Advancements in Machine Learning for Tuberculosis: A Systematic Analysis of Clinical Symptom Severity Datasets and Multimodal Fusion

Suresh S[1], Dhanalakshmi S[2]
[1]*Research Scholar, Department of Computer Science, Sri Krishna Arts and Science College*
[2]*Associate Professor, Department of Software Systems and AIML, Sri Krishna Arts and Science College*
doi.org/10.64643/IJIRTV13I1-205399-459

***Abstract*—In this systematic analysis, the findings concerning the development and utilization of datasets with symptom severity measures in clinical TB cases within the machine learning (ML) pipeline have been outlined. First, the aims encompass assessing the diversity and source of the datasets employed, benchmarking some ML approaches by using their symptom severity, as well as researching multimodal fusion strategies. Moreover, estimating labelling strategies and, ultimately, analysing the interpretability strategies adopted by the selected studies. In total, over 43 investigations conducted in numerous areas around the globe have been scrutinized, incorporating diverse modalities into the examination including clinical, acoustic, radiological, and genomic information. As per the results, fusing cough sound, clinical, and imaging data leads to improved diagnostic performance and more accurate predictions when employing advanced multimodal algorithms that produce AUCs above 0.9. Nevertheless, highly annotated datasets have significantly advanced ML algorithms' performance; however, patient adherence to the annotation guidelines and lack of consistency between each other is still a major challenge in the low resource settings. Techniques such as SHAP and feature importance assessment help build clinical trust but are, however, not extensively used across the literature. The information accumulated demonstrates that amalgamated longitudinal symptom intensity datasets possess significant promise for diagnosis and therapeutic oversight of tuberculosis utilizing machine learning. According to the review, data must be standardized, interpretable models must be developed to ensure scalable and clinically meaningful solutions in different healthcare settings.**

***Index Terms*—Tuberculosis; Machine Learning; Clinical Symptom Severity; Multimodal Data Fusion; Cough Sound Analysis; Explainable Artificial Intelligence (XAI).**

I. INTRODUCTION

Infectious diseases have the most significant burden of disease globally and among these is tuberculosis. There is an urgent demand for rapid and easy to use tests. As a result, development and use of clinical symptom severity datasets are an area of active ML research [1][2]. During the last decade, digital health technologies and artificial intelligence (AI) have been put to increasing use for tuberculosis (TB) screening & treatment monitoring through the collection and analysis of cough sound data and clinical metadata [3][4]. The conventional diagnostic tests aimed at early diagnosis like sputum microscopy or chest X-ray require advanced infrastructure and high resource utilization [5][6]. Newer alternatives are mobile technology and AI-driven cough sound analysis, which can offer scalable and non-invasive screening as well as suitable for low-resource settings.[2][4] The World Health Organization's (WHO) endorsement of the need for TB triage tests that meet defined sensitivity and specificity thresholds has lent momentum to further research activity in this area [2][7].

Although there have been advancements, the accurate collection and effective use of data related to the severity of clinical symptoms for TB diagnosis and prognosis continue to face major challenges. The existing datasets are often collected via heterogeneous methods, limited to a specific geographical region, and they overlook most of the relevant clinical features [10][11][12]. Although cough sound analysis is feasible for TB screening, over-reliance on patient self-reported symptoms and inconsistent use of a particular symptom severity scale have limited test

accuracy [8][13]. Many researchers are still evaluating ways to integrate the best combination of acoustic, clinical, and demographic features to optimize ML algorithm performance. Moreover, a more consistent nonexistence of large scale, standardised and clinically validated symptom severity datasets continue to act as a key barrier towards the development of AI models for reliable generalisation across different populations [16][17]. A lack of effective diagnostic tools provides inadequate access to timely TB detection, increased transmission and avoidable mortality on a broader level.

The foundation of this domain encompasses clinical symptom intensity alongside acoustic biomarkers and computational learning to facilitate early tuberculosis diagnosis and continuous assessment of therapeutic responsiveness [20][21]. Metrics of symptom intensity - cough frequency and severity correlated with systemic manifestations - can function as credible surrogates for disease load [22][13]. The application of acoustic cough biomarkers furnishes objective metrics that are non-invasive and bolster clinical symptomatology data [4][23]. Computational learning models capitalise on these diverse data streams through feature extraction and multimodal integration to forecast tuberculosis status and post-therapeutic outcomes [1][24]. This framework, which comprises integrated biomedical signal processing and clinical epidemiology, offers the scientific justification that underpins the connection between the datasets of symptom intensity and the diagnostic instrument that is propelled by artificial intelligence [25][26].

The aim of this comprehensive examination is to integrate contemporary data regarding the development and utilization of clinical symptom intensity repositories for tuberculosis within the framework of machine learning [9][11]. Through a meticulous examination of contemporary literature concerning investigations executed between the years 2008-2026, it delineates the deficiencies that promote the advancement of scalable, precise, and just TB screening mechanisms [7][27][28][14]. Investigations that encompassed clinical symptom intensity data and employed machine learning for TB diagnosis or prognosis were incorporated [29][30]. The results are structured thematically to make clear the trends found through data collection, the methods of integrating various modalities, and their relevance in clinical practice [31][32].

II. OBJECTIVES OF THE REVIEW

This analysis examines how datasets that encompass clinical symptom intensity of tuberculosis are constructed and utilised in machine learning applications. This assessment systematically probes into the growth of these data formats, their execution, and their merging within predictive strategies. The global predicament of tuberculosis remains acute, and there exists a substantial and untapped potential to enhance diagnosis, prognosis, and treatment follow-up through the utilisation of data on clinical symptom intensity with machine learning. By synthesising contemporary methodologies, dataset characteristics, and modelling techniques, this analysis delineates methodological deficiencies, standardisation prospects, and potential pathways for further advancements pertinent to the implementation of more effective and interpretable ML solutions in TB care.

The analysis objectives are specific:

- To evaluate existing evidence on the types and sources of clinical symptom severity datasets used in TB machine learning studies.
- To use data on symptom severity to evaluate existing machine learning models for TB diagnosis and treatment outcome prediction.
- To examine approaches for combining clinical, acoustic, radiological, and genomic data for assessing the severity of TB.
- The objective of this study is to analyse dataset labelling and annotation methodologies orienting around scoring symptom severity and longitudinal data collection in low-resource settings.
- Interpretability techniques can be useful to augment the transparency and clinical trust in the machine learning model outputs.

III. LITERATURE REVIEW

The main research question relating to key trends in the development and use of datasets for the severity of clinical symptoms for tuberculosis for machine learning is interpreted in this section and systematically decomposed into several detailed search statements. Breaking down the main research question into sub-questions allows for a better organized and informed literature search. This way, studies that address questions related to other dimensions of the topic will also be captured.

