East African Journal of Information Technology
eajit.eanso.org
Volume 7, Issue 1, 2024
Print ISSN: 2707-5346 | Online ISSN: 2707-5354
Title DOI: https://doi.org/10.37284/2707-5354

EAST AFRICAN NATURE & SCIENCE ORGANIZATION

Original Article

Explainable AI for Transparent and Trustworthy Tuberculosis Diagnosis: From Mere Pixels to Actionable Insights

Tibakanya Joseph[1], Male Henry Kenneth[1] & Nakasi Rose[1]*

[1] Makerere University, P. O. Box 7062, Kampala, Uganda.
* Corresponding Author ORCID ID: https://orcid.org/0000-0001-7368-3808; Email: jtibakanya@gmail.com

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

Date Published:

08 October 2024

Keywords:

Machine Learning, Transparency, Trustworthy AI, Explainable AI (XAI), Tuberculosis Diagnosis.

ABSTRACT

Building transparent and trustworthy AI-powered systems for disease diagnosis has become more paramount than ever due to a lack of understanding of black box models. A lack of transparency and explainability in AI-driven models can propagate biases and erode patients' and medical practitioners' trust. To answer this challenge, Explainable AI (XAI) is drastically emerging as a practical solution and approach to tackling ethical concerns in the health sector. The overarching purpose of this paper is to highlight the advancement in XAI for tuberculosis diagnosis (TB) and identify the benefits and challenges associated with improved trust in AI-powered TB diagnosis. We explore the potential of XAI in improving TB diagnosis. We attempt to provide a complete plan to promote XAI. We examine the significant problems associated with using XAI in TB diagnosis. We argue that XAI is critical for reliable TB diagnosis by improving the interpretability of AI decision-making processes and recognising possible biases and mistakes. We evaluate techniques and methods for XAI in TB diagnosis and examine the ethical and societal ramifications. By leveraging explainable AI, we can create a more reliable and trustworthy TB diagnostic framework, ultimately improving patient outcomes and global health. Finally, we provide thorough recommendations for developing and implementing XAI in TB diagnosis using X-ray imaging.

APA CITATION

Tibakanya, J., Kenneth, M. H. & Nakasi, R. (2024). Explainable AI for Transparent and Trustworthy Tuberculosis Diagnosis: From Mere Pixels to Actionable Insights. *East African Journal of Information Technology*, *7*(1), 341-354. https://doi.org/10.37284/eajit.7.1.2276

CHICAGO CITATION

Tibakanya, Joseph, Male Henry Kenneth and Rose Nakasi. 2024. "Explainable AI for Transparent and Trustworthy Tuberculosis Diagnosis: From Mere Pixels to Actionable Insights". *East African Journal of Information Technology* 7 (1), 341-354. https://doi.org/10.37284/eajit.7.1.2276.

HARVARD CITATION

Tibakanya, J., Kenneth, M. H. & Nakasi, R. (2024) "Explainable AI for Transparent and Trustworthy Tuberculosis Diagnosis: From Mere Pixels to Actionable Insights", *East African Journal of Information Technology*, 7(1), pp. 341-354. doi: 10.37284/eajit.7.1.2276.

IEEE CITATION

J. Tibakanya, M. H. Kenneth & R. Nakasi "Explainable AI for Transparent and Trustworthy Tuberculosis Diagnosis: From Mere Pixels to Actionable Insights.", *EAJIT*, vol. 7, no. 1, pp. 341-354, Oct. 2024.

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

MLA CITATION
Tibakanya, Joseph, Male Henry Kenneth & Nakasi Rose "Mtoto Wetu Growth and Temperature Monitoring System for Tanzania Child Healthcare Clinics". *East African Journal of Information Technology*, Vol. 7, no. 1, Oct. 2024, pp. 341-354, doi:10.37284/eajit.7.1.2276.

INTRODUCTION

Tuberculosis (TB) is one of the world's formidable global health challenges, causing death from a single infectious pathogen. The disease has a millenary history, which originated about 8000 years before Christ (B.C.) with lesions in men and animals known as "Pott's disease"(De Farias Gabriel et al., 2023). It is one of the deadliest diseases, infecting close to ten million people every year and killing over 14 per cent, second to COVID-19. Eight developing nations account for two-thirds of all newly infected cases (Ameen et al., 2021). In 2022 alone, it was diagnosed in 10.6 million human patients and caused 1.6 million fatalities (Acharya et al., 2024). The complex disease mainly affects the lungs (pulmonary TB) and may be treated in most instances with a prompt diagnosis and appropriate therapy (Mota Carvalho et al., 2023). Its complexities are exacerbated by factors such as the variety of clinical manifestations, the pervasiveness of drug-resistant strains (Yadav, 2023). The overlap of signs and symptoms with other related respiratory diseases accentuates the need for accurate and rapid diagnostic techniques (Marvin, & Alam, 2022).

While the disease is treatable if identified earlier, conventional diagnostic approaches lack in a few aspects. For instance, chest X-rays, sputum tests, and molecular assays have limitations regarding accuracy and speed in obtaining results. Moreover, Sputum smear microscopy, also known as bacilloscopic, is a technique to search for acid-fast bacilli (AFB) using the Ziehl-Nielsen (ZN) method. It appears to be more rapid to execute and affordable with a qualified expert who can read 300 fields on an image (negative) in over 30 minutes (Mota Carvalho et al., 2023). However, all these traditional approaches are time-consuming, require high-end expertise and may never provide reliable solutions, especially in developing economies. Consequently, artificial intelligence (AI) offers significant opportunities to transform TB diagnosis. AI has unprecedented potential to augment diagnostic accuracy, simplify clinical workflows, and eventually improve patient outcomes (Sachan et al., 2024). However, the patient and the clinical practitioner's trust in AI diagnostics remains to be explored.

