes precision, recall, and F1 scores for each class. Specifically for the *TB Not* class, precision is measured at 92%, recall at 99%, and F1 score at 95%, confirming the model's competence in this area. Similarly, for the *TB Likely* class, precision and recall rates are reported as 99% and 91%, respectively, leading to an F1 score of 95%. The overall accuracy rate of 95% speaks to the model's proficiency in detecting tuberculosis cases across classes.

Figure 3(a) and Table 2 present the confusion matrix and classification report respectively for SVC.

TABLE 2. Classification report for SVC.

	Precision	Recall	F1-Score
TB_Not_likely	0.92	0.99	0.95
TB_likely	0.99	0.91	0.95
Accuracy			0.95
Macro avg	0.95	0.95	0.95
Weighted avg	0.95	0.95	0.95

B. LOGISTIC REGRESSION

Another important technique employed in tuberculosis detection through cough analysis is logistic regression (LR), which effectively uncovers patterns that indicate the presence of this disease. The matrix that shows the confusion and the report on classification highlights the effectiveness of LR, emphasizing its ability to have precision and recall for both ''TB likely'' and ''TB not likely'' categories. The weighted average F1 score of 0.92 demonstrates that the model strikes a balance, between precision and recall, proving its capability to make predictions. Figure 3(b) and Table 3 present the confusion matrix and classification report respectively for LR.

TABLE 3. Classification Report for LR.

Class	Precision	Recall	F1-Score
TB_Not_likely	0.92	0.92	0.92
TB_likely	0.92	0.92	0.92
Accuracy			0.92
Macro avg	0.92	0.92	0.92
Weighted avg	0.92	0.9	0.92

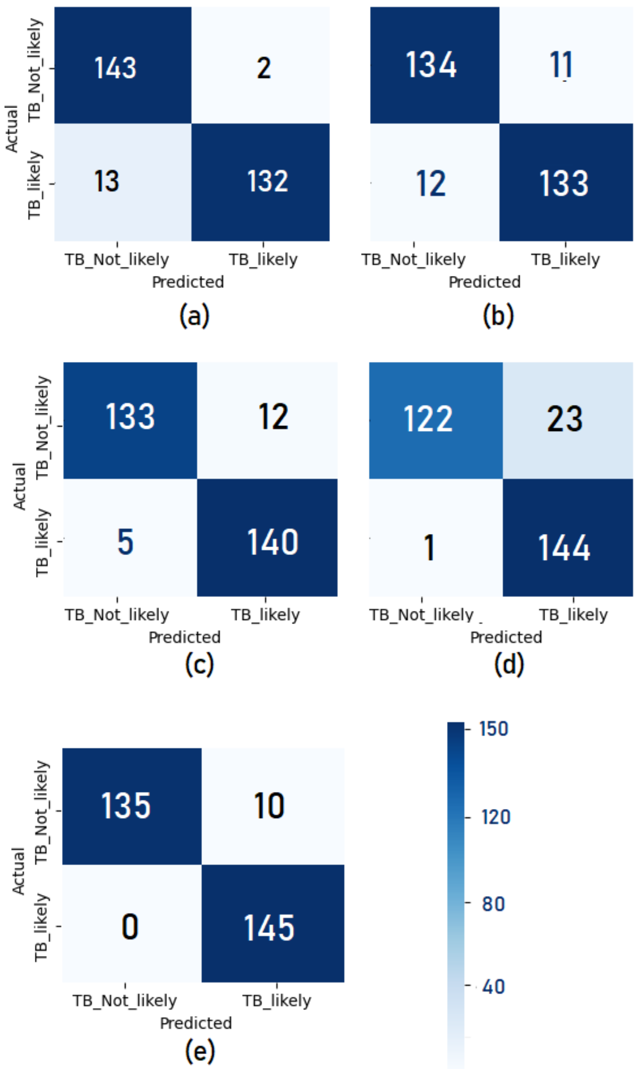


FIGURE 3. Confusion matrices for (a) SVC, (b) Logistic regression, (c) MLP, (d) KNN, and (e) Random forest.

C. MULTI-LAYER PERCEPTRON

MLP assumes a pivotal role due to its capacity to capture complex relationships within intricate datasets, particularly crucial for cough audio features and their connections to tuberculosis. The confusion matrix showcases MLP's ability to classify instances correctly, with high precision, recall, and F1 scores outlined in the classification report. The model's 94% accuracy underscores its competence in tuberculosis detection. Figure 3(c) and Table 4 present the confusion matrix and classification report respectively for MLP.