Transformed queries given below are based on original research question:

- A dataset of clinical symptom severity for tuberculosis: trends in development and application in machine learning.
- Multimodal datasets for TB symptom severity and prognosis include clinical, radiological, genomic and acoustic cough data; dataset development, labelling strategies, interpretability of ML models, and severity prediction and treatment outcomes.
- Datasets will consist of multimodal data including cough acoustics, clinical scores (TBscore, etc), imaging, and genomic data collected in low-resource settings (Tamilnadu, Mumbai) useful for severity prediction, prognosis, and screening. Data will also include labelling strategies, explainability of models, etc.
- Data on cough acoustics, continuous cough monitoring parameters, clinical scoring system, radiology, genomics for TB symptom severity and treatment response; so that suitable labelling strategies can be decided upon; can monitor longitudinal trajectories; model interpretability for screening, prognosis and treatment monitoring.

All transformed queries were run on a large academic database of over 270 million publications using the prescribed inclusion and exclusion criteria to select a well-defined collection of relevant research. This process gave an initial result of 78 research papers.

A. Overview of Research

This section characterizes research undertaken on important trends in the development and use of clinical symptom severity datasets for tuberculosis in machine learning. It focuses on how datasets have been constructed, used, and integrated into models. The literature reviewed provides ample scope for investigating CRD and cough, with studies undertaken employing cough analysis, clinical scoring, imaging and multimodal assessment. Studies are being conducted in high-burden TB regions in Africa, Asia, and South America. The synthesis focuses on dataset characteristics, the use of multimodal data, longitudinal data, labelling, model explanations.

Dataset Composition

- Over 30 studies have used different datasets. They include large multi-country cough sound collections, clinical symptom questionnaires, and multimodal clinical-radiological-genomic databases [10][26][24].
- Continuous cough monitoring during treatment can provide richer data for tracking symptoms, as several longitudinal data studies have stated [3][22].
- Many studies pointed out the limitations of relying on patient self-reporting of symptoms and suggested the implementation of scoring systems with acoustic biomarkers to augment data reliability [8][20].

Performance of the Model

- Close to 20 studies reported high-level diagnostic accuracy and found that most of the deep learning and ensemble techniques (e.g., ANN, CNN, and XGBoost) achieved above 85% accuracy with AUC value above 0.80 [9][34][1].
- Models depending on clinical symptom severity data as a standalone feature demonstrated good performance. However, performance was always improved when multimodal fusion was applied [16][35].
- Methods that focus on class imbalance, including cost-sensitive learning, enhance sensitivity and specificity in both verbal screening and clinical datasets [18][25].

Fusion of Multiple Modes

- About 15 studies performed fusion of cough audio with clinical, radiological or genomic data using late fusion, cross-attention modules, and graph neural networks [26][24][1].
- In drug-sensitivity prediction, fusion schemes consistently outperformed unimodal ones, with diverse data types being specifically beneficial [12][15][32].
- A promising advanced graph-based fusion frameworks typically demonstrate an ability to capture complex cross-modal dependencies that unimodal model cannot adequately represent [24][31].

Quality of Labelling and Annotation

- Several experiments use datasets that are well annotated with clinical characteristics, microbiological confirmation, and standardised

scoring systems for the severity of symptoms including a TBscore [20][13][16].
- Longitudinal tagging was used for monitoring continuous cough. However, adherence and technical issues affected overall completeness of data [3][40].
- The reviews observed that there were variation and inconsistency in the symptom self-reporting which indicates a need for annotation with more objectivity and standardization [8][7].

Interpretation of Model
- SHAP values, feature importance analyses, and uncertainty quantification were implemented in several studies to explain model predictions and improve clinical confidence [19][25][30].
- Certain models offered clinically validated interpretations with explanations suitable for use by frontline health [1][11].
- As interpretability becomes popular, complex graph-based architectures remain challenging for explainability [24][31].
- This view was reinforced by reviews underlining that in constrained resource settings, where front-line clinicians need to act on model output, interpretability becomes central [4].

This view was reinforced by reviews underlining that in constrained resource settings, where front-line clinicians need to act on model output, interpretability becomes central [4].

IV. ANALYSIS AND SYNTHESIS

The literature reviewed indicates significant improvement in the development and use of clinical symptom severity datasets for TB in machine learning applications. A primary theme is the use of multimodal data, particularly cough audio and clinical metadata, which has improved prediction and robustness of the model. Nonetheless, there are still issues with respect to dissimilarity in datasets, consistency of annotations, interpretability of models, etc. particularly in low-resource settings. While there many analyses have been reported with promising diagnosis performance, model generalizability issues, prospect of non-compliance with data collection protocols, and practicability of machine learning (ML) outputs in clinical care remain under-explored.

A. Dataset Characteristics and Quality

A number of sizeable, multi-country datasets, most particularly CODA TB, provide extensive cough sound recordings with rich clinical and demographic expertise, which enable strong model training and validation on diverse populations [10] [41] [16]. Certain investigations encompassed longitudinal data assemblages (for instance, the observation of chronic cough throughout treatment intervals). This will help us understand how symptom progression takes place over time [3][22]. Through de-identification and cleansing as used in TB spectrum, the usefulness of the data sets is increased without compromising patients' privacy [42][42]. However, despite all these strengths, many data sets have been adversely affected by differences in data collection resulting from such variables as recording devices, ambient noise conditions, and annotation techniques [36][43]. The inadequate system that makes it difficult to evaluate the severity of the symptoms poses challenges in making meaningful comparative studies [8]. Also, the challenges posed in longitudinal study observation come out as challenges towards achieving accurate and consistent data [3][40].

B. Multimodal Data Integration

Integrating cough audio with clinical, demographic, radiological, and genomic data has been substantiated to augment tuberculosis identification and disease prognosis [35][32][32][24]. The findings indicated that multiplexed graph neural networks and hybrid deep learning architectures exhibit superior performance by encapsulating intricate inter-modality relationships that singular-modality methodologies cannot [24][1]. Multimodal frameworks possess greater resilience against noise and variability that are intrinsic to a singular data type [34][15]. Nevertheless, considerable multimodal integration frequently necessitates datasets that are not uniformly accessible across environments, culminating in challenges associated with missing data and resultant model bias [26][24]. The comprehensive design of fusion models can, intrinsically, compromise clarity and introduce obstacles for their application in clinical situations [1]. Furthermore, differences in strategies and how characteristics are illustrated in multiple research projects impede the feasible establishment of consistent integration frameworks broadly [12].

C. Labelling and Annotation Techniques

Only some of the studies include the clinically validated scoring systems used in measuring symptom severity, such as the TB score and the Timika score, which objectively support the training of the algorithm as well as prediction - [20][13][5][6]. Functionality of the database of cough sound recordings relies on the involvement of humans or experts in annotating the noises, including segmentation of coughs [36] [40]. However, despite the unreliability of self-reported symptoms data and their status as false disease classifiers, the use of self-reported symptom data is widely prevalent [8]. The absence of standard procedures in labelling the input and output data leads to inconsistent inputs and outputs of studied algorithms [8][28].