Worldwide adoption of AI-powered TB diagnostics demands more than just accurate and rapid diagnosis. Due to the black-box nature of AI models, building transparent, interpretable, and trustworthy AI models is essential (Chaddad et al., 2023). AI models are opaque, they provide outputs from predictions without providing explanations for their thought processes (Adadi, & Berrada, 2018). Yet, the models' lack of transparency, interpretability and explainability can propagate biases and erode trust (Tchuente et al., 2024) in health regulatory bodies, patients, and medical practitioners. To answer this challenge, Explainable AI (XAI) is drastically emerging as a practical solution and approach to tackling ethical concerns in the health sector (Chaddad et al., 2023). XAI's primary aim is to bridge the gap between AI models' predictive power and the need for transparent, interpretable decision-making (Chadaga et al., 2023). Explaining the AI model's outcomes offers more actionable results (Marvin, & Alam, 2022). This can eventually enhance the trust, acceptance, and adoption of AI-powered TB diagnosis in medical practices worldwide.

Against this background, we examine the potential of XAI in the context of tuberculosis (TB) diagnosis, with a particular focus on increasing trustworthiness in AI-powered TB diagnostic tools. Drawing on insights from the intersection of AI, medical imaging, and public health, we hope to shed light on the challenges and opportunities of using XAI techniques to transform tuberculosis diagnosis—from mere pixels captured in X-ray images to actionable clinical explanations that inform evidence-based medical decisions. We argue that XAI is critical for reliable TB diagnosis by improving the

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

interpretability of AI decision-making processes and eventually enhancing trust in AI-powered TB diagnosis.

This work presents three primary contributions to the area of explainable AI (XAI) in tuberculosis (TB) diagnosis. Firstly, it thoroughly examines explainable artificial intelligence (XAI) approaches and how they are integrated into transparent, interpretable, and trustworthy AI-driven TB diagnoses. Furthermore, it assesses the influence of XAI on trust and transparency, illustrating how the provision of explanations increases assurance among healthcare practitioners and those seeking medical treatment. Finally, it examines the advantages and difficulties of using XAI in TB diagnosis, providing insights into enhanced diagnostic precision and the intricacies of integrating XAI into medical processes. We contribute to the World Health Organization (WHO) End Tuberculosis Global Strategy 2020–2035 (Mera et al., 2023). We envision a future in which AI augments, rather than replaces, healthcare providers' expertise in the fight against this global threat.

AI in TB Diagnosis

Kontsevaya et al. (2023) provide an overview of current methodologies for TB diagnosis. Immune-based diagnostics, X-ray, clinical symptoms and scores, cough detection, culture of Mycobacterium tuberculosis and identification of its resistance profile using phenotypic and genotypic methods, including next-generation sequencing, sputum—and non-sputum-based molecular diagnosis are highlighted. A few traditional approaches to TB diagnosis are accurate and rapid (Yadav, 2023). Consequently, many underdeveloped nations, such as India and Indonesia, are the most afflicted through underdiagnosis or a delay in obtaining specialised care (De Farias Gabriel et al., 2023). The challenges are accentuated by the development of drug-resistant TB strains that pose new challenges requiring new diagnostic methods. Conventionally, Chest X-rays, Sputum tests, and Molecular assays are the major diagnostic techniques for pulmonary tuberculosis. Sputum smear microscopy offers rapid and affordable results with a qualified expert able to read 300 fields on an image (negative) in 30 minutes (Mota Carvalho et al., 2023). Nonetheless, the chest X-ray image diagnostic technique is the cornerstone of all TB diagnoses regarding result accuracy (Hansun et al., 2023a). However, reading chest X-rays is subjective and prone to variable, necessitating the use of expert radiologists to ensure correct diagnosis (Ahmad et al., 2023). Sputum tests, which identify the presence of Mycobacterium tuberculosis germs in respiratory secretions, have limited sensitivity, especially in paucibacillary illness or extrapulmonary TB (Mera et al., 2023). Molecular assays, such as nucleic acid amplification tests (NAATs), provide more sensitivity and specificity than traditional sputum microscopy. However, they are resource-intensive and may not be accessible in resource-constrained environments (Kontsevaya et al., 2023).

METHODOLOGY

We employ a literature review approach to examine advancements in XAI for tuberculosis diagnosis (TB) and identify the benefits and challenges associated with improved trust in AI-powered TB diagnosis. The approach employed involved various stages: First, we explored various relevant databases to obtain relevant literature linked to keywords such as Machine Learning, Deep Learning, Tuberculosis Diagnosis, Explainable AI (XAI), Interpretability, Transparency, Trust, and Explainable Machine Learning Techniques. Secondly, the sources we obtained underwent a meticulous examination, considering their pertinence, reliability, and thoroughness of research. During this process, we put special emphasis on scholarly works that have been peer-reviewed, conference papers, and industry reports from reliable sources. Finally, after gathering a lot of information, a careful study was carried out to present a logical case for progress in XAI for tuberculosis diagnosis (TB) and to find out the pros and cons of building trust in AI-powered TB diagnosis.

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

RESULTS AND ANALYSIS

The Role of AI and ML in TB Diagnosis

In recent years, Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), have emerged as promising technologies for enhancing TB diagnosis. These technologies can potentially address many of the limitations associated with traditional diagnostic methods (Codlin et al., 2024; Hwang et al., 2024). Today, AI-based computer-aided detection for TB is already commercially accessible, and various studies have been undertaken to assess its efficacy in clinical settings (Okada et al., 2024). Both ML and DL are gaining popularity for the early identification of TB demonstrating remarkable accuracy, often surpassing the performance of human radiologists (Kumar et al., 2024). Studies indicate that AI-powered TB diagnostics have shown remarkable results comparable to or better than those of experienced medical personnel (Okada et al., 2024). ML techniques like DL enable the automated analysis of chest X-ray images, expediting screening and triaging of suspected TB patients (Hansun et al., 2023b; Murugesan et al., 2024; Tang et al., 2023). Furthermore, AI-powered molecular tests promise to increase the sensitivity and specificity of tuberculosis diagnosis by analysing complicated genomic data using powerful pattern recognition algorithms (Malik et al., 2023). The methods are essential for CT scan feature representation that mainly overcomes the limits of human interpretation of chest radiographs for detecting tuberculosis (Hansun et al., 2023c). The classification of this non-communicable illness using DL and ML uses discriminatory characteristics from extracted CT scan images and cough samples (Kumar et al., 2024). For instance, deep learning automatically pulls fundamental features from data with little previous knowledge (Huang et al., 2023) Thus, numerous research fields apply deep learning. Much recent research has employed deep learning to diagnose TB compared to traditional machine learning-based TB diagnosis research. Therefore, deep learning algorithms hold the future of TB diagnosis, improving the accuracy of the results. However, the interpretability and explainability of the models' thought processes remain a key challenge.