D. K-NEAREST NEIGHBORS

The K-Nearest Neighbors (KNN) Classifier excels in deciphering complex patterns, leading to an accuracy of 92%. The model showcases precision, recall, and F1 scores, emphasizing its ability to make accurate predictions and handle various cases.

TABLE 4. Classification report for MLP.

Class	Precision	Recall	F1-Score
TB_Not_likely	0.96	0.92	0.94
TB_likely	0.92	0.97	0.94
Accuracy			0.94
Macro avg	0.94	0.94	0.94
Weighted avg	0.94	0.94	0.94

Figure 3(d) and Table 5 present the confusion matrix and classification report respectively for KNN.

TABLE 5. Classification report for KNN.

Class	Precision	Recall	F1-Score
TB_Not_likely	0.99	0.84	0.91
TB_likely	0.86	0.99	0.92
Accuracy			0.92
Macro avg	0.93	0.92	0.92
Weighted avg	0.93	0.92	0.91

E. RANDOM FOREST

Lastly, the Random Forest Classifier proves indispensable in managing intricate data patterns. Its accuracy of 97% is substantiated by the confusion matrix and classification report, highlighting precision, recall, and F1 scores for both the "TB likely" and "TB Not likely" categories.

Figure3(e) and Table 6 present the confusion matrix and classification report respectively for RF.

TABLE 6. Classification report for random forest.

Class	Precision	Recall	F1-Score
TB_Not_likely	0.99	0.93	0.96
TB_likely	0.94	1.00	0.97
Accuracy			0.97
Macro avg	0.96	0.97	0.96
Weighted avg	0.96	0.97	0.97

VI. DISCUSSION

The use of SVC plays vital role, in detecting tuberculosis through the analysis of cough audio. The results obtained are quite impressive, showing a 95% accuracy rate on the test dataset. The confusion matrix accompanying the results effectively demonstrates the model's ability to distinguish between cases of TB with 143 negatives and 132 true positives. However, there were misclassifications observed including 2 false positives and 13 false negatives. To provide insight into these findings a detailed classification report is presented, which includes precision, recall, and F1 scores for each class. Specifically for the "TB Not" class precision is measured at 92% recall at 99%, and F1 score at 95%, confirming the model's competence in this area. Similarly, for the "TB Likely" class, precision and recall rates are reported as 99% and 91%, respectively, leading to an F1 score of 95%. The overall accuracy rate of 95% speaks to the model's proficiency in detecting tuberculosis cases across classes.

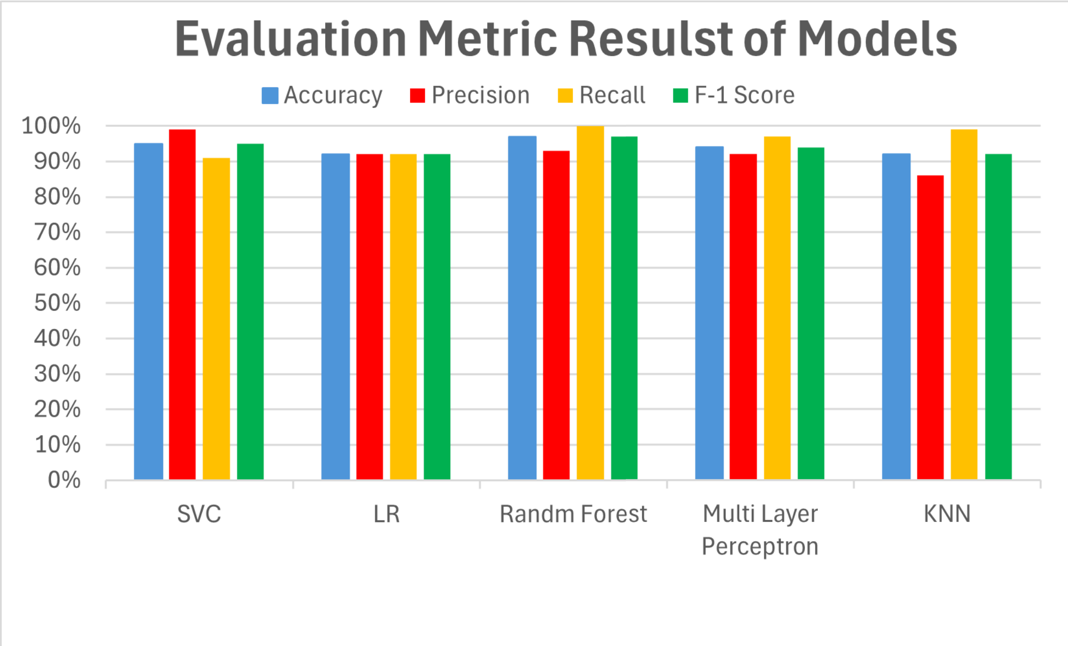


FIGURE 4. Comparison of results of classifiers.