D. Performance and Benchmarking of Machine Learning Models

Among different machine learning (ML) techniques studied, DNNs (Deep Neural Networks), GBTs (Gradient Boosting Trees), and CNNs (Convolutional Neural Networks) generated high diagnostic performance in several cases. Some of these approaches were found to achieve an AUROC (Area Under the Receiver Operating Curve) value greater than 0.9 [9][34][1][14]. An example of such study could be the CODA TB DREAM challenge that resulted in transparency due to open benchmarking of the methods and thus fast progress in the creation of models through community participation. Application of methods such as SMOTE that deals with data imbalance problem or cross-validation improves the performance of the model [9][18]. Despite the fact that many models show very good performance, they might be trained and validated using datasets from areas with limited geography and demography that raises a question about generalizability to other populations [16][15]. Moreover, some models, according to studies, have poor performance in real-life settings and just a few of these models have been prospectively validated in the clinical setting [15][4]. Standardization of assessments used in the studies can make a comparison between them complicated [12].

E. Explainability and Clinical Usefulness

There have been numerous studies that make use of various explainability techniques, specifically SHAP values, to provide information on the significance of the different features involved, which will allow for greater insight into the predictions of the models from a clinical perspective [25][30]. Combining the interpretable clinical features with the acoustic features will yield useful insights that can be used clinically [1][11]. Nonetheless, several high-performing models are still essentially black boxes, hindering the confidence of clinicians and their integration into routine workflows [12][1]. There is often insufficient funding consideration for the balance between complexity and interpretability, and user-centred design and implementation challenges are among the few things reported [7]. To date, the clinical effect of these models on patient outcome is not known [8].

F. Longitudinal Data and Treatment Monitoring

Constant monitoring of cough and tracking of symptoms over time can serve as dynamic biomarkers for treatment response with considerable potential for personalized TB care [3][22][40]. Some ML frameworks use time-series data to predict treatment outcomes and determine patients who are at risk of failure. On the other hand, longitudinal data collection is plagued by major practical obstacles, especially adherence problems, technical problems, and missing data problems all of which are far more severe in low resource settings [3][40]. So far, few studies have made use of temporal modelling and the combination of longitudinal symptom severity data with other clinical modalities has not yet been exploited [8]. The need for further prospective validation before endorsements can be made for the adoption of longitudinally derived ML predictions is on the table.

G. Low-Resource Settings Data Collection

Many research has established the possibilities of community-based data collection leveraging smartphone and wearable device for scalable TB screening in places not possible with regular diagnostic [3][40][2]. AI systems specifically designed for real-world, noisy auditory settings show promise for deployment under resource-constrained settings [34][15]. In spite of this, infrastructure problems, such as unstable power supply, unstable internet, and equipment maintenance constraints, pose challenges to continuous data collection [3][40]. Inconsistencies in the quality of data and compliance from participants may distort model training and limit generalisability. Ethical considerations concerning data privacy in these settings is another aspect that continues to remain underreported in the literature [39].

V. REVIEW OF LITERATURE THROUGH THE THEMATIC APPROACH

The clinical symptom severity datasets as it relates to tuberculosis for machine learning literature highlights a few major themes that deal with cough-related signal data collection and analysis, multi-modal assimilation of data, and predictive modelling for diagnosis and treatment outcome. There is a lot of focus on the generation of larger standardized cough sound datasets and the use of modern ML architectures for improving the accuracy of screening TB in low resource settings. Clinical scoring systems, image-based severity assessments, and immunological biomarkers were also employed, integrating with ML-based models for greater interpretability and prognosis prediction.

The article shows that data quality, model explainability, and longitudinal monitoring are emerging priorities for TB care optimization

A. Using Cough Sounds Datasets to Screen for Tuberculosis

The reviewed papers show that the theme of developing large-scale multi-country cough audio dataset for TB screening systematically is the most prevalent theme emerging out of 43 papers. Through the initiatives, CODA TB, Hyfe, datasets with hundreds of thousands of cough recordings and rich clinical and demographic metadata have been compiled which has enabled the training and validation of AI models on diverse populations.[10][41][16][17][36][40] A key aspect of these initiatives is their conscious commitment to actual use, including feasibility in low-resource settings and the inclusion of continuous monitoring in treatment. Attention in this case to ecological validity means that an increasingly attentive field to the difference between lab performance and real performance.

B. Machine learning models using clinical symptom severity

Out of the total studies, 43 studies were subject to evaluation and 19 were directly related to diagnosis and treatment outcome prediction. 19 of this total review utilized a number of ML algorithms on clinical symptom severity data. Machine learning models such as ANN, RFs and CNNs have outperformed traditional statistical models, and some of them have achieved accuracies above 90% [9][21][29][14][30]. The use of techniques such as SMOTE for imbalance correction and cross-validation for improving generalizability shows the sound methodology used by the authors. In summary, the findings reveal a substantial progress of the machine learning approaches in diagnosing TB cases compared to the traditional methods.

C. Integrating Multiple Data Types to Assess TB Severity

Though single-modality models have shown significant potential, 13 studies revealed that multiple data streams offer extra diagnostic value. The link of cough audio with clinical metadata, chest X-ray imaging and genomic data consistently improves the accuracy and interpretability of the TB severity assessments [14][35][5][32][24][31]. There are several different strategies for fusion from integrating deep learning at a late stage to graph neural networks that explicitly model cross-modal relationship patterns. This shift towards multimodality shows a growing trend in the field towards more holistic and patient-centred modelling of TB. The severity of TB cannot be captured with a single biomarker or data type.

D. Methodologies of Clinical Score and Annotation

The quality and consistency of clinical annotation is at the heart of either unimodal or multimodal modelling. The tuberculosis (TB) severity scoring system has been refined and validated in 11 studies, with the TBscore being the most commonly reviewed. The Timika score is another such standardization that can allow reproducible model training and outcome prediction [20][13][22][5]. Efforts are being made to adjust annotation methods for symptoms over time and comorbidities, including diabetes and HIV. This is to ensure it more accurately reflects what occurs in the clinic. The presence of well-annotated data is important for making sure that ML models are optimized against clinically meaningful outcomes.

E. Explaining and Making ML Models Interpretable

In addition to accuracy, clinical implementation of ML models depends on explainability and trustworthiness. Across 8 studies, the implementation of interpretability techniques represents the intention of mapping algorithmic output to clinical reasoning, most notably through major use of SHAP (from Shapley values) and model explanation [25][30][11][1]. Given the circumstance that these methodologies concentrate on recognizing attributes accountable for eliciting model

forecasts, such strategies enhance the categorization of hazards pertaining to therapeutic results and uncovering biomarkers pertinent to the medical environment. Applying these techniques becomes especially necessary in tuberculosis treatment as resources are not easily accessible.