The Black-Box Nature of ML and DL

While AI is bringing about remarkable innovations in TB diagnostics, it has come at the expense of making models less interpretable (Longo et al., 2024) and, therefore, less trustable. In other words, investigations of explainability should be essential components in AI-powered TB diagnostic research (Mamalakis et al., 2023). This is because the widespread adoption of AI and ML in TB diagnosis could be improved by the inherent black-box nature of the AI models (Naz et al., 2023; Weng et al., 2022). The models perform as opaque (Miró-Nicolau et al., 2022), providing outputs from predictions without explaining their thought processes (Adadi, & Berrada, 2018). In the context of medical diagnosis, where openness and interpretability are critical (Van Der Velden et al., 2022), AI's black-box character presents substantial barriers to its acceptability and reliability among practitioners, regulatory agencies, and patients (Marvin, & Alam, 2022; Patrício et al., 2023). To address this difficulty, Explainable AI (XAI) approaches must be developed and implemented, as well as the underlying processes driving AI predictions, interpretable insights into model outputs, and more transparency, accountability, and trust in AI-driven TB diagnostic systems. At all points, we must adhere to the European General Data Protection Regulations (GDPR) of “Right to Explanation”, which aligns well with the XAI aspirations to show stakeholders how the decision was reached (Gurmessa, & Jimma, 2023).

The Need and Benefits of XAI in TB Diagnosis

The pervasiveness of AI-powered TB diagnostics underscores the critical necessity for explainability to lead to usable and trustworthy systems (Panagoulias et al., 2024). XAI underscore the magnitude of transparency, interpretability, and trustworthiness in AI models, especially for health interventions where medical

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

decisions have essential real-world implications (Aranovich, & Matulionyte, 2023). Within medical imaging, XAI has enormous potential for increasing diagnostic accuracy, patient outcomes, and clinician confidence in AI-powered diagnostic systems (Krishnamoorthy et al., 2024).

To date, TB diagnosis mainly relies on medical image analysis, one of the most common AI applications in health (Chamola et al., 2023) to detect pulmonary abnormalities indicative of TB disease (Naz et al., 2023). Despite the success stories related to X-rays for TB detection, its interpretation is fundamentally subjective and prone to variation among radiologists (Weng et al., 2022). Furthermore, the intricacy of tuberculosis pathology, which includes a wide range of symptoms from mild infiltrates to cavitary lesions, makes accurate and consistent diagnosis using traditional imaging modalities difficult (Ameen et al., 2021; Ifty et al., 2024). XAI is a timely intervention to this challenge, assisting clinicians not only to achieve high-speed diagnostic results but also empowering them with actionable insights into the AI models' predictions (Ifty et al., 2024; Patrício et al., 2023). In other words, by providing interpretable reasons for AI-generated diagnoses, XAI improves transparency and trust in TB diagnosis, closing the gap between algorithmic predictions and clinical decision-making (Gurmessa, & Jimma, 2023) in many ways.

The use of XAI in TB diagnosis by chest X-ray imaging shows promise for many significant advantages.

- XAI approaches can potentially improve diagnostic accuracy by clarifying the essential traits and patterns that drive AI predictions (Maheswari, 2024). This would allow for more informed clinical evaluations and minimise diagnostic mistakes.

- XAI provides physicians with interpretable insights into AI-generated diagnoses (Sheu et al., 2023). It also increases model output confidence, resulting in better acceptance and use of AI-driven diagnostic tools in clinical practice (Larasati, 2023).

- XAI transparency allows doctors to assess and contextualise AI-generated diagnoses within the larger clinical context, allowing them to make evidence-based choices customised to patients' specific requirements (Aranovich, & Matulionyte, 2023). This transparency not only adds confidence to the system's projections but also allows healthcare practitioners to comprehend better and trust the diagnostic results (Patel et al., 2024).

Recent Case Studies on XAI for TB Diagnosis

Recent research shows that a wealthy class of ML and DL algorithms and XAI techniques used in TB diagnostics offer remarkable advantages in interpretability and explanatory power (Gurmessa, & Jimma, 2023). These can be classified based on diverse characteristics and use cases. XAI method is considered model agnostic if it can explain the predictions made by any machine learning (ML) model (Tchuente et al., 2024). Examples of such methods include SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) (Krishnamoorthy et al., 2024). On the other hand, a method is considered model-specific if it can only explain the predictions of a specific ML model (Sheu et al., 2023). An example is the importance of the feature embedded in implementing Random Forest. Local XAI techniques explain a single prediction, while global ones, on the other hand, offer explanations for the entire model (Mi et al., 2024). SHAP and LIME attribute importance scores to individual features explaining model predictions (Kaplan et al., 2024).

