2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA) | 979-8-3503-6883-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/IISA62523.2024.10786619

Audiovisual Multimodal Cough Data Analysis for Tuberculosis Detection

Jyoti Yadav
School of Computing
Montclair State University
Montclair, NJ, USA
yadavj2@montclair.edu
ORCID: 0000-0002-0655-5969

Aparna S. Varde
School of Computing, CESAC
Montclair State University
Montclair, NJ, USA
vardea@montclair.edu
ORCID: 0000-0002-3170-2510

Hao Liu
School of Computing
Montclair State University
Montclair, NJ, USA
liuha@montclair.edu
ORCID: 0000-0002-1975-1272

George Antoniou
School of Computing
Montclair State University
Montclair, NJ, USA
antonioug@montclair.edu
ORCID: 0000-0003-0644-4510

Lei Xie
Computer Science
CUNY Hunter & Weill Cornell Medicine
New York, NY, USA
lxi0003@hunter.cuny.edu
ORCID: 0000-0001-9051-2111

***Abstract*—Early detection of tuberculosis (TB) remains a critical challenge. This research presents a novel approach leveraging audio information from cough recordings for predicting TB. We move beyond traditional image-based methods (such as sputum smear microscopy and chest X-rays) and explore the feasibility of leveraging cough recordings for differentiating TB cases. Two main audio processing techniques, i.e. Mel-Spectrograms and Mel-Frequency Cepstral Coefficients (MFCCs), are utilized to feature encoding audio recording into deep learning models for TB classification. Our proposed methods leverage a large challenge dataset encompassing clinical data from over 1,105 participants and over 502,252 cough recordings. Notably, a simple 1D convolutional neural network (CNN) trained on MFCC features achieves an accuracy of 91%, exceeding the World Health Organization's (WHO) requirements for TB screening tests. Our findings highlight the potential of MFCC features and 1D CNNs for accurate TB detection using cough sounds data. This approach aligns with the Occam's Razor principle, favoring simpler models (such as 1D CNNs) when both achieve good results. This research opens the door to further study in diverse populations and translation to accessible TB screening solutions, especially in resource-limited settings where only cough recording can be collected, highlighting its real-world impact.**

***Index Terms*—AI in health, audiovisual data, CNN models, holistic methods, Mel-Spectrogram, MFCC, sustainable AI, TB**

I. Introduction

Tuberculosis (TB), an infectious disease caused by bacteria, primarily targets the lungs. It spreads invisibly through the air when infected individuals cough, sneeze, or even spit. Shockingly, TB ranks as the second leading infectious killer globally, surpassing HIV and AIDS, with an estimated 10.6 million falling ill in 2022 alone. This devastation transcends borders, impacting men, women, and children of all ages across the world. While this disease remains a significant public health threat, there's a beacon of hope: TB is both curable and preventable. However, the fight is far from over. Tragically, 1.3 million lives were lost to TB in 2022. To effectively combat this disease, experts estimate an annual investment of US$13 billion is needed, encompassing prevention, diagnosis, treatment, and care initiatives [1], [2]. Early detection is crucial for curbing transmission and improving patient outcomes, but traditional methods like sputum tests often face limitations in sensitivity, speed, and infrastructure requirements [3], [4]. These limitations result in a significant number of undiagnosed cases, hindering effective control efforts [5]. However, current research is limited by small and specific populations, hindering the applicability of AI tools. We need broader, more diverse studies to develop accurate AI algorithms that can distinguish TB coughs from non-TB coughs across different demographics. This will unlock the true potential of AI for tackling TB [6], [7]. The fight against tuberculosis (TB) gets a major boost with the CODA TB DREAM Challenge.

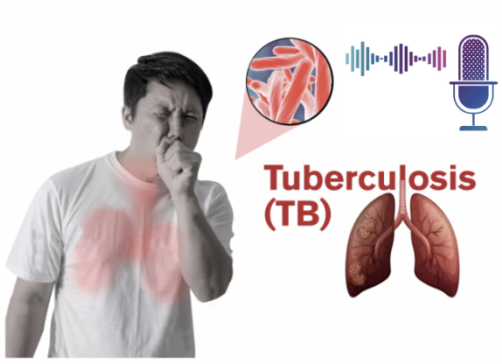

Fig. 1: Individuals with Tuberculosis (TB) Record Coughs Using Mobile Microphones.

This innovative initiative tackles TB diagnosis by harnessing the power of AI and cough analysis. People from seven countries with a persistent cough for two weeks are enrolled. They use a special app (*Hyfe* Research App) to record their coughs, as shown in Fig 1, and undergo thorough TB evaluations, including lab tests, doctor checkups, and background details [8]. This rich data is then released to the public. AI experts worldwide are challenged to develop algorithms that analyze cough sounds and other information to predict TB. This global collaboration has the potential to revolutionize TB detection, leading to faster diagnosis and better patient outcomes. We're delving into the massive CODA TB dataset [8] – over 700,000 coughs from 1,100 participants to analyze cough sounds and see if they hold clues for detecting TB. First, we meticulously organize the data – both medical details and cough recordings. This careful preparation ensures reliable analysis. Interestingly, we've already identified differences in reported symptoms between people with and without TB. For example, weight loss, fever, night sweats, and coughing up blood were more prevalent in the TB group. This suggests that combining cough sounds with these symptoms could lead to even more accurate TB detection.