F. Longitudinal Monitoring and Treatment Response

The longitudinal monitoring introduces an important element of time into the TB management process which is not possible to do using static systems and should work in tandem with the cross-sectional diagnosis method. From the 7-research conducted, it was discovered that cough sound analysis and regular collection of symptoms data have found a significant correlation between the metrics of cough and the microbiological results as well as treatment effectiveness. [3][22][40] The use of this period will make it easier to detect cases of treatment failure or bad responses to medications.

G. Obstacles and Practicality in Low-Resource Environments.

The effectiveness of these artificial intelligence-based TB detection systems hinges on their feasibility to function in the environments that suffer the most. Based on results from six studies conducted in low-and-middle income countries, the contradiction of technology is a perpetual challenge. According to [3][2][40][4], unreliable mobile internet connectivity, electricity supply issues, and noncompliance from participants are serious barriers. Nonetheless, this same body of evidence indicates that mobile-based screening tool is possible and that solutions can be scaled up when designed with local contextual constraints in mind. These findings reinforce the importance of co-designing AI tools with end-users and health systems and not adapting tools developed in high-income contexts.

H. Developing New Representations and Engineering Features

The development of high-quality acoustic feature representations, which can accurately distinguish between TB cough and coughs associated with other respiratory conditions, is a critical enabler of cough-based TB screening. Improvements in feature engineering including the use of mel-spectrograms and Mel-frequency cepstral coefficients (MFCCs) as well as NLP-style cough embeddings and noise-robust extraction methods significantly improve classification performance in difficult acoustic settings [23][37][37][34]. It should be acknowledged that there exist numerous methodologies that integrate the design paradigm concerning the management of sensitive auditory information within communities. Such advancements underscore the extent of progress attained by this field and its heightened emphasis on ethical considerations.

I. Radiological Imaging and Automated Severity Assessment

Alongside robust evaluation strategies, we also get the chance to investigate radiological imaging to automate our assessment of tuberculosis severity. In four autonomous research endeavours, deep neural networks have been determined to yield dependable and precise measurements regarding pulmonary involvement and cavitation identification, delivering an output that exhibits a strong correlation with extant clinical evaluations [5][6][38]. It can be asserted with conviction that convolutional neural networks can furnish successful outcomes when employed for the objective of radiological examination and consequently will contribute to mitigating the burden of radiologists operating under suboptimal conditions.

J. Diagnosis and prognosis of TB using biomarkers

Three studies, in addition to acoustic and imaging tools, examined the utility of machine learning for prediction of TB using clinician-collected biomarker data such as plasma immune profiles and routine laboratory indicators for predictive signatures of TB diagnosis and clinical outcomes [19][25]. Approaches that provide non-sputum-based diagnosis which is particularly relevant for patients unable to produce adequate sputum. This area is still in its infancy but mentioned in the reviewed literature, signalling growing enthusiasm for incorporating molecular and immunological information in AI-based TB care pathway.

K. Problems of Data Quality, Dataset Bias and Standardization

Virtually all thematic areas raise key issues relating to the quality and standardisation of data. As outlined in 3 studies, problems in dataset representativeness, variability of annotations, inconsistency in symptom reporting, are deep-rooted issues that can enable vulnerabilities in models [8][36]. Different studies

usually collect data through different methods. This makes comparison difficult and also meta-analysis. Tackling these problems with standardisation of data frameworks, annotation protocols, and reporting are not only technical priorities but requirements for responsible clinical translation of AI tools for the care of TB.

L. Innovation in AI: Advocacy of Open-Access Data for Public Good

A possible answer that may tackle data quality and access challenges includes open-access platforms and community-based benchmarking initiatives. As can be seen from the above, the CODA TB DREAM challenge has been showcased in two reviewed studies, thus gathering a number of scientists from various backgrounds and having them come up with TB screening algorithms based on their findings concerning exactly the same dataset. [16][16] While the activities might be interpreted as a competition, they will also result in a culture of replication and open science that may prove highly useful in an environment where data tends to be scarce and private. Contribution of the organization towards the future research in TB and their efforts in making AI development more democratic regarding TB is important.

M. Innovations in Techniques and Algorithms of AI

We think that the development of new algorithms that can strengthen multimodal models will determine the future of TB AI research. Two papers have noted that architectures like the use of cross-attention and balanced risk loss function were created specifically to improve TB risk prediction. They endeavour to provide all three attributes concurrently – specifically, instantaneous inference, elucidation, and appropriateness for resource-constrained implementation – an occurrence that is atypical among traditional methodologies. Although still nascent, these developments clearly point towards the emergence of the next wave of TB AI technologies – innovations that will be more precise as well as contextualized for deployment where they matter most.

VI. CHRONOLOGICAL ASSESSMENT OF LITERATURE

In recent years, the use of clinical symptom severity datasets for tuberculosis using machine learning has seen a number of advancements in relation to improvements in the data collection techniques and approaches towards creating multimodal datasets as well as advanced predictive models. This process started with clinical scoring methods and simple symptoms documentation that has since been replaced with the analysis of cough sounds in addition to other demographic and clinical information. In recent times, there has been more emphasis placed on the role of multimodal data fusion, model interpretability, ML architectures, and large datasets with multiple forms of data such as imaging and genomics.

A. Scoring of clinical symptom severity and conventional assessments from 2008 to 2013

Some of the earliest works done during this time were those aimed at creating and validating clinical severity scores for symptoms such as coughing, bleeding, and weight loss for purposes of tracking the treatment process and the patients' outcomes. These scores were especially helpful where resources were limited and would be the foundation on which later machine learning techniques could build their studies.

B. Timeframe 2014-2020: Implementation of ML and Acoustic Analyses to Diagnose TB

During the initial era, application of data mining and MLTs for TB diagnosis using patient symptoms and sputum test results together with cough audio characteristics was performed for the first time. Studies focused on examining basic classifiers and acoustic markers using small databases. Continuous obstacles appeared concerning complications in data, effectiveness evaluation, and the inadequacy of a stringent validation technique.

C. Cough monitoring and dataset generation, 2021-2023

The advancements in cough monitoring through smartphones allowed for a scientific evaluation of coughing in different environments. The new datasets, containing hundreds of recorded cough sounds with corresponding demographic and clinical data, provide considerable progress for training ML algorithms. The main emphasis during this period is placed on demonstrating the feasibility, creating effective preprocessing methods, and then training initial ML models.

D. Using integration and better prediction models

The scope of this study has greatly expanded to cover cough sounds along with clinical data, imaging data, and

genomic data, which are used to improve TB diagnosis and predict treatment outcomes. To bolster the effortless unification of these data assets, machine learning-adapted structural models, featuring convolutional neural networks, gradient boosting solutions, and graph neural networks have been designed. As the years progressed, researchers made notable advancements in the methodologies used for labelling strategies, symptom severity scoring systems, and longitudinal data collection methods. Furthermore, the research community became increasingly focused on the interpretability of models as well as their robustness to noise when deployed in the real world. The period of 2024-26 will emphasize complex multimodal fusion, interpretability as well as real-world tests.

