Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 150~164
e-ISSN: 2548-6861

Explainable Transformer and Machine Learning Models in Predicting Tuberculosis Treatment Outcomes. A Systematic Review

Shumirai S Sibanda [1]*, Belinda Ndlovu [2]*
* Informatics and Analytics Department, National University of Science and Technology, Bulawayo, Zimbabwe
n02218035d@students.nust.ac.zw [1], belinda.ndlovu@nust.ac.zw [2]

Article Info

Article history:

Received 2025-11-24
Revised 2026-01-01
Accepted 2026-01-20

Keywords:

Tuberculosis (TB),
Treatment Outcomes,
Transformer Models,
Explainable AI,
Machine Learning.

ABSTRACT

Tuberculosis (TB) remains a major health challenge, and predicting treatment outcomes continues to be difficult in real-world settings. Recent advances in Artificial Intelligence (AI), particularly transformer-based models, have shown promise in modelling longitudinal, multimodal, and heterogeneous TB data. However, their clinical adoption is constrained by limited interpretability, fairness concerns, and deployment challenges. This study presents a systematic literature review of explainable transformer and machine learning models used for TB prognosis. Following PRISMA guidelines, searches across ACM, IEEE Xplore, PubMed, and ScienceDirect identified 17 peer-reviewed studies published between 2020 and 2025 that met the inclusion criteria. The review synthesises evidence on predictive performance, explainability techniques, and deployment considerations. Findings indicate that transformer-based and deep learning models generally outperform conventional machine learning approaches on longitudinal and multimodal data. In contrast, traditional models remain competitive for tabular clinical datasets. Explainability approaches are dominated by feature importance methods and SHAP, with limited use of intrinsic transformer interpretability mechanisms. Persistent challenges include data scarcity, limited generalisability, computational overhead, insufficient evaluation of fairness, and weak alignment with real-world TB care workflows. Building on these findings, the study proposes the Explainable Transformer Adoption Model for TB Prognosis (ETAMTB) as a conceptual clinical adoption framework integrating multimodal transformers, explainability layers, clinician-facing interfaces, and deployment enablers. Overall, the review concludes that effective AI adoption in TB care requires balancing predictive performance, interpretability, and equity, and that explainable transformers should currently be viewed as promising but largely experimental tools rather than deployment-ready solutions.

I. INTRODUCTION

As the leading infectious disease killer after COVID-19, tuberculosis (TB) continues to be a serious global health concern, accounting for more deaths than HIV/AIDS[1][2]. TB, which is caused by Mycobacterium tuberculosis, mainly affects the lungs but can also affect other organs like the brain or spine.[1], [3]. When an infected individual coughs or sneezes, airborne particles are released into the air, causing transmission[4].TB can also be latent and asymptomatic, while other cases can be active, presenting symptoms such as fever, night sweats, weakness, and loss of appetite[5].

Drug-resistant strains (DR-TB) are also a big concern when dealing with TB because they have been linked to more complex diagnoses, lengthy and toxic therapies, and increased death rates[1], [9]. Moreover, TB is one of the contributing factors to deaths associated with HIV infections worldwide because comorbid conditions such as HIV can greatly enhance the progression of TB [7],[11].

One of the most crucial factors to measure the effectiveness of therapeutic efficacy in this case is to make a correct prediction about treatment outcomes, which can merely be classified into success or failure.[8], [12]. The emergence of data-driven approaches such as machine learning (ML) provides unique opportunities to investigate correlations between patient information, biomarkers, and socioeconomic variables and outcomes of prognosis for this particular disease[13]. Causes of treatment failure have been identified as including income, family size, patient knowledge, and the quality of care received. In addition to these factors, crowding and inadequate infection control within medical facilities contribute to disease transmission and poor outcomes [14]–[16].

Artificial Intelligence (AI) refers to computer systems capable of performing tasks that normally require human intelligence. Machine Learning (ML) is a subset of AI; which involves algorithms that learn from data to make predictions or classifications [17][18]. Machine Learning algorithms have been extensively researched for predicting treatment outcomes in various diseases [19], [20]. Predicting treatment outcomes using ML allows clinicians to identify potential treatment failure or relapse risks and the occurrence of reactions, allowing adjustments to treatment plans maximising treatment success rates [21]. Research on the application of ML algorithms for predicting TB treatment outcomes is limited compared to the research focused on TB diagnosis and spread [22]. Early studies used traditional algorithms such as logistic regression and random forests [23]. However, data has become more complex and multimodal, ML models, especially transformer architecture, have shown superior performance [13], [24], [25].

Since their introduction in the seminal 2017 paper "Attention is All You Need" [26], transformer models have revolutionised AI through their innovative self-attention mechanisms. This architecture, which includes integrated encoder-decoder components and a scalable structure, enables the models to contextualise input data and generate sophisticated predictions [26], [27]. Transformers have achieved tremendous success in healthcare regarding the analysis of complex data types, such as genomic sequences, sequential patient records, and medical images [26], [28]. Their self-attention mechanism provides a key structural advantage over recurrent architectures like Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) for modelling long-range dependencies in longitudinal data, as it processes sequences in parallel rather than sequentially and mitigates issues of vanishing gradients [27], [28]. This makes them particularly suited for the temporal and multimodal fusion tasks prevalent in TB prognosis. However, most of the challenges exist in the way of their clinical adoption. These models usually function like black boxes that provide accurate predictions without providing a clear explanation, which in turn depletes clinical confidence [7], [26]. Furthermore, transformers that have been trained on data from high-income settings often exhibit poor generalisability in low-income areas [28], [29]. However, their high hardware and cloud-computing requirements pose further obstacles to implementation in resource-constrained areas where the burden of tuberculosis is highest. [15], [27].

This systematic review represents a novel and imperative advance in these research avenues by focusing not only on diagnosis but also on prognostic capabilities through the application of explainable transformers. Although other systematic reviews have been conducted by other researchers, like [30] have extensively described the TB diagnosis application via artificial intelligence models in TB treatment and management, their focus remained broad and operative solely in terms of diagnosis and screening capabilities. They have not only restricted their discussions to comprehensive utilisation but have also often aggregated broad deep learning models like CNN without distinguishing newer architectures. Similarly, other systematic reviews, such as [31], have explored therapeutic efficacy but remained broad and general in their descriptions and encompassed adverse drug reactions and drug resistance without examining the architectural specifics of these adverse effects. Conversely, this systematic review for the first time specifically identifies and assesses these newly developed transformer models like Decoder Transformer (DT-THRE) and their effectiveness in TB treatment, in accurately forecasting patient consequences like treatment failure, recurrence and mortality. In addition to the aspects discussed above, this systematic study addresses critical gaps noted in other landmark studies, such as [32], which analysed prediction models developed prior to the widespread adoption of transformers and relied heavily on traditional statistical methods, such as logistic regression. This study deviates from these conventional approaches by synthesising and identifying these newly developed models, like explainable transformers and assessing their effectiveness to newly direct attention towards TB treatment forecasting methods, like treatment failure. In addition to the specifics discussed above, this systematic study further explores new areas related to newly developed methods, such as multimodal data fusion, synthesising how modern transformers integrate unstructured clinical and sociodemographic data with structured electronic health records to enhance predictive performance. Furthermore, the study moves beyond the superficial treatment of black-box limitations in previous systematic literature reviews by rigorously assessing architectural explainability. This review critically evaluates inherent interpretability mechanisms such as attention weights that are central to fostering clinical trust and autonomy in high-stakes decision-making, rather than merely noting the use of generic tools like SHAP. The primary contribution of this article is a systematic synthesis of explainable AI models and ML models for predicting TB treatment outcomes complemented by the ETAMTB framework as a synthesis-derived conceptual guide for clinical adoption rather than an empirically validated system. Unlike prior work, it uniquely connects predictive accuracy,

clinical interpretability, and adoption challenges in low-resource, high-burden settings.