Techniques such as Partial Dependence Plots (PDP) and Tree Explainer delve deeper, showing complex correlations between specific features and model predictions (Dosilovic et al., 2018). Anchors and saliency maps further dissect model predictions, displaying input data's most influential features and regions (Adadi, &

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

Berrada, 2018). Gradient-based methods, exemplified by Deep LIFT (Deep Learning Important FeaTures), and Guided Backpropagation, utilise gradient values to determine the overall significance of each feature in the prediction (Chaddad et al., 2023), (Tchuente et al., 2024). Grad-CAM, LIME and SHAP are the most common XAI techniques used in AI-powered pulmonary disease diagnostics (Sheu et al., 2023). Grad-CAM is mainly used for image classification at the same time for semantic segmentation (Bhandari et al., 2022). LIME reduces complicated models locally around particular cases to help explain their decisions (Chamola et al., 2023). SHAP, on the other hand, applies cooperative game theory to determine the relevance of features for individual forecasts (S, & V, 2024). These strategies show significant potential to improve transparency, build trust and make AI models more widely applicable (Gurmessa, & Jimma, 2023). Table 1 shows cases of research advances in the use of XAI techniques for TB diagnosis.

Table 1: Cases of XAI for TB Diagnosis, Techniques and Results

Ref	Aim	AI/XAI techniques	Results
(Ameen et al., 2021)	To classify TB using explainable Residual Network	ResNet-50, Grad-CAM	The model attained an accuracy of 99.5%, a sensitivity of 99%, a specificity of 100%, an F1-score of 99%, a precision of 99%, and an AUC of 99.5%.
(Marvin & Alam, 2022)	To develop a robust and explainable deep transfer learning model for assessing paediatric pulmonary health, including TB, with limited annotated paediatric chest X-ray images.	Deep transfer learning, Layer-Wise Relevance Propagation (LRP)	The model of accuracy of 97% for classifying Normal cases, 97% for COVID-19 cases, 70% for Tuberculosis cases, and 73% for Pneumonia cases. The recall for Pneumonia cases was 100%. The total accuracy of the model was 79%. These results were obtained after just 10 epochs.
(Maheswari, 2024)	To develop and verify a primary CNN architecture with fewer feature layers to provide consistent classification accuracy and strong interpretability.	Deep neural network, CNN, Activation maps (CAM) and (LIME), explainer systems	The receiver operating characteristic (ROC) curve shows classification accuracy, F1-score, sensitivity, and specificity of 0.95. CNN showed a peak area under the curve (AUC) value of 0.976.
(Jegatheeswaran et al., 2024)	To create an explainable Hybrid CNN called Swin-Transformer (Swin-T) designed explicitly for detecting TB in chest X-ray images.	Grad-CAM, hybrid CNN Swin-Transformer (Swin-T)	The model achieved a precision of 82.14%, a recall of 92.00%, a specificity of 82.76%, and an accuracy of 92.22%.
(Ifty et al., 2024)	To investigate a wide array of DL models, such as CNN, hybrid models, ensembles, transformers, and Big Transfer, to classify pulmonary diseases, including TB.	LIME, Grad-CAM, CNN, hybrid models, ensembles, transformers, and Big Transfer	According to the findings, the Xception model, which was fine-tuned using 5-fold cross-validation, achieved the most excellent accuracy of 96.21%.

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

(Dagnaw & Mouthadi, 2023a)	To XAI with a lightweight convolutional neural network (CNN) for improved classification accuracy of TB and Pneumonia and the capacity to provide explanations.	CNN, score-CAM XAI model	The classification accuracy is 97.5%, the precision is 97.4%, the sensitivity is 97.4%, the specificity is 98.6%, and the F1 score is 97.4%.
(Koyyada & Singh, 2023)	To create an explainable AI model to identify and diagnose lung illness in chest X-ray pictures.	DL, CNN-base model, LIME	The local and fusion models have shown great progress in achieving 99.6% accuracy with fewer epochs.
(Patel et al., 2024)	To propose a framework that employs EfficientNet-B4-based Transfer Learning for multi-disease classification in lung X-rays	EfficientNet-B4, Transfer learning, Grad-CAM	The result is a powerful multi-disease (including TB) classification system with 96% accuracy, supported by visualisations that emphasise key locations in X-ray pictures.
(Bhandari et al., 2022)	To present an XAI-based DL architecture for detecting and classifying COVID-19, pneumonia, and tuberculosis using CXR images.	DL, Grad-CAM, SHAP. LIME	During 10-fold cross-validation, the DL model had an average test accuracy of 94.31 ± 1.01% and a validation accuracy of 94.54 ± 1.33%.

Impact of XAI on Trust and Transparency

Many global systems will be prescribed by the use of algorithms, but just like humans, those systems must satisfy many assurances to boost trust (Dosilovic et al., 2018). Trust and transparency are essential aspects of medical diagnostics, too. The integration of AI into healthcare has the potential to increase diagnostic accuracy and efficiency significantly (Dagnaw, & Mouthadi, 2023b). However, AI systems must be clear and intelligible to be broadly accepted and used by healthcare professionals and patients. Explainable AI (XAI) meets this demand by offering clear, understandable insights into how AI models get their findings (Koyyada, & Singh, 2023). This transparency not only adds confidence to the system's projections but also allows healthcare practitioners to better comprehend and trust the diagnostic results (Patel et al., 2024).

Several studies have shown that XAI significantly increases trust in AI-powered medical diagnosis (Ali et al., 2023; Ifty et al., 2024; Mamalakis et al., 2023; Patrício et al., 2023). For example, in the context of tuberculosis (TB) diagnosis, studies have shown that healthcare personnel are more inclined to trust and depend on AI systems that provide clear and understandable explanations of their diagnostic procedures (Dagnaw, & Mouthadi, 2023b; Koyyada, & Singh, 2023, 2023). Visual and textual explanations, such as saliency maps and model-agnostic methods like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive Explanations), have been critical in explaining AI models' decision-making pathways, fostering a higher level of trust (Chamola et al., 2023; Marvin, & Alam, 2022).