In the latest times, we have seen complex multimodal fusion frameworks for integrating acoustic, imaging, clinical, and genomic features through multiplexed graph neural networks and cross-attention mechanisms. Strong focus on explainable AI, uncertainty quantification, and benchmarking across a range of populations paves the clinical adoption pathway. The rapid innovation of algorithms has been made possible through the use of large-scale open challenges and standardized protocols, with improvements in adherence and data quality in low resource deployment settings still ongoing.

VII. STUDIES THAT ARE SIMILAR BUT DIFFERENT

The studies reviewed in this article allude that the severity dataset of clinical symptoms in TB, more particularly the cough-based acoustic datasets, and machine learning of cough samples have shown to be increasingly relevant and feasible. Experts agree that multimodal data fusion, involving clinical-acoustic-radiological-genomic data, improves predictive capacity and facilitates precise patient stratification. Article average analytical bibliography created new public imputation dataset TB diagnose, prognose, cause, result and others. The background material based expert classical and model analysis machine learning. Such discrepancies illustrate challenges in aligning methods for wider clinical use, even as the field makes distinct strides toward more sensitive, accessible and interpretable diagnostics.

A. Configuration of the dataset

The majority of studies take advantage of large and multi-country datasets with highly detailed clinical symptom severity features i.e. cough sounds together with demographic and clinical metadata. There is a general agreement about the positive impact of diverse and multimodal data on improving TB prediction [10][41][16][17][11][32][32][24]. Differences in dataset quality emerge through annotated data quality: some datasets are constructed from clean and deidentified questionnaire data [42][42], whereas some datasets feature continuous longitudinal cough monitoring [3][40]. The representativeness of a dataset depends on its size and participation. The differences arise from the availability of resources, the level of the study (controlled vs. real-world), and the population studied (symptomatic vs. asymptomatic).

B. Models Efficiency

Typically, high-performing models use cough acoustic features which also include clinical metadata, and they often report AUROC values greater than 0.80 [12][34][35][15][1]. According to several studies found in the literature, XGBoost, Random Forest, and CNN are broadly effective machine learning approaches. It is noteworthy that the amalgamation of modality information consistently yields enhanced sensitivity and specificity [9][21][29][16][16][15]. However, other work reports a much lower predictive performance when solely based on cough audio (AUC ~0.70) [17][17][37], while using severity data in isolation claims over 99% accuracy [9][29]. Algorithms that perform the best differ from one study to the next, with some favouring deep learning [14], others tree-based methods [25] and others hybrid architectures [34]. This largely reflects differences in the size of the dataset, the quality of the features and the validation protocols.

C. Multimodal Fusion Techniques

Most studies agree that integrating clinical scores with other modalities, either acoustics from cough and radio-imaging or genomic information, will provide better TB diagnosis and prediction of prognosis [32][32][24][31]. Graph neural networks and gradient-boosted decision trees are being increasingly recognized as effective fusion methods. For fusion, approaches based on shared latent feature space outperform those relying on simple feature concatenation [19][26][24][1]. There is divergence regarding the type of fusion strategy - some

studies favour a late fusion [2], while other opt for intermediate or multiplexed graph fusion [24]. Differences in the selection of modality and the fusion timing are determined by the difference in availability of data, computational resource and research priorities. As novel frameworks such as multiplexed GNNs require large multimodal datasets, not all researchers have access to these datasets.

D. Labelling and Annotation Quality

Research studies globally agree on labelling with clinically relevant terms, including through validated symptom severity scoring systems such as TBscore and TBscore II, occurrence and longitudinal labelling to aid assessments of treatment response [20][13][3][22]. Annotation of cough data is done via experts or automated algorithms [36][43]. The continuation of some studies on self-reported symptoms [8], data de-identification practices that limit the granularity of annotation [42][42], and differences in the segmentation and classification of cough events across studies [36] are the causes of divergence. These differences may be related to study designs and use of different resources and definitions of severity of symptoms.

E. Interpreting models

Recent works are relying more and more on explainability, e.g. SHAP values, ranking of features, and interpretable model architecture to help clinicians trust their predictions [19][25][30][1][11]. Though there have been numerous advances, many high-performance deep learning models remain effectively opaque [34][14] with complex graph neural network fusions architectures proving particularly difficult [24]. The magnitude of interpretability examination across investigations spans from thorough frameworks to constrained documentation [9], exemplifying the compromise between the intricacy and lucidity of models. The urgent demand for reliable artificial intelligence in tuberculosis management is inciting continuous inquiry into more elucidative methodologies.

VIII. SIGNIFICANCE OF THE STUDY

A. Theoretical Consequences

The amalgamated results from contemporary research provide additional substantiation to the hypothesis that cough sound assessment, in conjunction with clinical symptom intensity, serves as a robust prognosticator for tuberculosis identification and surveillance. This engenders a pronounced dependence on sputum-derived data and self-reported symptoms, which have been demonstrated to be insufficient classifiers of pathological conditions [8][4][7].

The application of multi-faceted data encompassing acoustic cough attributes, clinical metadata, radiographic images, and genomic data enhances theoretical frameworks by supplying corroborative evidence that the severity of tuberculosis and therapeutic outcomes are optimally predicted through intricate data amalgamation from diverse sources rather than relying on a singular source [32][24][38].

ML algorithms, especially those using deep learning architectures and graph neural networks, have greater capability to capture non-linear relationships and time dynamics associated with severity of symptoms datasets. This confirms the theoretical assumption that advanced artificial intelligence techniques can discover patterns that classical statistics cannot [24][1][11].

Research has shown that cough characteristics vary between established patient subgroups, such as those with a history of tuberculosis (TB), and diabetic comorbidities. In addition, non-linear trajectories of increasing symptom severity have been observed after treatment initiation. In view of this observed heterogeneity, the authors suggest that more effective modelling approaches may require ongoing longitudinal methods as opposed to static cross-sectional methods [22][30][39].

As per the results, it's essential that ML models are interpretable to be in accordance with clinical reasoning. Overall, it suggests that XAI frameworks are important for both theoretical and practical acceptance in a clinical setting [25][1][30]

B. Practical Implications

The evidence that cough monitoring can feasibly and effectively be undertaken continuously in the community using mobile phone applications and artificial intelligence (AI) algorithms can provide a low-cost and scalable screening tool for resource-limited settings. Such a tool can be used at scale to transform TB case-finding and treatment monitoring [3][40][4].

The creation of open-access datasets and collaborative benchmarking challenges, such as the CODA TB DREAM challenge, has the potential to stimulate the development and evaluation of cough-based TB screening algorithms, demonstrating the applicability of

open-data and benchmarking for global health technology [16][16][34].