Despite the widespread applications of AI in TB management, there remains a gap in effectively utilising diagnostic success to inform prognostic actions. Presently, systematic reviews have primarily focused on reporting the aggregate performance of general approaches to diagnosis and diagnosis via broad deep learning models, but have not necessarily assessed newer models, such as transformers, in depth. Reports on approaches that utilise intrinsic methods of explanation, rather than general attention heat maps derived directly from these complex models, to promote clinicians' trust and enable immediate decision-making have not yet been explored in depth. Additionally, this review will address this gap by focusing solely on the role of explainable transformer models, specifically those designed for prognostic functionality and trustworthiness in TB treatment. To this end, the review is guided by the following research questions:

1. How do transformer-based models compare to traditional machine learning algorithms in predicting TB treatment outcomes
2. What explainability techniques are integrated into transformer-based models to enhance clinical interpretability in the context of TB treatment outcomes prediction.
3. What are the challenges faced in deployment and adaptability of transformer models for TB treatment outcome prediction.

To address these questions, this paper is structured as follows: Section II describes the review methodology, where the search strategy, eligibility criteria, study selection, quality assessment and inclusion and exclusion criteria used to obtain relevant studies are specified. Section III shows the results of the review and the discussions, implications, limitations and direction for future work of the study. Section IV concludes the review.

II. METHODS

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta Analysis (PRISMA). The PRISMA consists of a flow diagram divided into four parts: identification, screening, eligibility, and included.

A. Search Strategy

After the research questions were formulated, keywords that are relevant to the research were identified. The keywords were used to formulate search queries to identify articles relevant to this study. A search was conducted on the 22nd of September 2025, through the search papers were obtained from four main databases: ACM, IEEE Xplore, PubMed and ScienceDirect. The keywords were used as follows ("tuberculosis" OR "TB" OR "mycobacterium tuberculosis") AND ("treatment outcome" OR "treatment success" OR "treatment failure" OR "prognosis" OR "therapy response") AND ("transformer model" OR "BERT" OR "attention-based model" OR "encoder-decoder") AND ("explainable AI" OR "XAI" OR "model interpretability" OR "explainability" OR "attention visualisation" OR "transparency") AND ("machine learning" OR "artificial intelligence" OR "AI" OR "predictive model") AND (2020:2025[dp]) AND (English[lang]). A total of 205 articles were retrieved from the four databases: ACM (n = 33), PubMed (n = 46), IEEE Xplore (n = 63) and ScienceDirect (n = 63). These included conference papers, editorials, abstracts, preprints, peer-reviewed papers, empirical papers and reviews.

B. Screening

All retrieved articles were imported into Mendeley Reference Manager. Duplicates were automatically identified and removed. During title and abstract screening, studies were evaluated for the following indicators: presence of tuberculosis or TB treatment outcome in the title or abstract, evidence of AI, ML or Transformer model use and mention of explainability, interpretability or transparency in the methodology. Articles that did not meet these criteria were excluded. Out of 205 initial records, 113 duplicates were removed, leaving 92 papers. After title and abstract screening, 57 papers were excluded, leaving 35 for full-text review. Eighteen were excluded at this stage, resulting in 17 studies included for final analysis.

TABLE 1
INCLUSION & EXCLUSION CRITERIA

Inclusion	Exclusion
Studies predicting TB treatment outcomes using AI, ML or transformer models	Studies focused on TB diagnosis or detection
Integration of explainable AI (XAI) or interpretability	Studies without explainability or interpretability components
Peer-reviewed full-text journal, conference papers (2020-2025)	Preprints, Editorials, Book chapters and abstract only
Studies addressing clinical deployment challenges	Studies not addressing clinical application or deployment challenges
Studies written in English	Studies not written in English

C. Eligibility Criteria

The review focused on peer-reviewed empirical studies published between 2020 and 2025 that investigated TB treatment outcomes using transformer-based or other explainable AI models. Studies were considered eligible if they: (1) predicted definitive TB treatment outcomes, (2) applied a transformer architecture or integrated an explainable AI (XAI) technique with any ML model; (3) utilised relevant clinical data modalities (4) reported quantitative performance metrics or qualitative interpretability insights. Studies were

excluded if they focused only on TB diagnosis or detection, addressed other diseases unrelated to TB and were non-peer-reviewed materials.

D. Included

A total of 92 studies progressed to title and abstract screening after removal of duplicates. From these, 35 articles were selected for full-text review based on their relevance to tuberculosis treatment outcome prediction and the use of artificial intelligence methods. Following a detailed assessment, 18 studies were excluded for reasons such as focusing exclusively on TB diagnosis, lacking an explainability component, or not employing transformer-based or comparable predictive models. Consequently, 17 studies met all inclusion criteria and were included in the final qualitative synthesis of this systematic review.

E. Data Extraction and Synthesis

Studies were downloaded and retrieved from the web and saved into a file named SLR Reference 1. The folder was imported into Mendeley Reference Manager, where the duplicates were then removed. After the eligibility screening process was complete, the remaining studies were then saved in another folder named SLR references 2. An Excel sheet was developed by SS to extract relevant data from studies. The Excel sheet had columns that included author, year, study design, data modality, transformer or ML or AI model used, explainability method and relevance to research questions. These columns were used to extract key data from studies. The author SS conducted the initial data extraction and BN verified all extractions. Full-text papers were independently assessed by both reviewers against the predefined inclusion and exclusion criteria, leaving the final included studies for full analysis, which were now moved and saved in the folder FINAL SLR References.

F. Quality Assessment

The methodological quality, credibility, and relevance of the 17 included primary studies were assessed using the Critical Appraisal Skills Programme (CASP) checklist [31]. The CASP tool allows for a systematic assessment of research evidence by taking into consideration such aspects as the validity of the study, its methodological rigour, and the applicability of its findings [32]. As a result, each study was rated according to the criteria of clear research aims and objectives, methodological appropriateness, study design, data collection, data analysis, research ethics, and the research question addressed. Studies were considered low quality if they had severe methodological issues and incomplete descriptions; moderate quality if they met most criteria but had minimal limitations; and high quality if they exhibited clear aims, sophisticated or robust methodology, appropriate validation, and extensive reporting. Additionally, three studies employed multimodal data, whereas most used retrospective data. Lastly, the research involved numerous methodologies due to the differences in the research objectives of the studies.Approaches to model validation that enhance reliability were described in 11 articles. In terms of interpretability, four studies utilised formal explainability techniques, while six studies provided simple feature-importance analyses. Ethical considerations were well reported in nine studies, while six others mentioned them in passing or made no reference. Thus, the studies ranged in quality, with seven receiving a high-quality rating due to comprehensive methodology and validation, four rated moderate quality, and six categorised as low quality due to significant reporting or methodological gaps. Appendix A presents the adapted CASP quality assessment of the study whilst Appendix B presents the quality assessment score criteria.