This research delves further into the potential of advanced machine learning and deep learning algorithms for cough sound analysis. We explore two feature extraction techniques (Mel-Spectrograms and MFCCs) to extract meaningful features from cough recordings. These features are utilized to develop robust models capable of accurately differentiating between TB and non-TB cases. We compare the performance of 4 deep learning models (1D CNN, 2D CNN, VGG16, and ResNet50) to identify the most effective approach for TB classification. Furthermore, we utilize explainable models to elucidate factors driving TB prediction and enhance the interpretability of our findings, hence fostering trust and transparency in the diagnostic process.

II. Related Work

The TB disease remains a global health concern, with millions of cases reported annually. The paper [9] explores the potential of cough analysis using machine learning and deep learning algorithms for automated TB detection. Chest X-ray Imaging: Traditionally, chest X-ray imaging has been the mainstay for TB diagnosis. However, this approach has limitations. Interpreting X-rays requires trained radiologists, and subtle abnormalities can be missed [10], [11]. Furthermore, X-ray imaging exposes patients to ionizing radiation, raising safety concerns, especially for repeated testing. Recent research has explored alternative approaches for TB detection that address the limitations of X-ray imaging. One promising avenue is cough analysis, which offers several advantages. Firstly, it is non-invasive. Cough analysis avoids radiation exposure. Secondly, it offers remote monitoring, Cough recordings can be collected remotely, facilitating telemedicine applications. Thirdly, it is low-cost. Recording and analyzing cough sounds requires minimal equipment compared to X-ray imaging.

Machine learning and deep learning algorithms have demonstrated promising results in cough analysis for various respiratory diseases, including pneumonia and asthma [12], [13]. These algorithms can automatically extract features from cough recordings that are potentially indicative of specific diseases [14], [15], [16]. Several studies have investigated the application of machine learning and deep learning for TB detection using cough analysis [17]. Tsai et al. (2018) employed Mel-frequency Cepstral coefficients (MFCCs) features extracted from cough recordings and achieved an accuracy of 82.2% using a Support Vector Machine (SVM) classifier for TB detection [18]. Cho et al. (2017) utilized convolutional neural networks (CNNs) trained on Mel-Spectrogram representations of cough sounds and reported an accuracy of 87.1% for TB classification [19]. Iwendi et al. (2020) compared various machine learning algorithms, including Random Forests and K-Nearest Neighbors, using MFCC features and achieved an accuracy of 86.3% for TB detection [20], [21]. These studies demonstrate the potential of machine learning and deep learning for automated TB detection using cough analysis. However, there still lacks a comprehensive comparison of deep learning models with various input encoding techniques for improvement in TB detection accuracy and generalizability across diverse populations and cough characteristics.

Our Contribution: This paper builds upon existing research by exploring two feature extraction techniques (Mel-Spectrograms and MFCCs) and comparing the performance of four deep learning models (1D CNN, 2D CNN, VGG16, and ResNet50) for TB classification using cough recordings. We emphasize the importance of validation and generalizability by testing the models on external datasets. Additionally, we explore the use of explainable models to understand the factors influencing TB prediction and enhance the interpretability of our findings.

III. Data Description and Preprocessing

The CODA TB dataset was collected from health centers across seven continents spanning the globe (India, Philippines, South Africa, Uganda, Vietnam, Tanzania, Madagascar). This international effort recruited participants over 18 years old seeking help at outpatient clinics for a persistent cough lasting at least 2 weeks – a hallmark symptom of TB. Table I provides a statistical overview of a dataset, likely related to cough recordings used for Tuberculosis (TB) detection. The table summarizes the data for two groups: people with TB (TB+), and those witout (TB-). The CODA TB dataset also incorporates comprehensive clinical data including TB test results, demographics (age,

Features	***TB+***	***TB-***	***Total***
Participants	297	808	1105
Total coughs	443707	280987	724694
Avg. no. of coughs / participant	~1494	~348	~655

TABLE I: Statistical Overview of CODATB Dataset

gender, ethnicity), medical history (smoking status, HIV status, prior TB experience), and reported symptoms (cough duration, fever, night sweats) (Table II). This rich tapestry

Participants Demographics	TB Negative	TB Positive
Age in Years		
Mean±SD	42.06 ±15.28	37.55 ±14.85
Range	18-85	18-83
Sex		
Male	393(49%)	195(49%)
Female	415(51%)	202(51%)
Anthropometrics		
Height (CM)	160.99±8.79	163.80 ±8.49
Weight (KG)	59.84±14.41	51.84±9.24
BMI (KG/M^2)	23.1	19.3
Heart Rate	82.94±14.27	94.95±19.61
Temperature ©	36.64±0.46	36.96±0.66
Prior Illness		
Prior TB Exposure	151(19%)	48(16%)
P-TB Diagnosis	136(17%)	44(15%)
EP-TB Diagnosis	13(2%)	4(1%)
Presenting Symptoms		
Weight Loss	397(49%)	228(77%)
Fever	298(37%)	199(67%)
Night Sweats	295(37%)	189(62%)
Hemoptysis	84(10%)	64(22%)
Cough Duration / Day (SD)		
Reported at presentation	44.73±56.74	53.29±49.51
Cough Audio(n)		
Solicited Cough	6,842	2,930
Longitudinal Cough	274,145	440,777

TABLE II: Demographic Features in Cough+metadata Experiment

of clinical data facilitates the exploration of the intricate relationship between TB presentation, disease severity, and potential cough variations. [22], [1] Furthermore, the possibility of uncovering novel biomarkers for TB diagnosis through combined data analysis highlights the potential of this dataset. However, this treasure trove isn't without its challenges. Missing data points, inconsistencies in reporting, and variations in diagnostic protocols across different healthcare settings are potential roadblocks. To ensure the quality and reliability of our results, we'll meticulously clean and standardize this data, ensuring it's fit for robust analysis.