Linking cough sounds with routine clinical and demographic data will improve the accuracy and specificity of a diagnosis and supports the deployment of multimodal AI tools at primary healthcare level to triage patients for confirmatory testing and avoid delays [12][17][11].

Machine learning models trained on large and diverse datasets have been shown to have high predictive performance for TB diagnosis and treatment outcomes, indicative of potential readiness for clinical decision support applications that can inform individualised treatment strategies and resource allocation [26][27][29].

The creation of noise-robust, interpretable AI models for real-time inference on mobile devices will remove major barriers to deployment, thus facilitating use in remote and underserved communities, where standard diagnostics are not deployed.

Policy frameworks must consider AI-generated cough analysis integration as a useful adjunct for existing TB control programs. However, incorporation must undergo validation in different epidemiological situations and link with public health surveillance to ensure population-level gains [7][4][2].

IX. LIMITATIONS OF THE LITERATURE

Although this appraisal accentuates certain promising progressions, it harbours numerous constraints that necessitate contemplation when clarifying the results and for subsequent investigation.

Limited Sample Sizes: Due to minimal quantities of participants and/or datasets utilized by numerous scholars, there exists a deficiency of statistical power and generalizability in many investigations. With a reduced number of samples, there is a propensity for overfitting in models, culminating in diminished external validity and efficacy of algorithms [37][36][15].

Geographic Bias: The predominance of data collections employed for analysis and research conducted may exhibit a particular geographical locale or national predisposition, which restricts the applicability of the conclusions to alternative regions. Geographic skewness diminishes external validity and complicates the assurance of impartiality in the application of prediction algorithms [16][16][15][40].

Imbalanced Dataset: An imbalanced dataset signifies that the ratio of TB-positive and TB-negative subjects is disparate, yielding biases that culminate in suboptimal performance of the algorithm in practical applications [9][18][27].

Limited Data on Longitudinal Symptoms: A highly constrained number of investigations have considered longitudinal assessments and data pertaining to the magnitude of symptoms or cough surveillance. This restriction hampers the assessment of how the illness advances and how well the pharmaceutical treatments perform.

Limited Application of Multiple Methods: Numerous investigations have confronted the challenge of employing solely a singular methodology for tuberculosis detection, thereby constraining the effectiveness of evaluations, which concurrently signifies a reduction in the degree of accuracy of tuberculosis severity identification [35][24][32].

Un-standardized Methods of Labelling: The absence of standardization in severity scoring frameworks and labelling methodologies constitutes one of the methodological obstacles encountered by research in tuberculosis detection [20][13][36].

Technology and environmental limitations: Device variability, background noise, as well as compliance of subjects poses certain issues that negatively affect coughs data collection [3][40][15]. Therefore, it results in a drop in the standard of gathered data and the credibility of the model. This point is particularly significant in developing countries where the application is more urgent.

Limited External Validation: Many machine learning models lack extensive validation in other patients' population. Thus, it raises a question regarding generalizability of the model. Also, its clinical value outside the scope of initial research is questionable.

Over-reliance on Acoustic Characteristics: Using acoustic characteristics only for the detection of disease without taking into account clinical information or other biomarkers might lead to reduced accuracy in diagnosis and model interpretation [17][21][37] in cases of unusual manifestation of tuberculosis.

Model's Deficiency of Clarification: Elucidation of a model's outcomes is critical from a clinical perspective. Therefore, limited model explainability makes clinicians sceptical towards the implementation of such model [9][25][1].

X. GAPS AND DIRECTIONS FOR FUTURE RESEARCH

The findings generated by synthesizing literature on this topic have identified various gaps that will help pave the way to developing clinically meaningful and generalizable ML tools for TB management. These issues are discussed below.

Standardizing the Scoring and Labelling of Symptom Severity: One of the key obstacles to ensuring the ability to compare findings between studies and generalize the models using the datasets of TB patients has been the inconsistent application of the symptom severity scoring methods and labelling process [3][16][8][20]. For future research on the topic, it is critical that universally accepted standards for scoring and labelling the symptoms of TB be developed and agreed upon. Without such an approach, the creation of a harmonized database cannot be achieved.

Collection-and-use of-longitudinal-data: The failure to adequately exploit longitudinal symptom severity and cough monitoring data to model treatment response represents a considerable missed opportunity [3][40][22][26]. Subsequent work ought to be on the design of efficient longitudinal data collections mechanism with enhanced adherence support strategies. Further, the development of temporal ML models that incorporate longitudinal evolution of symptom severity with multimodal data for monitoring treatment effects in real-time. A continued investment in this area can greatly enhance personalisation of TB care.

Frameworks of Multimodal Data Fusion: To date, there are no widely adopted, interpretable, and scalable multimodal fusion methods for clinical, acoustic, radiological, and genomic data [12][24][1][31]. It is imperative that scientists examine and calibrate assorted fusion architectures such as graph neural networks and cross-attention modules, which include built-in explainability components. Also needed are frameworks that can tolerate missing modalities and heterogeneous input data. In absence of standardisation, clinical adoption of multimodal AI tools for TB will remain low.

Understanding Models and Gaining Clinical Confidence: Numerous models that perform well forget their interpretability. This leads to a lack of clinical validation and impacts trust. Moreover, it also impacts the adoption which occurs at the frontline of the health system. Further research should incorporate explainability methods, such as SHAP values, feature importance, and graph-based explanation techniques, the TB severity models. Moreover, user-centred studies should capture the required interpretations within specific clinical workflows and settings.

Diversity and generalizability of dataset: Current datasets often lack enough geographical, demographic, and clinical diversity for the model to be applied [16][15][11]. Next, more data from under-studied regions and populations should be gathered. Investment in studies that explicitly validate externally and in studies of domain adaptation should be made to assure model robustness and further equitable applicability.

Objective versus Self-Reported Symptom Data: The literature still largely relies on self-reported symptoms which have proven too inaccurate to classify TB disease state [8][20][4]. It is recommended that future studies assess objective measures of symptom severity (for example, continuous cough measurement and validated clinical scores) and develop hybrid models that combine both objective and subjective data with agreed reliability and predictive validity.

Gathering Data in Willing but Low-Resource Settings: Technical, infrastructural and adherence challenges continue to impede the collection of high-quality data in the highest burden settings for TB [3][40][2]. Future efforts must include innovations for the design of low-cost, low-energy-use, and user-friendly data collection devices; implementation of community engagement and adherence support strategies; and more thorough engagement with the ethical and privacy dimensions of data collection in these contexts.

Benchmarking and Evaluation Guidelines Keeping in Mind: The absence of standardized evaluation criteria and protocols makes fair and transparent comparison of TB ML models in the literature complicated [12][16][18]. The processes of evaluation should be agreed upon (for example: sensitivity, specificity, AUROC, uncertainty quantification) and open benchmarking challenges should occur with common datasets, facilitating more objective and useful evaluations.