III. RESULTS AND DISCUSSION

The systematic search and selection process, detailed in the PRISMA flow diagram. Given the relatively limited number of included studies (n=17) and their methodological heterogeneity, of varying methods, this synthesis identifies trends rather than firm conclusions as it presents their characteristics and findings.

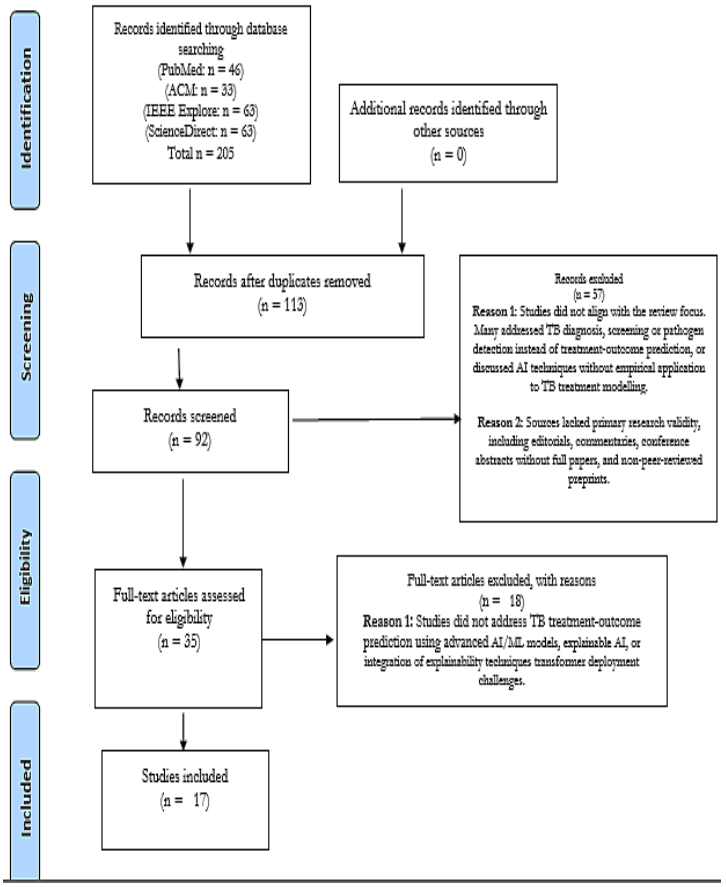

Figure 1: PRISMA Screening Result

TABLE 2
RESULTS TABLE

Author/ Year	Country	Model Type/ Name	Explainability Technique used	Performance Metrics	Key Predictors/ Data Modalities	Challenges/ Limitations	Deployment & Adaptation Context	Key Findings
[33]	China	Radiomic models Deep Learning (Gradient Boosting, Small Deep Learning Model	• Feature importance via radiomics • Fusion interpretability	AUC (0.764-0.867), internal/external validation	• Longitudinal CT scans • Demographic data • Clinical data	• Small dataset • High computational cost	Hospital level imaging integration	Fused radiomics, deep learning improved prediction of DR-TB outcomes early in treatment
[34]	D.R. Congo	CNN vs. classical ML (SVM, KNN, RF, Decision Tree)	• Implicit visual feature learning	Accuracy 94%, AUC 93%, Sensitivity 88%, F1-score 91.3%	• Clinical data • Demographic data	• Limited dataset • Lack of real world testing • Model interpretability	Academic evaluation, low cost implementable in Congolese clinics	CNN outperformed traditional models; suitable for early TB screening.
[35]	China	LASSO-Cox regression (clinical prognostic model)	• Coefficient Interpretation (transparent statistical model)	AUC 0.766 (train), 0.796 (validation)	• Blood • Biochemical markers	• Only Chinese cohorts • Manual data entry	Clinical use for prognosis prediction in hospital TB management	Clinical indicator based risk score effectively predicted TB treatment outcomes
[36]	Brazil/ USA	Logistic regression	• Coefficients • Nomogram (interpretable)	C statistic 0.77 :bootstrap validation	• HIV status, Hypertension • Drug use • Age • Education level	• Limited by missing data • No external validation	Web based point of care tool for TB prognosis	Simple clinical model predicted unsuccessful TB outcomes with good discrimination
[37]	Colombia	Machine learning (Random Forest, Natural Language Processing models, Data Fusion)	• Feature importance (clinician validated)	Sensitivity 73 %	• EMR text • Clinical data	• Limited structured EMR data	Designed for low resource, multi source diagnostic, prognosis detection system	AI models, especially clinical data driven, outperform traditional diagnostics, prognosis detection.
[30]	India	AI for TB diagnosis & treatment	• CNN heatmaps, saliency	Narrative synthesis	• Radiology genomics • Clinical data	• Limited explainability • Data bias • Ethical barriers	AI adoption in Indian clinics	AI can revolutionise TB detection and prognosis but ethical deployment is essential

[38]	Israel	Transformer Explainability model	• Deep Taylor • Decomposition based relevance propagation	Qualitative visual heat maps	• Vision text transformer features	• Complex implementation	Foundation model for visual explainability	Introduces explainability method for transformer beyond attention visualisation
[39]	China	AI models (CNN, Random Forest)	• SHAP • Feature Importance	Comparative narrative	• Clinical data • Imaging • Genomic • Treatment data	• Publication bias	General recommendation for AI use in therapy monitoring	AI models improve monitoring of Pulmonary TB treatment efficacy and drug resistance
[15]	South Africa	Logistic Regression	• Model Coefficients (interpretable)	Accuracy 64% Recall 95% F1 score 76%	• Comorbidities (HIV, obesity, hypertension)	• Limited sample size	Public health and primary care integration	Comorbidities strongly affect DR-TB treatment outcomes, integrated care is vital
[40]	Canada	DT-THRE (Decoder Transformer for Temporal Health Data	• Temporal attention embedding visualisation	Accuracy 78.5% Baseline 40.5%	• Sequential EHR data	• Model complexity	Prototype for decision support in disease prediction and prognosis	Incorporating temporal encoding significantly improves outcomes prediction accuracy
[11]	China	XGBoost Random Forest Boruta feature selection	• SHAP (Shapley Additive Explanations)	AUC = 0.928 (test set)	• resistance type, • Activated Partial Thromboplastin Time, Thrombin Time, Platelet Distribution Width, Prothrombin Time • clinical & CT data	• Limited external validation; single centre data	hospital Electronic Medical Record (EMR) data to predict treatment outcomes or risks in patients who have both tuberculosis (TB) and diabetes mellitus (DM).	XGBoost model with SHAP improved interpretability and early detection of treatment failure among TB-DM patients.
[41]	India	Decision Tree, Random Forest, SVM, Naïve Bayes	• Feature weight visualisation (implicit)	AUC = 0.909 Accuracy = 92.7%	• Clinical data	• Regional generalisability • Interpretability limited	Indian Randomised Controlled Trial (RCT) to forecast when a TB patient's sputum culture will turn negative during treatment.	Decision Tree outperformed others, showing high precision and recall; ML viable for clinical TB monitoring.
[23]	Malaysia	XGBoost (with hyperparameter tuning), Logistic Regression,	• Feature ranking (XGBoost gain)	Accuracy = 68.1%	• Demographic data • Clinical data	• Small sample • Single year dataset	Applicable to Penang State TB registry systems	Hyperparameter-tuned XGBoost yielded best accuracy; highlighted ML potential for