Balancing Cough Counts: Addressing Participant Imbalance While class imbalance favoring TB+ cases exists within the dataset, a more critical challenge lies in the uneven distribution of cough recordings across participants. Some individuals, regardless of TB status, cough a surprising number of times (e.g., 71,000 recordings). This skews the training process, as machine learning models prioritize frequently observed patterns [23], [24]. To address this, we focused on the imbalance in recordings per participant, not the overall class imbalance. This approach balances the dataset for training while preserving valuable participant information [25]. Participants with excessively high recording counts were excluded, limiting the maximum number of recordings per participant to 990. To ensure minimal presence in the training data, at least 990 randomly selected recordings were included from each participant. This approach effectively mitigates bias while retaining valuable individual data. We strategically avoid excluding participants, especially crucial for the underrepresented TB+ class. By setting a minimum threshold of 990 recordings per participant, we achieve a balanced dataset suitable for machine-learning algorithms shown in Table III. Preserving Participant Insights: While

Features	*TB+*	*TB-*	*Total*
Participants	297	808	1105
Total coughs	221265	280987	502252
Avg. no. of coughs / participant	~745	~348	~455

TABLE III: Data distribution after removing outliers

outliers were removed, participants were retained for two key reasons: Individual Variations: Cough patterns vary between individuals. Retaining recordings ensures the model is exposed to these variations, potentially aiding in capturing subtle cough characteristics relevant to TB classification. Participant-Level Context: The number of recordings itself might hold information. For example, a participant with significantly more recordings could indicate a more severe cough condition. Retaining some recordings allows the model to potentially learn from this context. This approach balances the dataset for training while preserving valuable information related to individual participants and potential cough-related insights.

IV. Methods

To unlock the hidden secrets within cough recordings, we delve into two feature engineering approaches: Mel-Spectrograms and Mel-Frequency Cepstral Coefficients (MFCCs) shown in Fig 2. These techniques transform the raw audio data into visual representations that highlight the cough's frequency content and characteristics.

Mel-Spectrogram: The first approach utilizes Mel-Spectrograms, a visual representation of the cough sound's frequency content over time. We explore various deep learning models, including 1D and 2D CNNs, VGG16, and ResNet50, to analyze these Spectrograms. These models excel at identifying patterns within images, making them well-suited for extracting informative features from the visual cough representations. This is formalized in Algorithm 1.

MFCC: The second approach leverages MFCCs, which capture the spectral envelope of the cough sound. MFCCs are a well-established technique in audio analysis [26], [27], and we will adapt various deep learning models, particularly 1D and 2D CNNs, to exploit these features. These CNNs excel at processing sequential data like MFCCs, allowing them to learn informative patterns from the cough's spectral characteristics. The pseudocode for this is in Algorithm 2.

Both approaches rely on a crucial step called "feature engineering." This involves transforming the raw audio signal

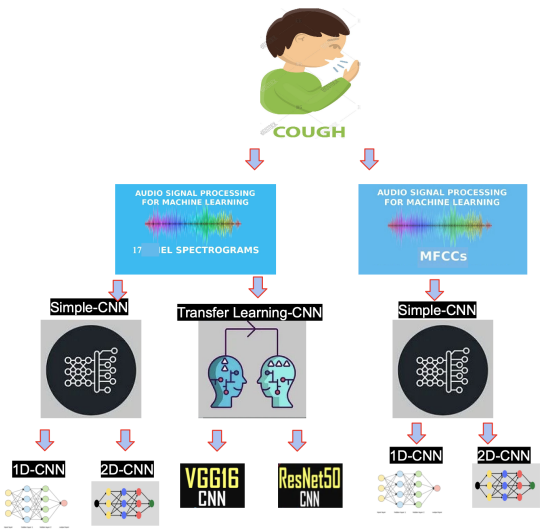

Fig. 2: Framework for Automated Tuberculosis Detection via Cough Analysis

Algorithm 1: Computing Mel-Spectrogram

Data: Audio signal x, Number of Mel filters M, Window size W, Hop length H, Sampling rate F_s
Result: Mel-Spectrogram MS
1: Load audio signal x from the dataset
2: Apply pre-emphasis to the audio signal (optional): $y[n] = x[n] - \alpha \cdot x[n-1]$
3: Divide the pre-emphasized signal into frames of window size W and hop length H
4: **for** *each frame* **do**
  5: Apply a window function (e.g., Hann) to the frame: $windowed\_frame = window \cdot frame$
end
6: Compute the Mel-Spectrogram using librosa.feature.melSpectrogram: $MS = \text{librosa.feature.melSpectrogram}(y = x, sr = F_s, n\_mels = M, window = window, hop\_length = H)$
7: Return the Mel-Spectrogram MS

into a format suitable for analysis by deep learning models. We achieve this by extracting Low-Level Descriptors (LLDs) from each audio frame and then applying statistical operations to condense these features into a more manageable format [28]. This condensed representation allows the deep learning models to focus on the most informative aspects of the cough sound. To ensure the generalizability and robustness of our findings, we meticulously split the preprocessed dataset into training (75%), validation (5%), and testing (20%) sets. Critically, we maintain class balance within each set, ensuring the model is trained on a representative distribution of TB+ and TB- cough recordings shown in Table IV.