Several case studies demonstrate the influence of XAI on improving transparency in TB diagnosis. Two main categories stand out:

- **Saliency Maps and Grad-CAM:** In clinical settings, saliency maps and Gradient-weighted Class Activation Mapping (Grad-CAM) are utilised to highlight areas of chest X-rays that the AI model concentrates on when recognising tuberculosis (TB)(Dagnaw, & Mouthadi, 2023a). Grad-CAM is mainly used for image classification but also

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

semantic segmentation (Bhandari et al., 2022). It works by generating the gradient of a differentiable output, such as a class score, based on a layer's convolutional features. Conversely, Saliency directly calculates input feature significance using gradient squared. Input may be graph nodes, edges, or features. The greater gradient value is assumed to represent the most relevant characteristics (Chaddad et al., 2023). These visual explanations enable radiologists to understand which regions the model believes are symptomatic of tuberculosis (Koyyada, & Singh, 2023), making verifying and accepting the AI's results more straightforward (Dagnaw, & Mouthadi, 2023b).

- **Model-Agnostic Approaches:** LIME and SHAP have been used in TB diagnosis models to explain instances (Bhandari et al., 2022). Local interpretable model-agnostic explanation (LIME) approximates a new simple model using black-box model predictions to interpret the old model. Results are interpreted using the new model (Chaddad et al., 2023). For example, LIME may provide interpretable approximations of the model's behaviour around a given prediction, allowing doctors to understand better the elements that contribute to a positive TB diagnosis in individual instances (Ali et al., 2023). On the other hand, SHAP averages marginal feature values to score model feature influence using Shapley values. Scores from anticipated images reveal each pixel's contribution and help explain categorization. Shapley values are calculated using all possible pulmonary disease characteristics (Bhandari et al., 2022). After pixelating the Shapley values, red pixels improve class prediction, whereas blue pixels decrease it.

This degree of information aids in the communication between sophisticated model projections and clinical decision-making (Raif et al., 2023). These case studies highlight the need to equip healthcare practitioners with tools that make AI decision-making clear and interpretable (Van Der Velden et al., 2022). Such technologies not only build trust but also improve cooperation between AI systems and physicians (Patel et al., 2024).

While XAI improves trust and transparency, it often results in trade-offs between model accuracy and interpretability. Highly complicated AI models like deep neural networks are more accurate but less interpretable. Simpler models like decision trees provide more transparency but may not achieve the same diagnostic accuracy. Balancing these trade-offs is a significant problem in developing XAI-based TB diagnostics. The objective is to strike an ideal equilibrium in which the model is accurate enough for trustworthy diagnosis and interpretable enough to be trusted by healthcare practitioners (Patel et al., 2024). To solve this difficulty, strategies include using hybrid models that incorporate the capabilities of both complex and simple models and continuously improving XAI approaches to increase explanatory power without sacrificing accuracy.

Ethical and Social Implications of AI-Powered TB Diagnosis

A significant proportion of the global population believes that there are general ethical concerns allied to using AI, especially for health. The ever-evolving technological landscape needs to be commensurate with the advancements in dealing with the ethical issues that arise from new developments (Gurmessa, & Jimma, 2023). The use of AI in clinical applications inevitably leads to the limitation of adding bias, the restrictions provided by privacy and security constraints and a need for more transparency and explainability of the networks (Mamalakis et al., 2023). When medical diagnostics are made for incorrect reasons, substantial ethical and policy concerns may be raised (Gurmessa, & Jimma, 2023). This is why investigations of explainability, ambiguity, security, privacy and bias should be essential components of any clinical AI tool research (Mamalakis et al., 2023), including AI-enabled TB diagnosis. Eventually, the trustworthiness of these AI systems can always be determined by the

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

depth of the assessment covering functional, operational, usability, safety, and validation aspects, as well as ethical concerns.

Data security and privacy are among the foremost ethical considerations for TB diagnostics. The use of medical imaging data, including chest X-rays, raises significant concerns regarding patient privacy and confidentiality (Gurmessa, & Jimma, 2023). Safeguarding such sensitive health information is of paramount importance. Therefore, robust data governance frameworks, secure data storage protocols, and stringent access controls are necessary to mitigate the risk of unauthorised disclosure or misuse of patient data. As AI-driven diagnostic tools become increasingly prevalent in clinical practice, it is imperative to address data privacy, bias, accountability, and equitable access to ensure that the benefits of technological advancements are ethically and socially responsible.

There are also social concerns in this case. Equitable access to AI-driven diagnostic tools is fundamental to realising the full potential of XAI in TB diagnosis and addressing global health disparities. Efforts to promote equitable access must prioritise the needs of underserved and marginalised communities, ensuring that diagnostic technologies are affordable, culturally sensitive, and tailored to the unique challenges faced by diverse patient populations. Moreover, mechanisms for benefit sharing and capacity building are essential to ensure that the benefits of technological innovation are equitably distributed and contribute to the advancement of public health globally. XAI's ethical and social implications in TB diagnosis are multifaceted and complex, requiring a holistic and interdisciplinary approach. By addressing concerns related to data privacy, bias, accountability, and equitable access, we can harness XAI's transformative potential to advance TB diagnosis while upholding the highest ethical standards and promoting social justice and equity in healthcare delivery.

Future Directions and Research Opportunities

As seen in Table 1, most of the XAI techniques used are Post hoc. This coincides (Chaddad et al., 2023) survey of XAI techniques in healthcare. We also postulate that they could be the easiest to apply XAI techniques. It is also reported that most medical practitioners trust LIME explanations. However, several questions remain unanswered, which future research endeavours need to critically investigate. Firstly, what is the effect of imaging techniques on explainability? interpretability? Incorporating modern medical imaging tools creates possibilities and problems for XAI. Subsequent research should examine how other imaging modalities, such as CT scans, MRIs, and advanced X-ray technologies, affect the explainability of AI models. Secondly, what is the effect of the mathematical and physical foundations on explanations of the models? This can involve evaluating the resilience and reliability of approaches like saliency maps, Grad-CAM, LIME, and SHAP when used on complex TB images. Thirdly, how do lighting changes affect medical images and explainability approaches? This research aims to investigate the effects of exposure, contrast, and lighting on the interpretability of AI models. Fourthly, what attributes influence the trust among both engineers and clinicians? This can be done by evaluating the perceived dependability, accuracy, and utility of XAI approaches from the viewpoints of both engineers who create these systems and medical practitioners who utilise them. Lastly, how do we establish collaborative efforts from Computer science, medical imaging, mathematics, physics, and healthcare researchers and practitioners? For example, merging experience in medical imaging with AI may result in more accurate and interpretable diagnostic models, while including input from healthcare experts helps guarantee these models suit clinical demands.