		Decision Tree comparisons						regional public health TB surveillance.
[42]	Moldova / USA	Neural Network, Random Forest, Logistic Regression	• Model feature importance (AUC based)	OC AUC = 0.87	• Demographic data • Clinical data • District level FLQ resistance data	• No external validation • Small dataset	Supports rapid empiric treatment guidance in low resource settings	Neural Network effectively predicted fluoroquinolone resistance in RR-TB using routine surveillance data.
[14]	India	AI driven multi model approach (ML + DL ensemble)	• Interpretivist AI framework	Accuracy = 87.5% Sensitivity = 88.2%	• Clinical records • Laboratory • Imaging features	• Ethical data privacy concerns • Limited real world testing	Academic proof of concept for AI decision support in TB care	AI models outperformed traditional diagnostics; emphasised ethical integration and personalised care potential
[22]	Malaysia / Brazil	Multinomial Naïve Bayes SMOTE for class imbalance	• Model transparency via probabilistic output	Accuracy	• Lab • Demographic data (Brazilian SINAN databases)	• Imbalanced classes • Data representativeness	Resource allocation support in public TB programmes	Naïve Bayes and SMOTE enhanced TB outcome prediction in imbalanced datasets; useful for targeted follow-up.
[43]	USA / Tanzania / Bangladesh / Siberia	Logistic Regression, Random Forest, XGBoost (stratified by regimen)	• Feature importance ranking per regimen	F1 score = 0.766; 0.667; 0.787 (regimen specific)	• Demographic data • BMI • Drug regimen • Comorbidities	• Sparse longitudinal data • Limited follow up	Research collaboration tool for MDR-TB across LMICs	Stratified XGBoost improved interpretability and performance; BMI a key predictor for TB recovery.

A. *Publication Trends*

The publication trend graph shows how research on explainable AI and ML has changed over time, in relation to tuberculosis treatment outcomes. This helps distinguish periods of greater scholarly activity and relate them to scientific vigour.

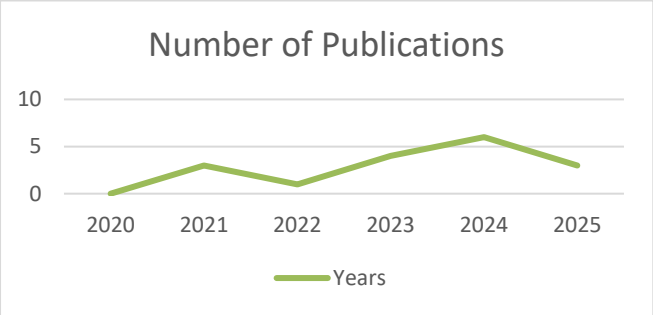

Figure 2: Publication Trends

The annual publication trends are depicted in Figure 2, which reflects a sharp increase in publication activity beginning in 2021 to a maximum publication rate of six papers per year during 2024. This is no surprise because there has been a resurgence of interest in analytics pertaining to TB across the globe due to care disruptions arising out of the COVID-19 pandemic. The nominal decrease inaccurately depicts a decrease in publication interest indicated to occur during 2025 and is most likely reflective of the natural evolution of this field into more niche domains like building Health Prediction Systems and Hyperparameter Optimisation Algorithms.

B. *Study Origin*

The knowledge about regional distribution helps to identify capacity growth trends regarding AI solutions within TB care, highlighting novelty produced due to high-volume regions.

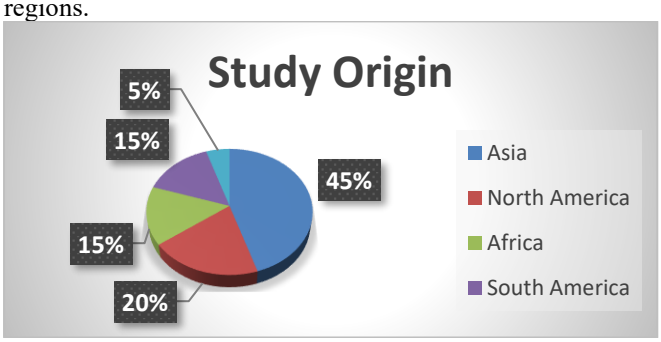

Figure 3:Study Origin

The research activity is identified in Figure 3, which gives a breakdown of study origins of the research studies included. The contribution of Asia is notable, with China, India, and Malaysia combining for almost half (45%) of the total research studies. The high prevalence rate of tuberculosis within this region, together with increased support for research on artificial intelligence, is not surprising. North America comprised 20% due to methodological excellence. Europe accounted for 5% of research studies, conducted through collaborative research partnerships in TB surveillance and modelling. South America and Africa, which are both high-prevalence regions, accounted for a surprisingly low 15% combined.

C. *Algorithms Found*

The studies included assessed and compared varied ML algorithms for predicting treatment outcomes, progression, and resistance in TB treatment. The algorithm, authors, and results are presented in TABLE 3.

TABLE 3
ALGORITHMS FOUND

Algorithm	Author(s)	Efficiency & Performance
Decision Tree	[23], [41]	Accuracy **92.72%**, AUC **0.909**, precision 95.9%
Random Forest	[34], [39], [41]–[43],[37]	AUC > 0.80
XGBoost (baseline, regimen stratified, tuned)	[11], [23], [43]	Accuracy **66.3%**, F1 scores 0.667–0.787 depending on regimen, Accuracy 68.1% (best in study)
Logistic Regression	[15], [34], [41], [42]	63.3% baseline; moderate performance
Neural Networks (CNN)	[42]	AUC **0.87** predicting FQ resistance

A comparative synthesis of key performance metrics across model categories reveals a nuanced picture. For tabular data, traditional ensemble methods like Random Forest and XGBoost consistently achieved high AUCs (0.80-0.93) and accuracy [11], [23], [41]. Transformer-based models (DT-THRE) demonstrated superior performance (Accuracy: 78.5%) on complex, sequential EHR data where they could leverage temporal attention, significantly outperforming baseline models [40]. Deep learning models (CNNs) applied to imaging data also showed high predictive value (AUC 0.76-0.87) [33]. Logistic Regression provided a strong, interpretable baseline but generally yielded lower discriminative performance (Accuracy ~64%, C-statistic ~0.77) [15], [36]. This supports a context-dependent model selection strategy. Direct quantitative comparison across all models was not feasible due to heterogeneity in datasets, outcome definitions, and validation strategies; therefore, performance trends are interpreted comparatively rather than as absolute superiority claims

D. Challenges

There are a number of persistent challenges to the application of these AI models in actual clinical settings. The main issues found are summarised in TABLE 4.

TABLE 4
CHALLENGES

Challenge	Author(s)	Description	Impact
Data Limitations	[33], [34], [36][14], [42], [43]	Small dataset size, missing data, sparse longitudinal data, and imbalanced classes.	Compromises model generalisability, increases overfitting risk, and reduces clinical reliability.
Limited Generalisability	[11], [35], [36], [41], [42]	Models trained on single centre or specific national cohorts (e.g., only Chinese patients).	Poor performance when applied to new populations with different demographic or clinical characteristics.
Computational Resources	[33], [34]	High computational cost of deep learning and transformer models.	Barriers deployment in resource constrained clinics common in high TB burden regions.
Interpretability Gaps	[30], [34], [41]	Complex models acting as black boxes or using implicit feature visualisation without formal XAI.	Hinders clinical trust and adoption, as clinicians cannot verify the model's reasoning.
Ethical & Privacy Concerns	[14], [30]	Data privacy issues and potential for algorithmic bias in model predictions.	Raises barriers to data sharing and necessitates rigorous ethical frameworks for deployment.