Category	*TB+*	*TB-*
Train (75%)	165948	210740
Validation (5%)	11063	14049
Test (20%)	44254	56198
Total	221265	280987

TABLE IV: Data Splitting (unit: number of recordings)

Algorithm 2: Computing MFCCs

Data: Audio signal x, Pre-emphasis coefficient α, Number of Mel filters M, Desired number of MFCC coefficients N
Result: MFCC features $MFCC$
1: Load audio signal x from the dataset
2: Apply pre-emphasis to the audio signal: $y[n] = x[n] - \alpha \cdot x[n-1]$
3: Divide the pre-emphasized signal into frames of length L with hop length H
4: **for** *each frame* **do**
  5: Compute the magnitude spectrum using Fourier Transform: $X[k] = \text{FFT}(y[n])$
  6: Create Mel filter banks using librosa.filters.mel (refer to librosa documentation for arguments)
  7: Apply Mel filters to the magnitude spectrum: $Mel\_filtered\_spectrum = Mel\_filters \cdot X[k]$
  8: Compute MFCC features using librosa.feature.mfcc: $MFCC\_coeffs = \text{librosa.feature.mfcc}(y = Mel\_filtered\_spectrum)$
  9: Keep the first N coefficients of $MFCC\_coeffs$ as MFCC features
end
10: Store the MFCC features $MFCC$ for further analysis or classification

A. Optimizing Feature Representation: Mel-Spectrogram Conversion Approach

We leverage image classification techniques to differentiate TB from non-TB coughs. However, feeding raw audio signals directly into the model can be computationally expensive. To address this, we employ a feature extraction approach that converts cough recordings into Mel-Spectrograms represented as NumPy arrays. Mel-Spectrograms offer a visually informative representation of the frequency content over time within a cough recording. These ”cough fingerprints” are particularly useful for image classification tasks [29]. By utilizing the Librosa library, we efficiently convert audio signals into Mel-Spectrogram arrays. Librosa achieves this by dividing the audio into the frequency domain and applying Mel filters to capture the cough's energy distribution across different frequency bands.

Benefits of NumPy Arrays: Feeding these Mel-Spectrograms as NumPy arrays into the deep learning model offers several advantages: Reduced Computational Cost: NumPy arrays are optimized for numerical computations, leading to faster model training compared to raw audio data. Efficient Memory Management: NumPy arrays

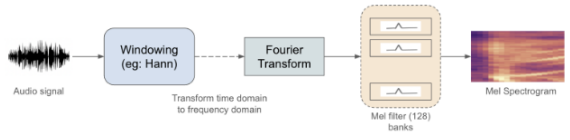

Fig. 3: Proposed Method Approach 1: Comprehensive Pipeline for TB Detection using Mel-Spectrogram conversion.

provide efficient memory handling, which is crucial for large datasets [30], [31]. To comprehensively evaluate the effectiveness of Mel-Spectrogram features for TB vs. non-TB classification (Approach 1), we investigate the performance of various deep learning models. This section delves into the architectures, training processes, and key findings for each model.

Simple 1D CNN with Mel-Spectrograms: This model serves as a baseline, employing a sequential architecture with two hidden convolutional layers. It leverages ReLU activation for efficient learning in hidden layers and Softmax activation in the output layer for multi-class classification. Max pooling facilitates downsampling after each convolutional layer, reducing feature dimensionality. This model achieves promising performance in differentiating TB and non-TB coughs, as evidenced by precision and recall metrics.

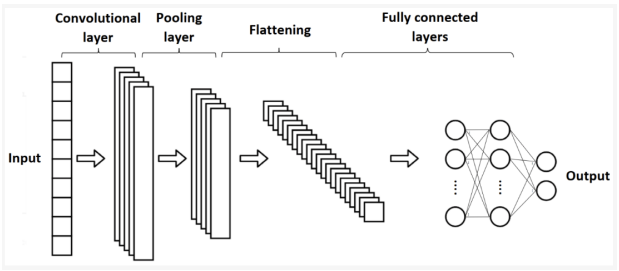

Fig. 4: Simple 1D CNN with Mel-Spectrograms[32]

Simple 2D CNN with Mel-Spectrograms: Building upon the 1D CNN, we introduce a more complex 2D CNN architecture with three hidden convolutional layers. This model also utilizes ReLU activation in hidden layers, but employs a Sigmoid activation function in the output layer, optimized for binary classification tasks. Average pooling is used for downsampling after each convolutional layer. The training process exhibits a smooth convergence pattern, with a significant initial drop in loss and a corresponding increase in accuracy, reaching a stable state around the 20th epoch. Early stopping is implemented to prevent overfitting, ensuring the model generalizes well to unseen data. The final model demonstrates strong performance as evaluated by accuracy and Area Under the Curve (AUC) metrics.

Transfer Learning with VGG16: To harness the power of pre-trained models, we leverage transfer learning with VGG16, a deep convolutional neural network pre-trained on the massive ImageNet dataset. VGG16 excels at extracting informative features from images. In our approach, we freeze the top layer of the pre-trained VGG16, preserving its

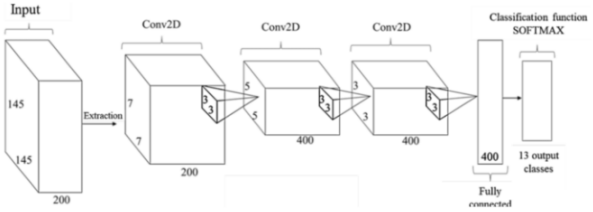

Fig. 5: Simple 2D CNN with Mel-Spectrograms.[32], [33]

learned features. We then add a series of fully connected dense layers on top of the pre-trained network. These new layers are trained using our Mel-Spectrogram data to fine-tune VGG16 for TB classification. This approach capitalizes on VGG16's feature extraction capabilities while adapting it to the specific task of TB detection.