CONCLUSION

The overarching purpose of this paper is to highlight the advancement in XAI for tuberculosis diagnosis (TB) and identify the benefits and challenges associated with improved trust in AI-powered TB diagnosis. While XAI is a critical topic in responsible AI discourses, its application in TB diagnosis constitutes a paradigm change with far-reaching consequences. In this work, we show the importance of XAI in improving transparency, interpretability, and trustworthiness in AI-driven TB diagnostic models. We make significant suggestions for deploying XAI-driven TB diagnostic tools, emphasising the need for multidisciplinary cooperation, data quality, ethical concerns, and unique XAI technique applications. Future research and development efforts are required to improve current XAI approaches, investigate new applications, use other XAI approaches and techniques and handle growing obstacles in clinical implementation. Multidisciplinary partnerships among AI researchers, medical practitioners, politicians, and public health specialists are critical for driving development and ensuring the appropriate incorporation of XAI into clinical practice. Furthermore, addressing inequities in access to AI-powered diagnostic technology, especially in impoverished and marginalised populations, is critical to ensuring equitable healthcare delivery and fostering health equality. Adopting transparent, interpretable, and trustworthy AI in TB diagnostics, we may realise XAI's full potential to battle the disease and improve health outcomes for people and communities throughout the globe.

REFERENCES

Acharya, V., Choi, D., Yener, B., & Beamer, G. (2024). Prediction of Tuberculosis From Lung Tissue Images of Diversity Outbred Mice Using Jump Knowledge Based Cell Graph Neural Network. IEEE Access, 12, 17164– 17194. https://doi.org/10.1109/ACCESS.2024.3359989

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

Ahmad, H. K., Milne, M. R., Buchlak, Q. D., Ektas, N., Sanderson, G., Chamtie, H., Karunasena, S., Chiang, J., Holt, X., Tang, C. H. M., Seah, J. C. Y., Bottrell, G., Esmaili, N., Brotchie, P., & Jones, C. (2023). Machine Learning Augmented Interpretation of Chest X-rays: A Systematic Review. Diagnostics, 13(4), 743. https://doi.org/10.3390/diagnostics13040743

Ali, S., Akhlaq, F., Imran, A. S., Kastrati, Z., Daudpota, S. M., & Moosa, M. (2023). The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. Computers in Biology and Medicine, 166, 107555. https://doi.org/10.1016/j.compbiomed.2023.107555

Ameen, Z. S., Saleh Mubarak, A., Altrjman, C., Alturjman, S., & Abdulkadir, R. A. (2021). Explainable Residual Network for Tuberculosis Classification in the IoT Era. 2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT), 9–12. https://doi.org/10.1109/FoNeS-AIoT54873.2021.00012

Aranovich, T. de C., & Matulionyte, R. (2023). Ensuring AI explainability in healthcare: Problems and possible policy solutions. Information & Communications Technology Law, 32(2), 259– 275. https://doi.org/10.1080/13600834.2022.2146395

Bhandari, M., Shahi, T. B., Siku, B., & Neupane, A. (2022). Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI. Computers in Biology and Medicine, 150, 106156. https://doi.org/10.1016/j.compbiomed.2022.106156

Chadaga, K., Prabhu, S., Bhat, V., Sampathila, N., Umakanth, S., Chadaga, R., Shashi Kumar, G. S., & Swathi, K. S. (2023). An explainable multi-class decision support framework to predict COVID-19 prognosis utilizing biomarkers. Cogent Engineering, 10(2), 2272361. https://doi.org/10.1080/23311916.2023.2272361

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. Sensors, 23(2), Article 2. https://doi.org/10.3390/s23020634

Chamola, V., Hassija, V., Sulthana, A. R., Ghosh, D., Dhingra, D., & Sikdar, B. (2023). A Review of Trustworthy and Explainable Artificial Intelligence (XAI). IEEE Access, 11, 78994– 79015. https://doi.org/10.1109/ACCESS.2023.3294569

Codlin, A. J., Vo, L. N. Q., Garg, T., Banu, S., Ahmed, S., John, S., Abdulkarim, S., Muyoyeta, M., Sanjase, N., & Wingfield, T. (2024). Expanding molecular diagnostic coverage for tuberculosis by combining computer-aided chest radiography and sputum specimen pooling: A modeling study from four high burden countries. https://www.researchsquare.com/article/rs-3813705/latest

Dagnaw, G. H., & Mouthadi, M. E. (2023a). Towards Explainable Artificial Intelligence for Pneumonia and Tuberculosis Classification from Chest X-Ray. 2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), 55–60. https://doi.org/10.1109/ICT4DA59526.2023.10302183

Dagnaw, G. H., & Mouthadi, M. E. (2023b). Towards Explainable Artificial Intelligence for Pneumonia and Tuberculosis Classification from Chest X-Ray. 2023 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), 55–60. https://doi.org/10.1109/ICT4DA59526.2023.10302183

De Farias Gabriel, A., Kirschnick, L. B., Só, B. B., Schuch, L. F., Silveira, F. M., Martins, M. A. T., Wagner, V. P., Lopes, M. A., & Martins, M. D. (2023). Oral and maxillofacial tuberculosis: A systematic review. Oral Diseases, 29(7), 2483–2492. https://doi.org/10.1111/odi.14290

Dosilovic, F. K., Brcic, M., & Hlupic, N. (2018). Explainable artificial intelligence: A survey. 2018 41st International Convention on Information and Communication Technology , Electronics and Microelectronics (MIPRO), 0210– 0215. https://doi.org/10.23919/MIPRO.2018.8400040