The synthesis revealed a number of recurrent issues, which are listed in Table 4. Data restrictions, which included small sample sizes, missing data, and class imbalance, were the most significant obstacle, mentioned in more than one-third of the research (6/17). Concerns about limited generalisability (5/17), when models trained on certain national cohorts, for example, China, Brazil, demonstrated ambiguous performance in other populations, immediately followed this. Significant barriers to real-world clinical application were also repeatedly identified, especially in LMICs, which include interpretability limitations, computational resource needs, and ethical and privacy issues.

E. Explainability Approaches Used

Figure 4 visualises the range of explainability methods applied across the studies included in this review. As transparency and clinician trust are essential for TB treatment decision support, analysing which techniques are used and how frequently they highlight the maturity and direction of explainable AI in this domain.

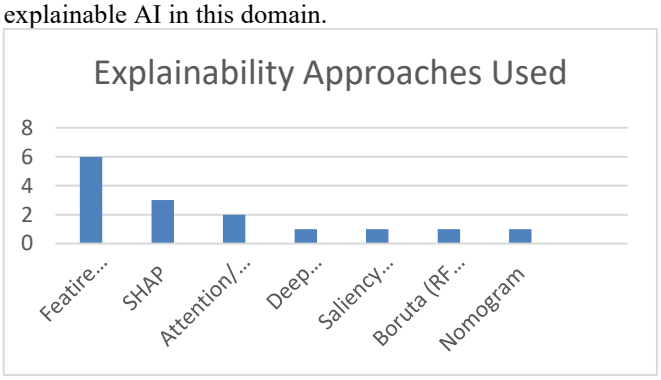

Figure 4: Explainability Approaches

There is a definite preference for explainability strategies that strike a balance between clinical intuitiveness and computational efficiency, according to the analysis. The most popular methods, used in six out of the seventeen research (35%), were feature importance and model coefficients. With three studies (18%), SHAP was the next most popular method, indicating a shift toward more reliable, instance-level explanations. More complex approaches like Layer wise Relevance Propagation (LRP) (n = 1) and attention visualisations (n = 2) were relatively rare and typically limited to studies involving transformer architectures and complex multimodal data. This distribution results from a sense of discipline to facilitate clinically useful, interpretable results as more sophisticated models increasingly adopt advanced concepts within XAI.

F. Discussion

Compared with prior SLRs, which focus either on TB diagnostics and CNN-based imaging [29] or conventional ML models without considerations of interpretability [20], this review advances the field by assessing transformer architectures through a three-dimensional lens, which includes comparative performance, explainability for clinical trust and feasibility of deployment in resource-constrained TB health systems. No existing review combines these dimensions, so this paper represents a novel sociotechnical perspective on explainable transformers for TB prognosis.

RQ1: Performance Comparison: Transformer-Based Models vs Traditional Machine Learning Algorithms in TB Treatment Outcome Prediction

For structured, cross-sectional clinical data, traditional ML algorithms, especially Random Forest, being the most popular algorithm having been employed in six studies and XGBoost have proved outstanding performance with tremendous efficacy [23], [41]. Some major advantages reinforce their ongoing relevance, namely computational effectiveness, reduced training data volume requirements, and some inherent interpretability via feature importance metrics [19], [39], [44]. These attributes make them a more viable alternative in many real-world clinical environments with restricted resources, in which IT support could be limited, especially in areas with few data [15], [28]. For example, simplicity was highlighted in the logistic regression model proposed by [36], [45], demonstrating its utility for clinical prediction with a C-statistic of 0.77.

Nevertheless, it is crucial to note, from the analysis, that there is a definite and powerful advantage to transformers and other deep learning architectures in scenarios that demand modelling rich, longitudinal or highly multimodal data [26], [28]. What stands out about the performance of the DT-THRE model from [40], which significantly demonstrates the transformer's core strength, is its capacity to model complex temporal dependencies in patient records, which static models simply cannot model. It achieved a remarkable accuracy of 78.5% on sequential EHR data, representing a substantial improvement over the benchmark model's 40.5%. It is further validated by studies such as [33], illustrating movements in performance by architectures being specially modified for dynamic high-dimensional data, for which a deep learning model on longitudinal CT scan data improved AUC value to 0.764-0.867. Another validation is [34], illustrating improved performance for a CNN in healthcare, outperforming traditional ML, reaching an accuracy rate of 94% on its clinical predictive tasks, demonstrating improved performance in healthcare ecosystems with increasing multimodal, temporarily dynamic data [24], [25], for which there will be substantiation in their importance.

These findings indicate that in model selection, there is a need for a complex paradigm. Traditional algorithms are effective and efficient for structured, tabular data. In contrast, transformer models, with their parallel self-attention architecture, excel at processing sequential or multimodal data by capturing long range dependencies [27], [28]. Transformers require large datasets and high computational power, often unavailable in high burden settings, making traditional ML models more practical for static predictions in such environments [46].

1. *Random Forest*: Random Forest (RF) is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputting the mode of the classes of the individual trees [47]. It introduces randomness through bagging, which is bootstrap aggregating and random feature selection, which combats overfitting and enhances generalisability [19]. Random Forest was the most frequently employed algorithm across the reviewed studies. Its ensemble structure demonstrated strong performance with AUCs consistently above 0.80, which was attributed to its robustness against overfitting and capacity to handle nonlinear relationships in clinical data [19], [39]. However, despite providing global feature importance scores, its intrinsic lack of transparency for individual predictions presents a significant limitation for clinical deployment [41].

2. *XGBoost*: XGBoost (eXtreme Gradient Boosting) is a sophisticated and efficient implementation of gradient boosting[23]. It sequentially builds an ensemble of trees, with each new tree designed to correct the errors of the previous ones, using a gradient-based optimisation process. XGBoost was found to be remarkably proficient at structured clinical datasets owing to its inherent error-corrective regularisation technologies [23]. XGBoost performed remarkably well in complex scenarios, with AUC values amounting to 0.928 for treatment failure predictions. [11], [23]. Yet, it requires SHAP analysis for output interpretation due to its inherent complexity in clinical contexts [11], [43]

3. *Logistic Regression*: A fundamental mode in statistics for binary classification, logistic regression employs a logistic function to calculate predictions based on probability [42]. It is a linear model, mapping a transformation of inputs into their linear combination, followed by passing it through the sigmoid function [15]. Some works employed logistic regression as a simple model for comparison due to its interpretability, simplicity and accuracy. Though being constrained to assume linearity, it reduces its capacity for complex pattern identification in comparison with other algorithms; its coefficient interpretability is unambiguous [42], [43].

4. *Convolutional Neural Networks (CNNs):* Convolutional Neural Networks (CNNs) are a subclass of deep neural networks that are most frequently used for visual imagery analysis. They are essential for radiology image analysis because they employ convolutional layers to automatically and adaptively learn spatial hierarchies of features from input images [3]. In the context of TB, [34] used a CNN-based Supervised Deep Learning model (SDLM) to examine longitudinal CT scans and predict treatment outcomes in DR-TB patients [48]. Moreover, [34] also validated the supremacy of CNN over conventional ML for anticipating clinical information [34]. As far as the capability to distil the complex patterns implicit in high-dimensional information is concerned, the CNNs currently have no equal. Their limitations are being a black box model, requiring greatly intense computational power and necessitating large amounts of annotated information during the learning phase, which

can be a rather severe limitation in the vast majority of resource-limited environments [26].