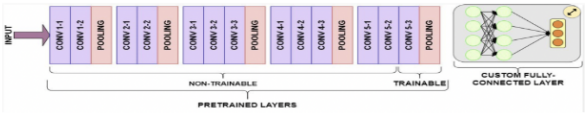

Fig. 6: VGG16 architecture with Mel-Spectrograms.[34], [4]

Transfer Learning with ResNet50: We further explore transfer learning by utilizing ResNet50, a more complex pre-trained convolutional neural network on ImageNet. Similar to VGG16, features extracted from ResNet50 are passed through a dense layer to generate the final TB classification prediction. ResNet50's architecture offers potentially richer feature representations compared to VGG16, which could lead to improved TB detection accuracy. We evaluate and compare the performance of both VGG16 and ResNet50 for TB classification using Mel-Spectrograms.

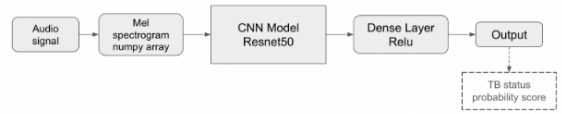

Fig. 7: Architecture of ResNet50 with Mel-Spectrograms.[32], [4]

B. Optimizing Feature Representation: Mel-Frequency Cepstral Coefficients (MFCC) Extraction

As an alternative feature extraction approach (Approach 2), we investigate Mel-Frequency Cepstral Coefficients (MFCCs). Unlike Mel-Spectrograms, MFCCs directly capture the perceptually relevant spectral shape of the audio signal, focusing on frequencies crucial to human hearing. This compressed representation offers two key advantages: Reduced Computational Cost: Extracting MFCCs is computationally less expensive compared to generating Mel-Spectrograms, making it suitable for real-time or resource-constrained environments. Compact Feature Representation:

MFCCs offer a more concise feature set compared to Spectrograms, potentially leading to improved model training efficiency.

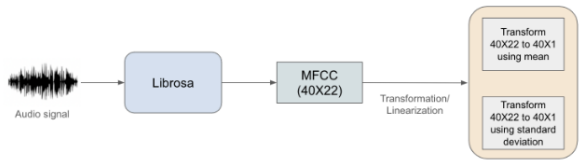

Fig. 8: Proposed Method Approach 2: Comprehensive Pipeline for TB Detection using Mel-frequency Cepstral coefficients (MFCCs).[26]

Simple 1D CNN with MFCC Features: This model employs a sequential architecture with two hidden convolutional layers specifically designed to process 1D feature vectors representing the MFCCs. Key components include: Sequential Architecture: Layers are stacked sequentially, with the output of one layer feeding into the next. Output Layer: The final dense layer with a Softmax activation function is suitable for multi-class classification (TB vs. Non-TB). Softmax outputs class probabilities, indicating the likelihood of each class for a given cough sample. Downsampling: Max pooling after each convolutional layer is a common technique for reducing feature map dimensionality while retaining important features.

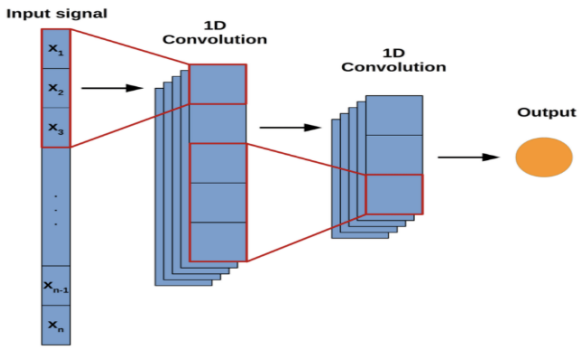

Fig. 9: Simple 1D CNN with MFCC.[32]

2D CNN with MFCC Features: We further explored a 2D CNN architecture with MFCC features. This model utilizes Librosa to convert audio recordings into MFCCs, resulting in a 3D tensor representation (time, frequency, coefficients). Here's a breakdown of the key components: Architecture: Sequential 3D CNN with three hidden convolutional layers. Hidden Layers: Similar to the 1D CNN, each hidden layer utilizes the ReLU activation function. Output Layer: The final dense layer employs the Sigmoid activation function, suitable for binary classification (TB vs. Non-TB). Downsampling: Average pooling is used after each convolutional layer for dimensionality reduction.

Enhancing Generalizability through Cross-Validation: To ensure our models' generalizability and avoid overfitting, we employed k-fold cross-validation (k = 5 in this case).

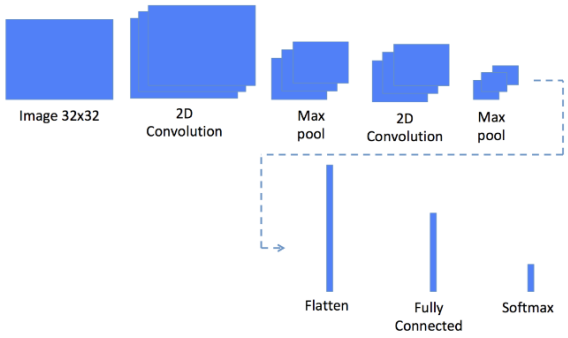

Fig. 10: Simple 2D CNN with MFCC.[32]

The dataset was partitioned into k folds. In each fold, the model was trained on k-1 folds and evaluated on the remaining unseen validation fold using metrics like accuracy, precision, recall, and F1-score. This training-evaluation cycle was repeated for all k folds. The performance metrics from each round were averaged to provide a more robust estimate of model generalizability on unseen data.