Combination of Treatment Outcome Prediction: Models integrating multimodal data and temporal dynamics for treatment outcome prediction have not yet integrated symptom severity datasets [26][32][30]. Future investigations should entail the development as well as prospective validation of predictive models integrating symptom severity with clinical, radiological and genomic data longitudinally that produce therapeutic

success/failure; allowing earlier as well as tailored clinical interventions.

Extraction of Features While Preserving Privacy: Several studies have failed to address privacy concerns related to cough audio data in sufficient depth [37][38]. Such shortcomings limit data sharing and collaborative model development. Efforts in the future should aim to design and validate privacy-preserving feature extraction methods - like an NLP-style cough embedding - that retain diagnostic utility while protecting patient identity, to enhance access to audio-based datasets.

XI. OVERALL INTEGRATION AND CONCLUSION

The body of literature reviewed in this paper indicates substantial progress in developing a clinical symptom severity dataset for TB and its application within machine learning frameworks. The development and application within ML frameworks have clearly focused on multimodal application. Datasets combining cough audio, clinical metadata, radiological imaging, and genomic information that are large-scale and geographically diverse allow for better and more robust prediction models that are more generalizable than unimodal approaches. The use of ongoing or longitudinal cough monitoring has supplied essential temporal data regarding the cough disease dynamic (e.g. change made by the treatment) but patient adherence and the quality of data in low resource settings remains a challenge that is significant. Clinical scoring systems that are objective, like the TBscore, have been established as reliable standards for annotation. They can either be used to support inconsistent self-reported symptom data or replace them to enhance the quality of datasets. Moreover, they also help improve the validity of the model inputs.

Machine learning systems using ensemble methods, deep learning architecture and advanced fusion techniques consistently demonstrate high diagnostic accuracies. In general, multimodal systems perform better than unimodal systems for various evaluation metrics. Methods of fusion, like late fusion and multiplexed graph neural networks, can efficiently capture complex cross-modal interactions that can enhance TB detection and treatment outcome prediction. Notwithstanding all these modifications, challenges pertaining to diverse data collection methodologies, absent data, and varying attributes' representations persist as impediments on the pathway towards standardization and inter-study comparison. Although recent models demonstrate insufficient prospective clinical validity and lack geographic representativeness, the development of open data initiatives and benchmarks for algorithmic assessment drives the advancements in Machine Learning solutions.

One of the most important areas, which remains poorly developed, concerns the interpretability of predictions. As novel applications of Shapley values, feature importance studies, and uncertainty analysis techniques emerge, they facilitate clinicians in gaining trust towards their use for making decisions. Machine Learning models providing interpretable results and targeting health professionals can be highly useful for practical implementation. It is undeniable that this is especially significant for domains confronting monetary limitations, as their models ought to be both workable and transparent. Nonetheless, numerous effective models lack interpretability, and finding a balance in this aspect is challenging.

The combination of longitudinal assessment of symptoms' severity, along with the use of multimodal features, although novel, creates a promising pathway towards achieving personalization in monitoring therapy response and being able to detect early signs of therapy failure. The research proves that objective counts of coughing episodes and a global clinical score are potentially valuable and measurable biomarkers. However, there is still much work to be done to overcome the problems of low adherence and availability of technological means for gathering information from such communities. Thus, although the field of ML applications in tuberculosis diagnosis and treatment is steadily progressing towards achieving accuracy, interpretability, and scalability of ML approaches, further research should also pay attention to developing guidelines for assessing symptoms' severity and validating findings in heterogeneous populations.

ACKNOWLEDGMENT

The author sincerely acknowledges the support and facilities provided by Sri Krishna Arts and Science College for enabling the successful completion of this research work.

REFERENCES

[1] Lu Z, et al. "DeepGB-TB: A Risk-Balanced Cross-Attention Gradient-Boosted Convolutional Network for Rapid, Interpretable Tuberculosis Screening." arXiv, vol. abs/2508.02741, Aug. 2025. doi:10.48550/arxiv.2508.02741.

[2] Rudraraju G. "Development and Clinical Validation of Swaasa AI Platform for Screening and Prioritization of Pulmonary TB." Sept. 2022. doi:10.1101/2022.09.19.22280114.

[3] Raberahona M, et al. "Continuous Digital Cough Monitoring During 6-Month Pulmonary Tuberculosis Treatment." ERJ Open Research, Oct. 2024. doi:10.1183/23120541.00655-2024.

[4] Zimmer A, et al. "Making Cough Count in Tuberculosis Care." June 2024. doi:10.60692/v5h11-kr197.

[5] Ghanem M, et al. "Percent of Lung Involved in Disease on Chest X-Ray Predicts Unfavorable Treatment Outcome in Pulmonary Tuberculosis." medRxiv, Aug. 2024. doi:10.1101/2024.08.19.24311411.

[6] Kantipudi K, Bui V, Jaeger S, Yaniv Z. "Automated Pulmonary Tuberculosis Severity Assessment on Chest X-Rays." Apr. 2024. doi:10.1007/s10278-024-01052-7.

[7] Chowdary GJ, et al. "Can We Triage for Pulmonary Tuberculosis from the Sound of a Cough? A Comprehensive Technical Review of Artificial Intelligence-Based Approaches." Jan. 2026. doi:10.31224/6230.

[8] McCreesh N, MacPherson P, Bampi JVB, Engel N, Kranzer K, Khan P. "Reported Tuberculosis Symptoms: An Inadequate Classifier of Disease State." Aug. 2025. doi:10.31219/osf.io/fhpeu_v1.

[9] Sampath S, Dhanalakshmi S. "Tuberculosis Prediction: Performance Analysis of Machine Learning Models for Early Diagnosis and Screening Using Symptom Severity Level Data." International Journal of Basic and Applied Sciences, vol. 14, no. 1, pp. 435–444, 2025. doi:10.14419/parmkr90.

[10] Huddart S, et al. "A Dataset of Solicited Cough Sound for Tuberculosis Triage Testing." Scientific Data, vol. 11, no. 1, 2024. doi:10.1038/s41597-024-03972-z.

[11] Parihar DS, et al. "Neural Network-Based Identification of Easily Obtainable Demographic and Clinical Characteristics to Identify People with Tuberculosis." Oct. 2025. doi:10.1101/2025.10.23.25338536.

[12] Kafentzis GP, Selisios E. "Tuberculosis Screening from Cough Audio: Baseline Models, Clinical Variables, and Uncertainty Quantification." Jan. 2026. doi:10.48550/arxiv.2601.07969.

[13] Wejse C, et al. "TBscore: Signs and Symptoms from Tuberculosis Patients in a Low-Resource Setting Have Predictive Value and May Be Used to Assess Clinical Course." Scandinavian Journal of Infectious Diseases, vol. 40, no. 2, pp. 111–120, 2008. doi:10.1080/00365540701558698.