RQ2: Explainability Techniques for Clinical Interpretability in Transformer-Based TB Treatment Outcome Prediction.

A core aim of explainability is to build clinical trust and promote AI adoption in TB care [49]. However, while widely assumed, direct empirical evidence that XAI enhances clinician trust or improves decision making in TB prognosis remains scarce. The studies reviewed show a range of approaches, with a definite preference for those offering intuitive, practically useful insights. Explainability is progressing beyond simple attention visualisation for transformer-based models, which are intrinsically complex [38]. The author [38] proposed a novel framework that incorporates Deep Taylor Decomposition and Layer wise Relevance Propagation (LRP) to generate more robust and faithful explanations than attention maps alone, which is crucial for verifying model reasoning in high-stakes clinical predictions [38].

However, feature significance approaches continue to be the most popular across all model types because they are obvious for doctors who are used to evaluating risk variables. The author [11] introduced a game theory based technique called SHAP (SHapley Additive exPlanations) to a XGBoost model, which revealed important indicators such as certain blood coagulation markers for treatment failure in TB Diabetes patients [11]. Similarly, [35] used the coefficients from a LASSO Cox regression model to construct a transparent, clinical indicator-based risk score, which is inherently interpretable [39]. For imaging based models, such as the radiomics and deep learning fusion model by [33], feature importance via radiomics provided insights into which imaging biomarkers drove the predictions [33]. Explainable AI methods such as SHAP (SHapley Additive exPlanations) improve interpretability by providing insights into "why" a model makes a particular prediction [50],[51]. The trend indicates that while advanced XAI for transformers is emerging, the field currently relies heavily on model agnostic techniques like SHAP and intrinsic model interpretability to bridge the transparency gap. Attention visualisations are intuitive but non-causal and open to misinterpretation [38], while methods like Layer-wise Relevance Propagation (LRP) remain too computationally complex for routine clinical use and SHAP can be computationally expensive and may produce unstable explanations with correlated features. This underscores the urgent need for robust, clinically validated explanation methods that are both faithful to the model and actionable for practitioners.

1) *SHAP (SHapley Additive exPlanations) :* SHAP is a unified approach to interpreting the output of any ML model based on Shapley values from cooperative game theory [39]. It works by computing the marginal contribution of each feature to the prediction outcome across all possible combinations of features, assigning each an importance value for a specific prediction [39]. SHAP's strong theoretical reinforcements, capacity to offer both local and global individual prediction and overall model behaviour interpretability are its main advantages [38], [52]. Its major deficiency is the high computational cost that makes it slow in real-time clinical applications, especially for models with many characteristics or complicated ensembles [28].

2) *Feature Importance and Model Coefficients:* This method identifies the contribution of each input variable to the model's performance based on the model's internal characteristics, which can be the weights of a linear model or the feature importance of a tree-based model [35]. In the case of linear models such as Logistic Regression and LASSO regression, the magnitude and sign of the model's coefficients provide direct information about the contribution of each input variable [37]. The main advantage of this method is its interpretability without requiring a deep understanding of AI concepts. However, its disadvantage can be the case when the input variables are correlated in the model and it fails to account for why a specific prediction was made [43].

3) *Attention Visualisation:* Attention visualisation is a technique unique to attention-based models such as transformers [29]. It visualises attention weights to reveal which input sequence elements the model mainly focused on while making a certain prediction. Based on this, the author [40] applied the temporal attention embedding visualisation for their DT-THRE model to clearly describe how the model emphasises certain points in time in sequential health data [40]. The strength of this method is that it gives a straightforward intuitive look into the model's internal decision process and aligns well with sequential data. According to [38], attention weights can be difficult to interpret in models with multiple attention layers, and they are not always correct explanations for the model's decision making process. High attention does not always equate to causal importance [38].

4) *Layer wise Relevance Propagation (LRP) :* LRP is a technique for explaining the predictions of deep neural networks by redistributing the prediction output backwards through the network's layers to the input, assigning a relevance score to each input feature [34]. It works by using a set of propagation rules to trace the contribution of each neuron back to the input. The key strength of LRP is its ability to generate detailed, pixel-wise or feature-wise explanations for complex deep learning models, and its weakness lies in its complexity and computational intensity, requiring specialised expertise to implement and interpret correctly, which can be a significant barrier in routine clinical practice [28].

RQ3: Challenges in Deployment and Adaptability of Transformer Models for TB Treatment Outcome Prediction

In this review, fairness is understood as the absence of systematic performance disparities across clinically relevant subgroups, including gender, HIV status, socioeconomic context, and geographical setting, consistent with group-based and distributional fairness perspectives in healthcare AI. A significant obstacle to successful AI that remains is high-quality representative data. Notable flaws in the data, such as small dataset sizes [34], [42], the presence of missing data points and imbalanced class distributions [22], [43], have been noticed in many studies. These naturally reduce the robustness of the model and result in overfitting [34], [48]. A condition called dataset shift arises when ML models trained on specific populations demonstrate substantial drops in performance on new demographical or clinical environments, exhibiting the challenge presented by data [53]. This is especially apparent in studies being performed on particular national cohorts [35], thereby raising several questions regarding the fairness associated with using such models on various scenarios in different settings of healthcare [43].

The resource demands of deep AI create a deployment paradox in low-resource, high-burden settings. Key barriers include unreliable internet for cloud inference, absent digitised EHRs, scarce technical support, and variable staff capacity for complex dashboards. Furthermore, the challenge of fairness, ensuring models do not exhibit biased performance across patient subgroups (gender, ethnicity, or HIV status), is rarely addressed. None of the reviewed studies conducted formal fairness audits, a critical omission for equitable deployment [43].

However, difficulties also arise on the human side of adoption. This is especially true, even if highly effective models are developed and since there is a lack of interpretability originating from their black box approach, this undermines trust among clinicians to use predicted outcomes within practice [34]. Lack of trust is further worsened by legitimate concerns for ethics regarding privacy and security and also for fairness regarding algorithmic bias to vulnerable populations [14], [30]. Absence of strong governance structures and properly defined ethical frameworks further restricts the development of trust essential for clinical adoption[43]

The application of transformer AI into high-burden settings like LMICs brings unique ethical deployment challenges. The most critical issue is fairness: models trained on data from High Income Countries (HICs) are often inaccurate when applied to the local population of TB patients, whose clinical and socioeconomic profiles differ significantly from those in HICs [43]. This lack of local validation may result in biased or unequal care and calls for mandatory local fairness audits. Secondly, privacy and integrity of sensitive patient data are threatened by the model's reliance on multimodal data fusion, which can be overcome with the adoption of federated learning that securely enables collaborative model training without centralising sensitive patient data[37]. Finally, to guarantee accountability, the complexity of transformer logic needs to be made auditable using inherently robust XAI methods so that local clinicians confidently use the system[38].