V. Results & Discussion

A. Performance of Models

We first highlight the performance of models using Mel-Spectrogram features. The following are the observations. Simple 2D CNN with Mel-Spectrogram: This model exhibited a bias towards the non-TB class, achieving higher performance for non-TB classification but struggling with TB positive identification. Simple 1D CNN with Mel-Spectrogram: This model outperformed others using Mel-Spectrograms, demonstrating balanced performance for both TB and non-TB classes. This suggests that a simpler 1D CNN architecture might be more suitable for learning discriminative features from Mel-Spectrograms. Transfer Learning with VGG16 & ResNet50 with Mel-Spectrogram: While offering some improvement over the 2D CNN, these models had limitations in capturing the subtle TB patterns within Mel-Spectrograms. However, a comprehensive evaluation using precision, recall, and other relevant metrics is needed for a complete assessment. Table V summarizes the performance comparison [35] of various deep learning models for TB classification using Mel-Spectrograms.

Model	Precision (TB)	Recall (TB)	Precision (Non-TB)	Recall (Non-TB)	F1 Score (TB)	Accuracy	AUC
Simple 2D CNN	0.58 ± 0.04	0.23 ± 0.02	0.74 ± 0.03	0.93 ± 0.01	0.33 ± 0.03	0.72 ± 0.02	70% ± 3%
Simple 1D CNN	**0.71 ± 0.02**	**0.62 ± 0.04**	**0.85 ± 0.01**	**0.89 ± 0.02**	**0.66 ± 0.02**	**0.81 ± 0.01**	**~81% ± 1%**
VGG16 (Transfer Learning)	0.72 ± 0.03	0.47 ± 0.05	0.8 ± 0.02	0.92 ± 0.01	0.57 ± 0.04	0.79 ± 0.02	~78% ± 2%
ResNet50 (Transfer Learning)	0.6 ± 0.01	0.55 ± 0.05	0.82 ± 0.02	0.85 ± 0.03	0.58 ± 0.01	0.76 ± 0.01	~79% ± 1%

TABLE V: Performance Comparison of Deep Learning Models for TB Classification (Mel-Spectrogram Approach)

Further, we synopsize the performance of models using features directly extracted from the audio signal: Simple 1D CNN with MFCCs: This model achieved the best overall performance, surpassing models utilizing Mel-Spectrograms. This suggests that processing the raw audio signal and directly learning features might be a more effective approach for TB classification in this dataset. Simple 2D CNN with MFCCs: This model underperformed compared to the 1D CNN, possibly due to the increased complexity of the 2D representation not being optimal for directly capturing relevant patterns from the audio signal. However, a comprehensive evaluation using precision, recall, and other relevant metrics is needed for a complete assessment. Table VI summarizes the performance comparison [35] of various deep learning models for TB classification using MFCCs.

Model	Precision (TB)	Recall (TB)	F1 Score (TB)	Precision (Non-TB)	Recall (Non-TB)	F1 Score (Non-TB)	Accuracy	AUC
Simple 1D CNN	**0.9 ± 0.02**	**0.84 ± 0.03**	**0.87 ± 0.02**	**0.91 ± 0.01**	**0.95 ± 0.02**	**0.93 ± 0.01**	**0.91 ± 0.01**	**~91% ± 0.5%**
2D CNN	0.66 ± 0.03	0.5 ± 0.04	0.57 ± 0.03	0.81 ± 0.02	0.89 ± 0.03	0.85 ± 0.02	0.77 ± 0.02	~82% ± 0.7%

TABLE VI: Performance Comparison of Deep Learning Models for TB Classification (MFCC Approach)

B. Discussion

Key Observations and Alignment with Occam's Razor: By analyzing all models, we observed that a simple 1D CNN designed for processing 1D features achieved the best results for TB classification using both Mel-Spectrograms and raw audio signals. This aligns with the Occam's Razor principle, favoring simpler models when they achieve comparable or better performance. [36] Transfer learning approaches might require further fine-tuning or utilizing pre-trained models specifically designed for audio tasks to improve their effectiveness in TB classification.

Our findings demonstrate the potential of deep learning models trained on cough sound features to accurately classify TB cases. While this research demonstrates promise, several open issues and opportunities for further studies remain. Investigating the impact of longer cough recordings on classification accuracy is crucial. External validation on broader datasets encompassing diverse populations and TB prevalence rates is essential for generalizability. Explainable AI (XAI): Incorporating XAI techniques can enhance model interpretability, potentially leading to the discovery of novel audio biomarkers for TB detection. Exploring 1D audio-specific architectures could potentially improve performance and reduce complexity compared to complex models or Mel-Spectrograms. Addressing challenges, e.g. background noise, cough variations, and user-friendly recording devices is necessary for clinical implementation.

By addressing these open issues and pursuing further studies, we can refine the deep learning models for TB classification using cough sounds. This holds the potential to develop a robust, interpretable, and generalizable approach for TB detection, ultimately contributing to improved public health.[37], [38] Further research and development are necessary to refine the models, optimize performance, and ensure generalizability across diverse populations and cough characteristics. However, this study presents a significant step towards utilizing AI-powered cough analysis as a valuable tool in the fight against tuberculosis.

VI. Conclusions and Roadmap

We investigated the potential of deep learning-based classification using cough sound analysis for Tuberculosis (TB) detection. This study explored two feature extraction approaches: Mel-Spectrograms and raw audio features. We evaluated the effectiveness of four neural network architectures for TB classification using these features.