Gurmessa, D. K., & Jimma, W. (2023). A comprehensive evaluation of explainable Artificial Intelligence techniques in stroke diagnosis: A systematic review. Cogent Engineering, 10(2), 2273088. https://doi.org/10.1080/23311916.2023.2273088

Hansun, S., Argha, A., Liaw, S.-T., Celler, B. G., & Marks, G. B. (2023a). Machine and Deep Learning for Tuberculosis Detection on Chest X-Rays: Systematic Literature Review. Journal of Medical Internet Research, 25, e43154. https://doi.org/10.2196/43154

Hansun, S., Argha, A., Liaw, S.-T., Celler, B. G., & Marks, G. B. (2023b). Machine and Deep Learning for Tuberculosis Detection on Chest X-Rays: Systematic Literature Review. Journal of Medical Internet Research, 25, e43154. https://doi.org/10.2196/43154

Hansun, S., Argha, A., Liaw, S.-T., Celler, B. G., & Marks, G. B. (2023c). Machine and Deep Learning for Tuberculosis Detection on Chest X-Rays: Systematic Literature Review. Journal of Medical Internet Research, 25, e43154. https://doi.org/10.2196/43154

Huang, C., Wang, W., Zhang, X., Wang, S.-H., & Zhang, Y.-D. (2023). Tuberculosis Diagnosis Using Deep Transferred EfficientNet. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 20(5), 2639–2646. https://doi.org/10.1109/TCBB.2022.3199572

Hwang, E. J., Jeong, W. G., David, P.-M., Arentz, M., Ruhwald, M., & Yoon, S. H. (2024). AI for detection of tuberculosis: Implications for global health. Radiology: Artificial Intelligence, 6(2), e230327.

Ifty, T. T., Shafin, S. A., Shahriar, S. M., & Towhid, T. (2024). Explainable Lung Disease Classification from Chest X-Ray Images Utilizing Deep Learning and XAI (arXiv:2404.11428). arXiv. http://arxiv.org/abs/2404.11428

Jegatheeswaran, T., Kumaralingam, L., Selvarajah, S., & Ratnarajah, N. (2024). Explainable Hybrid CNN Swin-Transformer Network for Tuberculosis Diagnosis in Chest X-Rays of Sri Lankan Patients. 2024 4th International Conference on Advanced Research in Computing (ICARC), 85–90. https://doi.org/10.1109/ICARC61713.2024.10499756

Kaplan, S., Uusitalo, H., & Lensu, L. (2024). A unified and practical user-centric framework for explainable artificial intelligence. Knowledge-Based Systems, 283, 111107. https://doi.org/10.1016/j.knosys.2023.111107

Kontsevaya, I., Cabibbe, A. M., Cirillo, D. M., DiNardo, A. R., Frahm, N., Gillespie, S. H., Holtzman, D., Meiwes, L., Petruccioli, E., & Reimann, M. (2023). Update on the diagnosis of tuberculosis. Clinical Microbiology and Infection. https://www.sciencedirect.com/science/article/pii/S1198743X23003403

Koyyada, S. P., & Singh, T. P. (2023). An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images. Healthcare Analytics, 4, 100206. https://doi.org/10.1016/j.health.2023.100206

Krishnamoorthy, T. V., Venkataiah, C., Rao, Y. M., Prasad, D. R., Chowdary, K. U., Jayamma, M., & Sireesha, R. (2024). A novel NASNet model with LIME explanability for lung disease classification. Biomedical Signal Processing and Control, 93, 106114. https://doi.org/10.1016/j.bspc.2024.106114

Kumar, S., Bhagat, V., Sahu, P., Chaube, M. K., Behera, A. K., Guizani, M., Gravina, R., Di Dio, M., Fortino, G., Curry, E., & Alsamhi, S. H. (2024). A novel multimodal framework for early diagnosis and classification of COPD based on CT scan images and multivariate pulmonary respiratory diseases. Computer Methods and Programs in Biomedicine, 243, 107911. https://doi.org/10.1016/j.cmpb.2023.107911

Larasati, R. (2023). Trust and explanation in Artificial Intelligence systems: A healthcare application in disease detection and preliminary diagnosis [PhD Thesis, The Open University]. https://oro.open.ac.uk/88778/

Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Ser, J. D., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., Jiang, R., Khosravi, H., Lecue, F., Malgieri, G., Páez, A., Samek, W., Schneider, J., Speith, T., & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. Information Fusion, 106, 102301. https://doi.org/10.1016/j.inffus.2024.102301

Maheswari, B. U. (2024). Explainable deep-neural-network supported scheme for tuberculosis detection from chest radiographs.

Malik, H., Anees, T., Al-Shamaylehs, A. S., Alharthi, S. Z., Khalil, W., & Akhunzada, A. (2023). Deep Learning-Based Classification of Chest Diseases Using X-rays, CTScans, and Cough Sound Images. Diagnostics (Basel, Switzerland), 13(17), 2772. https://doi.org/10.3390/diagnostics13172772

Mamalakis, M., Dwivedi, K., Sharkey, M., Alabed, S., Kiely, D., & Swift, A. J. (2023). A transparent artificial intelligence framework to assess lung disease in pulmonary hypertension. Scientific Reports, 13(1), 3812. https://doi.org/10.1038/s41598-023-30503-4

Marvin, G., & Alam, Md. G. R. (2022). Explainable Augmented Intelligence and Deep Transfer Learning for Pediatric Pulmonary Health Evaluation. 2022 International Conference on Innovations in

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

Science, Engineering and Technology (ICISET), 272– 277. https://doi.org/10.1109/ICISET54810.2022.9775845