A coordinated approach that recognises their interdependence is required to address these problems. Developing lean implementations of frugal AI to improve computationally efficiency [54], as well as partnerships for diverse data acquisition tasks, are examples of structural solutions to be combined with different solutions like data augmentation [48], synthetic minority oversampling [22], and domain adaptation techniques. Above all, explanations should not remain an addition but be fundamental principles [38], [40], and at the same time, developed ethics and governance frameworks ensure patient safety and equity against risks imposed by advancements in technology [27].

G. Conceptual Framework

Synthesising the identified challenges and requirements, we propose the Explainable Transformer Adoption Model for TB care (ETAMTB) as a conceptual framework for integration. It is crucial to clarify that ETAMTB is not an empirically validated tool but a synthesis-derived roadmap outlining the necessary components and workflow for responsible development and deployment. The framework suggests a structured clinical course through which transformer-based AI can be integrated into the decision-making process of tuberculosis treatment. It begins with the capture of multimodal TB data.

These further undergo pre-processing and TB-specific feature engineering to combine these heterogeneous data sources for model ingestion. At the heart of this framework lies a transformer-based prognosis model that influences attention mechanisms in learning clinical dependencies within the dataset for predicting treatment outcomes, severity progressions or response likelihoods. ETAM TB goes ahead to emphasise that such predictions should not be black box outputs but rather a feed into a dedicated explainability layer using SHAP values, attention weighting, and feature relevance visualisation to produce clinically interpretable reasoning behind model predictions. Such explanations are channelled through to a clinician-facing prognosis dashboard designed to support risk stratification and inform the need for treatment adjustment, hence empowering health workers rather than replacing clinical judgment.

Finally, the framework introduces operational enablers like federated learning are implemented for preserving privacy, model distillation for low-resource deployment, and governance mechanisms addressing ethics, fairness, and regulatory compliance. All these put together position ETAM TB as a practical, trustworthy, and context sensitive outline on how explainable transformers can be used in real world TB care.

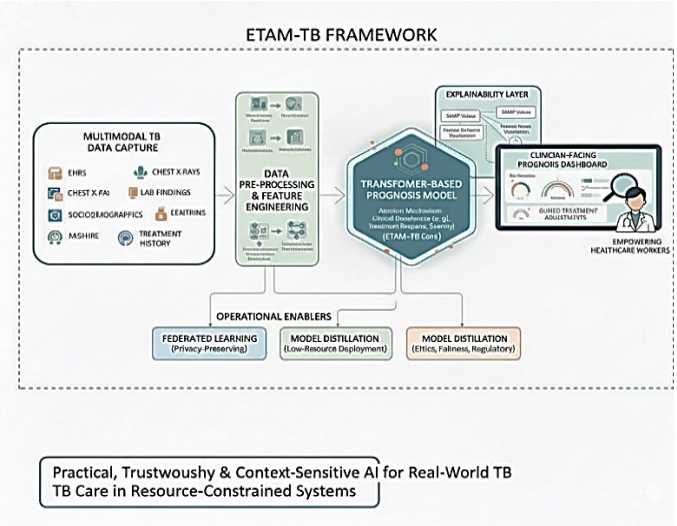

Figure 5: ETAM TB Framework

H. *Implications of the Study*

The conclusions of this review also have key implications for practice and future studies; the development of context-sensitive AI pathways should address all stated criteria and, therefore, take first priority for health systems at this point in time. It is clearly essential that researchers have cost-effective transformer architecture designs suited to settings where tuberculosis is still predominant.

1) *Practical Implications:* Results highlight that the strategy of application for AI needs to be context-specific. Even though resource-intensive models perform equally well on longitudinal data for well-resourced hospitals, their application to primary care clinics in high-burden areas is far from reality [33], [40]. Low-cost, lightweight AI tools using public data, tailored to the specific constraints of the low-resource environment, are therefore urgently needed [35], [39]. Furthermore, explainability faces a significant gap in the operational integration of its implementation. Techniques like SHAP and nomograms are technologically feasible [11], [36], but their clinical efficacy depends on partnership with clinicians in order to co-design the AI interface to make explanations trustworthy and therapeutically actionable to overcome the black box uncertainty of doctors [30].

   a) *Practical Recommendations:* Clinical systems should integrate XAI dashboards like attention relevance maps and SHAP to validate the clinician side. The Ministries of Health should develop federated data frameworks for TB that may improve transformer generalisability across regions. Hospitals from LMICs should consider the use of lightweight, distilled transformers consistent with offline execution on more limited computing infrastructure. TB programmes should mandate external validation and gender sensitive fairness testing before fielding AI models.

2) *Theoretical Implications:* From a theoretical standpoint, the analysis advocates a paradigm shift from optimising particular models to designing AI for equality and generalisation instead. Poor generalisability remains prevalent [11], [35], which poses a crucial limitation: locally trained models tend to break down in new populations. These indicators highlight the urgent need for federated learning strategies and novel domain adaptation capable of learning trustworthy representations from different international TB data without jeopardising patients' privacy. Furthermore, the area urges digging deeper into theoretical explanations for explanations themselves and developing reputable instruments to determine whether an explanation genuinely improves clinician comprehension and decision making or it is just a technical output. The high computational cost of advanced models, given the small, vague datasets typical of many TB programs, presents an important research direction towards frugal AI by model distillation and efficient architectural search[28].

3) *Policy Implications:* These findings have direct policy implications. The World Health Organisation and national TB programmes should establish data management frameworks that protect patient privacy while enabling data sharing for AI development. Policymakers and funders must support the creation of large, representative national datasets to monitor model performance and prevent disparities. Crucially, policy must mandate fairness evaluations and transparency, requiring AI developers to report and address performance gaps across key subgroups before deployment.

I. *Limitations of the Study*

Despite the very rigorous approach adopted for this systematic review, several limitations exist which should also be appreciated: while focusing on four large databases and English language publications, this may have excluded significant studies published in local journals and non-English publications, thus possibly also infusing geographical bias. Additionally, putting into focus on publications after 2020 while ensuring currency, leaving out earlier foundational work, lack of homogeneity in reporting among selected studies prevented direct algorithmic comparisons and meta-analysis, though dual reviewer techniques have minimised subjectivity, there was interpretive judgement in this rapidly evolving field. The synthesis of only 17 studies limits robust quantitative comparisons and alongside a geographic concentration in Asia, reduces applicability to other high burden regions.

J. Future Work

Future research should establish an integrated sociotechnical paradigm, starting with the basic work required. This calls for the constitution of international consortia in order to create large-scale, multimodal data sets from genomic, imaging, and socioeconomic elements representing diversity originating from the high burden regions. Architectural innovation should then focus on the creation of domain-specific lightweight transformers using knowledge distillation, precisely benchmarked against frugal ML/AI models for their capability to assure vigorous and efficient offline operation and the ability to suggest clinically tailored interventions. Regarding explainability, future work should focus on co-designing clinically verifiable XAI dashboards with clinicians, ensuring that the visual interface showing SHAP values and attention heatmaps is effortlessly integrated into Electronic Medical Records (EMRs). In turn, this can make static predictions dynamic clinical partners. Lastly, on clinical integration and ethics, future research should ensure deployability and equity in LMICs by adopting federated learning and domain adaptation. To this end, well-structured ethical and policy frameworks should be established to ensure uniform reporting and a comprehensive multicentre validation process within national TB programs.