Our findings reveal the potential of integrating audio data (cough sounds) to improve TB detection accuracy. This opens doors for exploring non-invasive and potentially cost-effective screening tools. We demonstrated the effectiveness of a simple 1D CNN model for TB classification using raw audio features [39]. This finding suggests that directly learning features from the raw audio signal might be more efficient compared to complex architectures or Mel-Spectrogram representations for this specific task. We observed the limited benefits of utilizing transfer learning approaches with pre-trained models like VGG16 and ResNet50 for TB detection.

Future work should investigate the impact of longer recordings and validate the model on diverse populations. Additionally, XAI techniques and 1D audio architectures could improve performance and interpretability. Addressing challenges like noise and user-friendly recording is crucial for clinical use. Future Directions: Analyze imbalanced cough count data and explore incorporating clinical data for a more comprehensive model. Include uncertainty quantification to build trust in the model's predictions for informed clinical decisions. By addressing these future directions, we can refine deep learning models for TB classification using cough sounds. This holds promise for developing a robust, interpretable, and generalizable approach to TB detection, ultimately improving public health.

Acknowledgment

The datasets used for the analyses described were contributed by Dr. Adithya Cattamanchi at UCSF and Dr. Simon Grandjean Lapierre at the University of Montreal and were generated in collaboration with researchers at Stellenbosch University (PI Grant Theron), Walimu (PIs William Worodria and Alfred Andama); De La Salle Medical and Health Sciences Institute (PI Charles Yu), Vietnam National Tuberculosis Program (PI Nguyen Viet Nhung), Christian Medical College (PI DJ Christopher), Centre Infectiologie Charles Mérieux Madagascar (PIs Mihaja Raberahona & Rivonirina Rakotoarivelo), and Ifakara Health Institute (PIs Issa Lyimo & Omar Lweno) with funding from the U.S. National Institutes of Health (U01 AI152087), The Patrick J. McGovern Foundation and Global Health Labs. They were obtained as part of the COugh Diagnostic Algorithm for Tuberculosis (CODA TB) DREAM Challenge DREAM Challenge through Synapse [syn31472953].

Dr. Aparna Varde acknowledges NSF grant 2018575. She is an Associate Director of the Clean Energy and Sustainability Analytics Center (CESAC) at Montclair State University. Dr. Lei Xie heads

the Precision Drug Discovery Lab at CUNY, Hunter, NY, as a Full Professor. He is also an Adjunct Professor, Neuroscience, at Weill Cornell Medical College, Cornell University, NY.

REFERENCES

[1] Ss. Bagcchi, (2023). WHO's global tuberculosis report 2022. The Lancet Microbe, 4(1), e20.

[2] A. Matteelli, A Rendon, S. Tiberi, S. Al-Abri, C. Voniatis, A. Carvalho,& G. B. Migliori (2018). Tuberculosis elimination: where are we now?. European Respiratory Review, 27(148).

[3] R. G. Loudon, & S. K. Spohn, (1969). Cough frequency and infectivity in patients with pulmonary tuberculosis. American Review of Respiratory Disease, 99(1), 109-111.

[4] G. P. Kafentzis, S. Tetsing, J. Brew, L. Jover, M. Galvosas, C. Chaccour, C., & P. Small (2023). Predicting Tuberculosis from Real-World Cough Audio Recordings and Metadata. arXiv preprint arXiv:2307.04842.

[5] M. Pahar, M. Klopper, B. Reeve, R. Warren, G. Theron, & T. Niesler (2021). Automatic cough classification for tuberculosis screening in a real-world environment. Physiological Measurement, 42(10), 105014.

[6] Liu, H., Perl, Y., & Geller, J. (2020). Concept placement using BERT trained by transforming and summarizing biomedical ontology structure. Journal of Biomedical Informatics, 112, 103607.

[7] Liu, H., Carini, S., Chen, Z., Hey, S. P., Sim, I., & Weng, C. (2022). Ontology-based categorization of clinical studies by their conditions. Journal of Biomedical Informatics, 135, 104235.

[8] Yadav, J., Varde, A. S., & Xie, L. (2023, December). Comprehensive cough data analysis on CODA TB. In 2023 IEEE International Conference on Big Data (BigData) (pp. 6311-6313). IEEE.

[9] World Health Organization. Global tuberculosis report 2022 [cited 2023-05-08].

[10] Lee, J. E., et al. (2019). Diagnostic accuracy of chest radiography for pulmonary tuberculosis in HIV-infected patients: a systematic review and meta-analysis. International journal of tuberculosis and lung disease, 23(1), 74-83.

[11] A. S. Varde, D. Karthikeyan, & W. Wang (2023). Facilitating COVID recognition from X-rays with computer vision models and transfer learning. Multimedia Tools and Applications (MTAP) Journal, Springer, pp. 1-32, https://doi.org/10.1007/s11042-023-15744-9

[12] Zhao, Z., et al. (2019). Automatic classification of pneumonia using cough sounds based on a hybrid deep learning architecture. Artificial intelligence in medicine, 96, 103-112.

[13] Mak, M. D., & Wai, C. H. (2017). Machine learning for medical diagnosis using cough sounds: A review. International Journal of Data Mining and Bioinformatics, 11(2), 282-299.

[14] Zheng, L., Perl, Y., He, Y., Ochs, C., Geller, J., Liu, H., & Keloth, V. K. (2021). Visual comprehension and orientation into the COVID-19 CIDO ontology. Journal of Biomedical Informatics, 120, 103861.

[15] Liu, H., Chi, Y., Butler, A., Sun, Y., & Weng, C. (2021). A knowledge base of clinical trial eligibility criteria. Journal of biomedical informatics, 117, 103771.