Another important technique employed in tuberculosis detection through cough analysis is Logistic Regression (LR), which effectively uncovers patterns that indicate the presence of this disease. The matrix that shows the confusion and the report on classification highlights the effectiveness of LR, emphasizing its ability to have precision and recall for both "TB likely" and "TB not likely" categories. The weighted average F1 score of 0.92 demonstrates that the model strikes a balance, between precision and recall, proving its capability to make predictions.

Multi-Layer Perceptron (MLP) assumes a pivotal role due to its capacity to capture complex relationships within intricate datasets, particularly crucial for cough audio features and their connections to tuberculosis. The confusion matrix showcases MLP's ability to classify instances correctly, with high precision, recall, and F1 scores outlined in the classification report. The model's 94% accuracy underscores its competence in tuberculosis detection. The K-Nearest Neighbors (KNN) Classifier excels in deciphering complex patterns, leading to an accuracy of 92%. The model showcases precision, recall, and F1 scores, emphasizing its ability to make accurate predictions and handle various cases.

Lastly, the Random Forest Classifier proves indispensable in managing intricate data patterns. Its accuracy of 97% is substantiated by the confusion matrix and classification report, highlighting precision, recall, and F1 scores for both the "TB likely" and "TB Not likely" categories.

Figure 4 provides a comparison of various applied models against evaluation metrics like accuracy, precision, recall and F1 score 4.

VII. CONCLUSION AND FUTURE WORK

In conclusion, this research investigated the use of cough audio to diagnose tuberculosis utilizing advanced machine learning techniques to derive meaningful insights from this unique diagnostic approach. Among the models evaluated, the Support Vector Classifier (SVC) and Random Forest emerged as the most effective, significantly enhancing the accuracy and interpretability of tuberculosis diagnosis.

The SVC demonstrated exceptional capability in capturing intricate, nonlinear patterns within high-dimensional data, achieving an accuracy of 95%. Its ability to effectively distinguish between tuberculosis and non-tuberculosis cases underscores its robustness inaccurate classification. On the other hand, the Random Forest Classifier, with its impressive accuracy of 97%, proved to be a standout model. It excelled in managing complex data patterns, mitigating overfitting, and providing valuable insights into feature importance. The Random Forest's performance highlights its potential to transform tuberculosis detection by improving accuracy and reliability in cough audio analysis.

This research sets the stage for several promising future directions. Refining the current models and exploring hybrid approaches that integrate the strengths of multiple algorithms could further enhance accuracy and robustness. Expanding the dataset to include a broader and more diverse sample could improve the generalizability of the models, making them more applicable to real-world scenarios. Additionally, investigating new features and employing advanced audio processing techniques may reveal hidden patterns and lead to even more precise diagnostics.

Incorporating explainable AI can further improve the transparency and clinical applicability of these models. This would allow a smooth and acceptable integration into healthcare settings. As technology advances, real-time monitoring systems and portable devices could revolutionize tuberculosis detection, making it faster and more non-invasive. The progress made in this research represents just the beginning of a future where machine learning and medical diagnostics are seamlessly integrated, offering the potential to significantly advance tuberculosis detection and other medical fields. Tuberculosis (TB) is a disease that takes a long time to treat. Our research so far has looked at how TB shows up in people with a productive cough (coughing with mucus). However, many TB patients have an unproductive or voluntary cough (a dry cough). Our future work will focus on detecting TB from these kinds of coughs, as they are harder to identify.We plan to gather more data on unproductive and voluntary coughs from TB patients. Next, we will develop computer programs to analyze these cough sounds and see if we can find patterns that indicate TB. We will then test these programs to make sure they are accurate and work well with different groups of people. Finally, we want to see if these tools could be used in real-life TB testing to help find the disease earlier and improve treatment.By focusing on this new approach, we hope to fill an important gap in TB diagnosis and make it easier to detect TB in people who don't show obvious symptoms.

DATA AVAILABILITY

Data consisting of cough recordings was collected from a local hospital after fulfilling all formalities. However, the data will be provided to researchers on demand, but will not be made publicly available.