Mera, H. B., Wagnew, F., Akelew, Y., Hibstu, Z., Berihun, S., Tamir, W., Alemu, S., Lamore, Y., Mesganaw, B., & Adugna, A. (2023). Prevalence and Predictors of Pulmonary Tuberculosis among Prison Inmates in Sub-Saharan Africa: A Systematic Review and Meta-Analysis. Tuberculosis Research and Treatment, 2023. https://www.hindawi.com/journals/trt/2023/6226200/

Mi, J.-X., Jiang, X., Luo, L., & Gao, Y. (2024). Toward explainable artificial intelligence: A survey and overview on their intrinsic properties. Neurocomputing, 563, 126919. https://doi.org/10.1016/j.neucom.2023.126919

Miró-Nicolau, M., Moyà-Alcover, G., & Jaume-i-Capó, A. (2022). Evaluating Explainable Artificial Intelligence for X-ray Image Analysis. Applied Sciences, 12(9), 4459. https://doi.org/10.3390/app12094459

Mota Carvalho, T. F., Santos, V. L. A., Silva, J. C. F., Figueredo, L. J. D. A., De Miranda, S. S., Duarte, R. D. O., & Guimarães, F. G. (2023). A systematic review and repeatability study on the use of deep learning for classifying and detecting tuberculosis bacilli in microscopic images. Progress in Biophysics and Molecular Biology, 180– 181, 1– 18. https://doi.org/10.1016/j.pbiomolbio.2023.03.002

Murugesan, S. S., Velu, S., Golec, M., Wu, H., & Gill, S. S. (2024). Neural Networks based Smart e-Health Application for the Prediction of Tuberculosis using Serverless Computing. IEEE Journal of Biomedical and Health Informatics, 1–12. https://doi.org/10.1109/JBHI.2024.3367736

Naz, Z., Khan, M. U. G., Saba, T., Rehman, A., Nobanee, H., & Bahaj, S. A. (2023). An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs. Cancers, 15(1), 314. https://doi.org/10.3390/cancers15010314

Okada, K., Yamada, N., Takayanagi, K., Hiasa, Y., Kitamura, Y., Hoshino, Y., Hirao, S., Yoshiyama, T., Onozaki, I., & Kato, S. (2024). Applicability of artificial intelligence-based computer-aided detection (AI–CAD) for pulmonary tuberculosis to community-based active case finding. Tropical Medicine and Health, 52(1), 2.

Panagoulias, D. P., Virvou, M., & Tsihrintzis, G. A. (2024). A novel framework for artificial intelligence explainability via the Technology Acceptance Model and Rapid Estimate of Adult Literacy in Medicine using machine learning. Expert Systems with Applications, 248, 123375. https://doi.org/10.1016/j.eswa.2024.123375

Patel, A. N., Murugan, R., Srivastava, G., Maddikunta, P. K. R., Yenduri, G., Gadekallu, T. R., & Chengoden, R. (2024). An explainable transfer learning framework for multi-classification of lung diseases in chest X-rays. Alexandria Engineering Journal, 98, 328– 343. https://doi.org/10.1016/j.aej.2024.04.072

Patrício, C., Neves, J. C., & Teixeira, L. F. (2023). Explainable Deep Learning Methods in Medical Image Classification: A Survey (arXiv:2205.04766). arXiv. http://arxiv.org/abs/2205.04766

Raif, P., Suchanek-Raif, R., & Tkacz, E. (2023). Explainable AI (XAI) in Healthcare—Tools and Regulations. 2023 IEEE EMBS Special Topic Conference on Data Science and Engineering in Healthcare, Medicine and Biology, 151– 152. https://doi.org/10.1109/IEEECONF58974.2023.10405096

Sachan, A., Kakadiya, R., Mishra, S., Kumar-M, P., Jena, A., Gupta, P., Sebastian, S., Deepak, P., & Sharma, V. (2024). Artificial intelligence for discrimination of Crohn’s disease and gastrointestinal tuberculosis: A systematic review. Journal of

Article DOI: https://doi.org/10.37284/eajit.7.1.2276

Gastroenterology and Hepatology, 39(3), 422–430. https://doi.org/10.1111/jgh.16430

Sheu, R.-K., Pardeshi, M. S., Pai, K.-C., Chen, L.-C., Wu, C.-L., & Chen, W.-C. (2023). Interpretable Classification of Pneumonia Infection Using eXplainable AI (XAI-ICP). IEEE Access, 11, 28896– 28919. https://doi.org/10.1109/ACCESS.2023.3255403

Tang, N., Yuan, M., Chen, Z., Ma, J., Sun, R., Yang, Y., He, Q., Guo, X., Hu, S., & Zhou, J. (2023). Machine Learning Prediction Model of Tuberculosis Incidence Based on Meteorological Factors and Air Pollutants. International Journal of Environmental Research and Public Health, 20(5), 3910. https://doi.org/10.3390/ijerph20053910

Tchuente, D., Lonlac, J., & Kamsu-Foguem, B. (2024). A methodological and theoretical framework for implementing explainable artificial intelligence (XAI) in business applications. Computers in Industry, 155, 104044. https://doi.org/10.1016/j.compind.2023.104044

V, J. P., & S, A. A. V. (2024). A multi-aspect framework for explainable sentiment analysis. Pattern Recognition Letters, 178, 122– 129. https://doi.org/10.1016/j.patrec.2024.01.001

Van Der Velden, B. H. M., Kuijf, H. J., Gilhuijs, K. G. A., & Viergever, M. A. (2022). Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Medical Image Analysis, 79, 102470. https://doi.org/10.1016/j.media.2022.102470

Weng, F., Meng, Y., Lu, F., Wang, Y., Wang, W., Xu, L., Cheng, D., & Zhu, J. (2022). Differentiation of intestinal tuberculosis and Crohn’s disease through an explainable machine learning method. Scientific Reports, 12(1), 1714. https://doi.org/10.1038/s41598-022-05571-7

Yadav, P. (2023). Challenges & Solutions for Recent Advancements in Multi-Drugs Resistance Tuberculosis: A Review. Microbiology Insights, 16, 117863612311524. https://doi.org/10.1177/11786361231152438