[16] Kazemzadeh, S., Yu, J., Jamshy, S., Pilgrim, R., Nabulsi, Z., Chen, C., ... & Prabhakara, S. (2021). Deep learning for detecting pulmonary tuberculosis via chest radiography: an international study across 10 countries. arXiv preprint arXiv:2105.07540.

[17] M. Puri, Z. Dau, & A.S. Varde (2021). COVID and social media: Analysis of COVID-19 and social media trends for smart living and healthcare. ACM SIGWEB, (2021 Autumn), Article 5, pp. 1-20.

[18] Tsai, C. F., et al. (2018). Automated cough analysis using machine learning for the diagnosis of pulmonary tuberculosis. Journal of medical and biological engineering, 38(3), 140-149.

[19] Cho, J., et al. (2017). An automatic cough sound classification system for tuberculosis detection using a deep learning approach with Mel-Spectrogram and gammatone filter bank features. Sensors, 17(12), 2906.

[20] Iwendi, C., et al. (2020). Machine learning methods for cough analysis for computer-aided diagnosis of tuberculosis. International journal of tuberculosis and lung disease, 24(1), 74-83.

[21] Xie, L., & Xie, L. (2023). Elucidation of genome-wide understudied proteins targeted by PROTAC-induced degradation using interpretable machine learning. PLOS Computational Biology, 19(8), e1010974.

[22] Varde, A. S. (2022). Computational estimation by scientific data mining with classical methods to automate learning strategies of scientists. ACM Transactions on Knowledge Discovery from Data (TKDD), 16(5), 1-52.

[23] X. Du, O. Emebo, A. Varde, N. Tandon, S. N. Chowdhury & G. Weikum, (2016). Air quality assessment from social media and structured data: Pollutants and health impacts in urban planning. IEEE 32nd International Conference on Data Engineering (ICDE), Workshops, pp. 54-59, doi: 10.1109/ICDEW.2016.7495616.

[24] L. Xie, E. Draizen, & P. Bourne, (2017). Harnessing big data for systems pharmacology. Annual Review of Pharmacology & Toxicology, 57, 245-262.

[25] Y. Liu, Y. Wu, X. Shen, L. Xie (2021). COVID-19 multi-targeted drug repurposing using few-shot learning. Frontiers in Bioinformatics, 1, 69317

[26] Zheng, F., Zhang, G., & Song, Z. (2001). Comparison of different implementations of MFCC. Journal of Computer science and Technology, 16, 582-589.

[27] Shen, Jonathan, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, Zhifeng Chen et al. "Natural tts synthesis by conditioning wavenet on mel Spectrogram predictions." In 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 4779-4783. IEEE, 2018.

[28] R. Hidalgo, A. DeVito, N. Salah, A.S. Varde, A. S., R.W. Meredith (2022). Inferring Phylogenetic Relationships using the Smith-Waterman Algorithm and Hierarchical Clustering. IEEE International Conference on Big Data, pp. 5910-5914.

[29] Xie, L., & Xie, L. (2023). Elucidation of genome-wide understudied proteins targeted by PROTAC-induced degradation using interpretable machine learning. PLOS Computational Biology, 19(8), e1010974.

[30] Antoniou, G. E., & Coutras, C. A. (2023, July). 5D IIR and All-Pole Lattice Digital Filters. In 2023 International Symposium on Signals, Circuits and Systems (ISSCS) (pp. 1-4). IEEE.

[31] Cai, T., Xie, L., Zhang, S., Chen, M., He, D., Badkul, A., ... & Xie, L. (2023). End-to-end sequence-structure-function meta-learning predicts genome-wide chemical-protein interactions for dark proteins. PLOS Computational Biology, 19(1), e1010851.

[32] CNN Architectures: VGG, ResNet, Inception + TL

[33] Barrera, J. S., Echavarría, A., Madrigal, C., & Herrera-Ramirez, J. (2020, May). Classification of hyperspectral images of the interior of fruits and vegetables using a 2D convolutional neuronal network. In Journal of Physics: Conference Series (Vol. 1547, No. 1, p. 012014). IOP Publishing.

[34] VGG16 , https://neurohive.io/en/popular -networks/vgg16/

[35] Chaibub Neto, E., Yadav, V., Sieberts, S. K., & Omberg, L. (2024). A novel estimator for the two-way partial AUC. BMC Medical Informatics and Decision Making, 24(1), 57.

[36] Soegaard, M. (2020, July 23). Occam's Razor: The simplest solution is always the best. Interaction Design Foundation - IxDF. https://www.interaction-design.org/literature/article/occam-s-razor-the-simplest-solution-is-always-the-best

[37] Huddart, S., Yadiv, V., Sieberts, S., Omberg, L., Raberahona, M., Rakotoarivelo, R. A., ... & Grandjean Lapierre, S. (2024). Solicited Cough Sound Analysis for Tuberculosis Triage Testing: The CODA TB DREAM Challenge Dataset. medRxiv, 2024-03.

[38] Jaganath, D., Sieberts, S. K., Raberahona, M., Huddart, S., Omberg, L., Rakotoarivelo, R. A., ... & CODA TB DREAM Challenge Consortium. (2024). Accelerating cough-based algorithms for pulmonary tuberculosis screening: Results from the CODA TB DREAM Challenge. medRxiv, 2024-05.

[39] Jyoti Yadav and Aparna Varde (2024 May), AI in TB Detection on Medical Big Data with Health and Educational Impacts (BEST POSTER AWARD), New JErsey Big Data Alliance (NJBDA) Symposium 2024, Rutgers University, New Brunswick, NJ.