REFERENCES

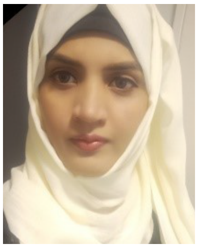
[1] K. Sayood, *Introduction to Data Compression*. Boston, MSA, USA: Newnes, 2012.
[2] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[4] J. Lewis, "Fast normalized cross-correlation," in *Vision Interface*. QC, Canada: Canadian Image Processing and Pattern Recognition Society, 1995, pp. 120–123.
[5] T.-W. Loong, "Understanding sensitivity and specificity with the right side of the brain," *Brit. Med. J.*, vol. 327, no. 7417, pp. 716–719, 2003.
[6] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.
[7] K.-H. Thung and P. Raveendran, "A survey of image quality measures," in *Proc. Int. Conf. Tech. Postgraduates (TECHPOS)*, Dec. 2009, pp. 1–4.
[8] H. Sheikh and A. C. Bovik, "Information theoretic approaches to image quality assessment," in *Handbook of Image and Video Processing*. Canada: Elsevier, 2005, pp. 975–989.
[9] 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, May 2024, Art. no. e0302651.
[10] A. Soni, A. Rai, and S. K. Ahirwar, "Mycobacterium tuberculosis detection using support vector machine classification approach," in *Proc. 10th IEEE Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Jun. 2021, pp. 408–413.
[11] M. O. Ramkumar, D. Jayakumar, and R. Yogesh, "Multi res U-net based image segmentation of pulmonary tuberculosis using CT images," in *Proc. 7th Int. Conf. Smart Struct. Syst. (ICSSS)*, Jul. 2020, pp. 1–4.
[12] M. Pahar, M. Klopper, B. Reeve, R. Warren, G. Theron, and T. Niesler, "Automatic cough classification for tuberculosis screening in a real-world environment," *Physiol. Meas.*, vol. 42, no. 10, Oct. 2021, Art. no. 105014.
[13] 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," 2023, *arXiv:2307.04842*.
[14] M. Pahar, G. Theron, and T. Niesler, "Automatic tuberculosis detection in cough patterns using NLP-style cough embeddings," in *Proc. Int. Conf. Eng. Emerg. Technol. (ICEET)*, Oct. 2022, pp. 1–6.
[15] G. Frost, G. Theron, and T. Niesler, "TB or not TB? Acoustic cough analysis for tuberculosis classification," 2022, *arXiv:2209.00934*.
[16] P. PrasannaKumari and B. P. Rao, "Abnormality detection of TB using instance learned classifier on lung CT images," *Int. J. Recent Technol. Eng.*, vol. 8, pp. 979–982, 2019.
[17] D. Jaganath et al., "Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM challenge," *medRxiv*, pp. 1–30, May 2024. [Online]. Available: https://www.medrxiv.org/content/10.1101/2024.05.13.24306584v1
[18] M. Sharma, V. Nduba, L. N. Njagi, W. Murithi, Z. Mwongera, T. R. Hawn, S. N. Patel, and D. J. Horne, "TBscreen: A passive cough classifier for tuberculosis screening with a controlled dataset," *Sci. Adv.*, vol. 10, no. 1, pp. 1–14, Jan. 2024.
[19] A. Fawaz, M. B. Ali, M. Adan, M. Mujtaba, and A. Wali, "A deep learning framework for efficient high-fidelity speech synthesis: Styletts," *iKSP J. Comput. Sci. Eng.*, vol. 1, no. 1, pp. 1–10, 2021.
[20] Cadence PCB Solutions. (2020). *What is Signal to Noise Ratio and How to Calculate It?*. Accessed: Oct. 10, 2024. [Online]. Available: https://resources.pcb.cadence.com/blog/2020-what-is-signal-to-noise-ratio-and-how-to-calculate-it
[21] B. McFee, C, Raffel, D. Liang, D. P. W. Ellis, M. McVica, E. Battenberg, and O. Nieto. (2023). *Librosa: Audio and Music Signal Analysis in Python*. Accessed: Oct. 10, 2023. [Online]. Available: https://librosa.org/doc/main/feature.html
[22] A. Wali, M. Ahmad, A. Naseer, M. Tamoor, and S. A. M. Gilani, "StynMedGAN: Medical images augmentation using a new GAN model for improved diagnosis of diseases," *J. Intell. Fuzzy Syst.*, vol. 44, no. 6, pp. 10027–10044, Jun. 2023.
[23] A. Shahzad and A. Wali, "Computerization of off-topic essay detection: A possibility?" *Edu. Inf. Technol.*, vol. 27, no. 4, pp. 5737–5747, May 2022.
[24] S. Serrano, L. Patane, and M. Scarpa, "Obstructive sleep apnea identification based on VGGish networks," in *Proc. ECMS*, Jun. 2023, pp. 556–561.
[25] Y. Song, X. Sun, L. Ding, J. Peng, L. Song, and X. Zhang, "AHI estimation of OSAHS patients based on snoring classification and fusion model," *Amer. J. Otolaryngol.*, vol. 44, no. 5, Sep. 2023, Art. no. 103964.
[26] A. Wali and M. Saeed, "Biologically inspired cellular automata learning and prediction model for handwritten pattern recognition," *Biologically Inspired Cognit. Archit.*, vol. 24, pp. 77–86, Apr. 2018.
[27] R. Imran, N. Hassan, R. Tariq, L. Amjad, and A. Wali, "Intracranial brain haemorrhage segmentation and classification," *iKSP J. Comput. Sci. Eng.*, vol. 1, no. 2, pp. 52–56, 2021.
[28] R. Shahid, A. Wali, and M. Bashir, "Next word prediction for Urdu language using deep learning models," *Comput. Speech Lang.*, vol. 87, Aug. 2024, Art. no. 101635.
[29] S. Serrano, L. Patanè, O. Serghini, and M. Scarpa, "Detection and classification of obstructive sleep apnea using audio spectrogram analysis," *Electronics*, vol. 13, no. 13, p. 2567, Jun. 2024.
[30] X. Sun, L. Ding, Y. Song, J. Peng, L. Song, and X. Zhang, "Automatic identifying OSAHS patients and simple snorers based on Gaussian mixture models," *Physiol. Meas.*, vol. 44, no. 4, Apr. 2023, Art. no. 045003.
[31] H. M. Hamza and A. Wali, "Pakistan sign language recognition: Leveraging deep learning models with limited dataset," *Mach. Vis. Appl.*, vol. 34, no. 5, p. 71, Sep. 2023.
[32] A. Wali, R. Shariq, S. Shoaib, S. Amir, and A. A. Farhan, "Recent progress in sign language recognition: A review," *Mach. Vis. Appl.*, vol. 34, no. 6, p. 127, Nov. 2023.
[33] Z. Jiang, W. Zaheer, A. Wali, and S. A. M. Gilani, "Visual sentiment analysis using data-augmented deep transfer learning techniques," *Multimedia Tools Appl.*, vol. 83, no. 6, pp. 17233–17249, Jul. 2023.