[14] Landry G, Malumba RN, Kabutakapua FCB, Mangata BB. "Performance Comparison of Classical Algorithms and Deep Neural Networks for Tuberculosis Prediction." Jurnal Techno Nusa Mandiri, vol. 21, no. 2, pp. 126–133, 2024. doi:10.33480/techno.v21i2.5609.

[15] Ma N, et al. "AI-Enabled Tuberculosis Screening in a High-Burden Setting Using Cough Sound Analysis and Speech Foundation Models." Sept. 2025. doi:10.48550/arxiv.2509.09746.

[16] Jaganath MD, et al. "Accelerating Cough-Based Algorithms for Pulmonary Tuberculosis Screening: Results from the CODA TB DREAM Challenge." medRxiv, May 2024. doi:10.1101/2024.05.13.24306584.

[17] Kafentzis GP, et al. "Predicting Tuberculosis from Real-World Cough Audio Recordings and Metadata." 2023. doi:10.48550/arxiv.2307.04842.

[18] Septiandri AA, Aditiawarman, Tjiong R, Burhan E, Shankar AH. "Cost-Sensitive Machine Learning Classification for Mass Tuberculosis Verbal Screening." arXiv, 2020.

[19] Yao F, et al. "Plasma Immune Profiling Combined with Machine Learning Contributes to Diagnosis and Prognosis of Active Pulmonary Tuberculosis." Emerging Microbes & Infections, 2024. doi:10.1080/22221751.2024.2370399.

[20] Rudolf F. "The Bandim TBscore – Reliability, Further Development, and Evaluation of Potential Uses." Global Health Action, vol. 7, no. 1, 2014. doi:10.3402/GHA.V7.24303.

[21] Ayano YM, Debelee TG. "Robust Cough Analysis System for Diagnosis of Tuberculosis Using Artificial Neural Network." 2023. doi:10.1007/978-3-031-31327-1_1.

[22] Lee GO, et al. "Cough Dynamics in Adults Receiving Tuberculosis Treatment." PLOS ONE, vol. 15, no. 6, 2020. doi:10.1371/journal.pone.0231167.

[23] Yadav J, Varde AS, Liu H, Antoniou GE, Xie L. "Audiovisual Multimodal Cough Data Analysis for Tuberculosis Detection." 2024. doi:10.1109/IISA62523.2024.10786619.

[24] D'Souza NS, et al. "Fusing Modalities by Multiplexed Graph Neural Networks for Outcome Prediction in Tuberculosis." 2022. doi:10.1007/978-3-031-16449-1_28.

[25] Liu ZZ, et al. "Validation and Interpretation of Machine-Learning Models for Rapid Identification of Active Tuberculosis Infection Using Routine Laboratory Indicators." Frontiers in Cellular and Infection Microbiology, vol. 15, 2025. doi:10.3389/fcimb.2025.1718614.

[26] Wang L, Campino S, Clark TG, Phelan JE. "A Multi-Stage Machine Learning Framework for Stepwise Prediction of Tuberculosis Treatment Outcomes." 2025. doi:10.21203/rs.3.rs-7558046/v1.

[27] da Silva MHLF, et al. "Machine Learning Classification of Favorable vs Unfavorable Tuberculosis Treatment Outcomes Using Clinical and Sociodemographic Data from Brazil's SINAN-TB (2001–2023)." 2025. doi:10.21203/rs.3.rs-7502054/v1.

[28] Rusdah, Winarko E. "Review on Data Mining Methods for Tuberculosis Diagnosis." 2013.

[29] Priyono E. "Prediction of Tuberculosis Patients with Machine Learning Algorithms." JIPI, vol. 9, no. 4, pp. 2349–2356, 2024. doi:10.29100/jipi.v9i4.5486.

[30] Peng AZ, et al. "Explainable Machine Learning for Early Predicting Treatment Failure Risk Among Patients with TB-Diabetes Comorbidity." Scientific Reports, vol. 14, 2024. doi:10.1038/s41598-024-57446-8.

[31] Suda C. "Early Detection of Tuberculosis with Machine Learning Cough Audio Analysis: Towards More Accessible Global Triaging Usage." arXiv, 2023. doi:10.48550/arxiv.2310.17675.

[32] Sambarey A, et al. "Integrative Analysis of Multimodal Patient Data Identifies Personalized Predictors of Tuberculosis Treatment Prognosis." iScience, vol. 27, 2024. doi:10.1016/j.isci.2024.109025.

[33] Yadav J, Varde AS, Xie L. "Comprehensive Cough Data Analysis on CODA TB." Proc. IEEE Big Data, pp. 6311–6313, 2023. doi:10.1109/BigData59044.2023.10386805.

[34] Kesarwani S. "AcousticTB: A Hybrid Deep Learning and Gradient Boosting Framework for Noise-Robust Tuberculosis Screening from Cough Audio." 2025. doi:10.36227/techrxiv.175742700.07240558/v1.

[35] Buyung RA, Nugroho W. "Multimodal Deep Learning for Tuberculosis Detection Using Cough Audio and Clinical Data with Health Acoustic Representations (HeAR)." International Journal of Advanced Computer Science and Applications, vol. 16, no. 10, 2025. doi:10.14569/IJACSA.2025.0161036.

[36] David HJ. "TBscreen: A Passive Cough Classifier for Tuberculosis Screening with a Controlled Dataset." 2023. doi:10.5281/zenodo.10431328.

[37] Pahar M, Theron G, Niesler T. "Automatic Tuberculosis Detection in Cough Patterns Using NLP-Style Cough Embeddings." 2022. doi:10.1109/ICEET56468.2022.10007261.

[38] Wang Q, et al. "Analysis of a Large Patient-Level Dataset to Predict Outcome of Treatment for Drug-Resistant Tuberculosis." medRxiv, 2022. doi:10.1101/2022.09.14.22279738.

[39] Pandey SK, et al. "Prediction of Tuberculosis Disease Progression with AI Analysis of Clinical Data." 2023. doi:10.1109/ICAIIHI57871.2023.10489091.

[40] Huddart S. "Feasibility and Analysis Considerations of Continuous Community-Based Cough Monitoring in Low- and Middle-Income Settings." 2022. doi:10.1101/2022.07.06.22277224.

[41] Huddart S, et al. "Solicited Cough Sound Analysis for Tuberculosis Triage Testing: The CODA TB DREAM Challenge Dataset." medRxiv, 2024. doi:10.1101/2024.03.27.24304980.

[42] Rehan S, et al. “TB Spectrum Dataset (Cleaned, De-identified).” 2024. doi:10.6084/m9.figshare.27634866.

[43] Sharma M, et al. “TBscreen: A Passive Cough Classifier for Tuberculosis Screening with a Controlled Dataset.” Science Advances, vol. 10, 2024. doi:10.1126/sciadv.adi0282.