HAROON MAHMOOD received the master's and Ph.D. degrees in computer engineering from Politecnico di Torino, Italy. He is currently an Assistant Professor with Al Ain University, Al Ain, United Arab Emirates. Before joining this role, he was an Assistant Professor with the National University of Computer and Emerging Sciences, Lahore for nine years. His specialization is information security, especially the design of reliable and secure the IoT systems.

MANAL IFTIKHAR received the M.S. degree from the Department of Data Science, FAST-NUCES, Lahore Campus. She is currently a Freelancer. Her area of work and research includes machine learning and speech processing.

AAMIR WALI received the Ph.D. degree in computer science from the Department of Computer Science, National University of Computer and Emerging Sciences. He has been teaching with the Department of Computer Science, National University of Computer and Emerging Sciences, since 2004. His research interests include font development, writing systems, machine learning, image processing, human–computer interaction, and virtual/augmented reality.

ARSHAD ALI received the M.Sc. degree in computer science from Punjab University, Lahore, in 2003, the master's degree in information technology from the University of Avignon, France, in 2009, and the joint Ph.D. degree in telecommunication and computer science from the Institute of Telecom SudParis and UPMC (Paris VI), in 2012. He is currently an Associate Professor and the Head of the Cyber Security Department, National University of Computer Science and Emerging Sciences, Lahore Campus, Pakistan. He is a Postdoctoral Researcher with Orange Laboratories, Paris. His research interests include mobile ad-hoc networks, AI with cyber security, NLP, and AI in healthcare and agriculture.



MARYAM GULZAR received the B.S. degree in computer science from the CS Department, University of Lahore (UOL), Pakistan, in 2016, and the M.S. (SPM) degree from the School of Computing, National University of Computer and Emerging Sciences (NUCES), Lahore Campus, Pakistan, in 2018. She is currently pursuing the Ph.D. degree with the Department of Software Engineering, LUT University, Finland. She is a Junior Researcher. She was a Lecturer with the Department of Software Engineering, University of Lahore, Pakistan. Her research interests include digital transmission, information systems, and machine learning. She won a research grant from the Higher Education Commission of Pakistan for four years to complete her Ph.D. studies.