IV. CONCLUSION

This review finds that while traditional models handle structured TB data well, transformers better manage complex, longitudinal datasets. However, adoption is hindered by limited interpretability, unaddressed fairness, high computational costs, and poor generalisability. Accuracy alone is insufficient for clinical use; current explainability methods like SHAP need more robust, context-aware evaluation. The bias toward studies from high-resource settings raises equity concerns for high-burden regions. The proposed ETAMTB framework serves as a conceptual roadmap to bridge these gaps. Ultimately, explainable transformers show promise but remain experimental, requiring diverse data, explicit fairness audits, efficient design, and stakeholder collaboration to become deployable, equitable tools in real-world care.

REFERENCES

[1] J. Chakaya *et al.*, "Global Tuberculosis Report 2020 – Reflections on the Global TB burden, treatment and prevention efforts," *Int. J. Infect. Dis.*, vol. 113, pp. S7–S12, Dec. 2021, doi: 10.1016/j.ijid.2021.02.107.

[2] *Global Tuberculosis Report 2023*. World Health Organization, 2023.

[3] S. I. Nafisah and G. Muhammad, "Tuberculosis detection in chest radiograph using convolutional neural network architecture and explainable artificial intelligence," *Neural Comput. Appl.*, vol. 36, no. 1, pp. 111–131, Jan. 2024, doi: 10.1007/S00521-022-07258-6/METRICS.

[4] "Global tuberculosis report 2024 World Health Organization," Nov. 2024.

[5] Y. Luo *et al.*, "Development of diagnostic algorithm using machine learning for distinguishing between active tuberculosis and latent tuberculosis infection," *BMC Infect. Dis.*, vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12879-022-07954-7.

[6] I. Shah, V. Poojari, and H. Meshram, "Multi-Drug Resistant and Extensively-Drug Resistant Tuberculosis," *Indian J. Pediatr.*, vol. 87, no. 10, pp. 833–839, Oct. 2020, doi: 10.1007/S12098-020-03230-1/METRICS.

[7] M. C. Hosu, L. M. Faye, and T. Apalata, "Predicting Treatment Outcomes in Patients with Drug-Resistant Tuberculosis and Human Immunodeficiency Virus Coinfection, Using Supervised Machine Learning Algorithm," *Pathogens*, vol. 13, no. 11, Nov. 2024, doi: 10.3390/pathogens13110923.

[8] L. S. Peetluk *et al.*, "A Clinical Prediction Model for Unsuccessful Pulmonary Tuberculosis Treatment Outcomes," *Clin. Infect. Dis.*, vol. 74, no. 6, pp. 973–982, Mar. 2022, doi: 10.1093/cid/ciab598.

[9] A. Gupta, V. Kumar, S. Natarajan, and R. Singla, "Adverse drug reactions & drug interactions in MDR-TB patients," *Indian J. Tuberc.*, vol. 67, no. 4, pp. S69–S78, Dec. 2020, doi: 10.1016/J.IJTB.2020.09.027.

[10] V. A. Dartois and E. J. Rubin, "Anti-tuberculosis treatment strategies and drug development: challenges and priorities," *Nat. Rev. Microbiol.*, vol. 20, no. 11, pp. 685–701, Nov. 2022, doi: 10.1038/S41579-022-00731-Y;SUBJMETA.

[11] A. Z. Peng *et al.*, "Explainable machine learning for early predicting treatment failure risk among patients with TB-diabetes comorbidity," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-57446-8.

[12] L. Pantaleon, "Why measuring outcomes is important in health care," *J. Vet. Intern. Med.*, vol. 33, no. 2, pp. 356–362, Mar. 2019, doi: 10.1111/jvim.15458.

[13] A. Sambarey *et al.*, "Integrative analysis of multimodal patient data identifies personalized predictors of tuberculosis treatment prognosis," *iScience*, vol. 27, no. 2, Feb. 2024, doi: 10.1016/j.isci.2024.109025.

[14] S. K. Pandey, K. U. Singh, R. S. Dingankar, K. Jadhav, K. Gupta, and R. K. Yadav, "Prediction of Tuberculosis Disease Progression with AI Analysis of Clinical Data," 2023, doi: 10.1109/ICAIIHI57871.2023.10489091.

[15] M. C. Hosu, L. M. Faye, and T. Apalata, "Comorbidities and Treatment Outcomes in Patients Diagnosed with Drug-Resistant Tuberculosis in Rural Eastern Cape Province, South Africa," *Diseases*, vol. 12, no. 11, Nov. 2024, doi: 10.3390/diseases12110296.

[16] M. C. Hosu, L. M. Faye, and and T. Apalata, "Predicting Treatment Outcomes in Patients with Drug-resistant Tuberculosis and HIV Co-infection Using a Supervised Machine Learning Algorithm." Sep. 2024, doi: 10.20944/preprints202409.1747.v1.

[17] O. A. Hussain and K. N. Junejo, "Predicting treatment outcome of drug-susceptible tuberculosis patients using machine-learning models," *Informatics Heal. Soc. Care*, vol. 44, no. 2, pp. 135–151, Apr. 2019, doi: 10.1080/17538157.2018.1433676.

[18] S. Hadebe, B. Ndlovu, and K. Maguraushe, "Managing Diabetes Using Machine Learning and Digital Twins," *Indones. J. Innov. Appl. Sci.*, vol. 5, no. 2, pp. 145–162, 2025, doi: 10.47540/ijias.v5i2.1981.

[19] Y. Wang, Y. Xu, Z. Yang, X. Liu, and Q. Dai, "Using Recursive Feature Selection with Random Forest to Improve Protein Structural Class Prediction for Low-Similarity Sequences," *Comput. Math. Methods Med.*, vol. 2021, 2021, doi: 10.1155/2021/5529389.

[20] Y. Yang, L. Xu, L. Sun, P. Zhang, and S. S. Farid, "Machine learning application in personalised lung cancer recurrence and survivability prediction," *Comput. Struct. Biotechnol. J.*, vol. 20, pp. 1811–1820, Jan. 2022, doi: 10.1016/j.csbj.2022.03.035.

[21] M. Asad, A. Mahmood, and M. Usman, "A machine learning-based framework for Predicting Treatment Failure in tuberculosis: A case study of six countries," *Tuberculosis*, vol. 123, p. 101944, Jul. 2020, doi: 10.1016/J.TUBE.2020.101944.

[22] W. L. W. Foh, S. L. Ang, C. Y. Lim, A. A. L. Alaga, and G. H. Yeap, "Prediction of Tuberculosis Patients' Treatment Outcomes Using Multinomial Naive Bayes Algorithm and Class-Imbalanced Data," 2023, doi: 10.1109/GlobConET56651.2023.10150132.

[23] Y. S. Zakaria, N. A. Ariffin, A. Ahmad, R. Rainis, A. M. Muslim, and W. M. M. Wan Ibrahim, "Optimizing Tuberculosis Treatment Predictions: A Comparative Study of XGBoost with Hyperparameter in Penang, Malaysia," *Sains Malaysiana*, vol. 54, no. 1, pp. 3741–3752, Jan. 2025, doi: 10.17576/jsm-2025-5401-22.

[24] B. Nansamba, J. Nakatumba-Nabende, A. Katumba, and D. P. Kateete, "A Systematic Review on Application of Multimodal Learning and Explainable AI in Tuberculosis Detection," *IEEE Access*, vol. 13. Institute of Electrical and Electronics Engineers Inc., pp. 62198–62221, 2025, doi: 10.1109/ACCESS.2025.3558878.

[25] A. Sambarey *et al.*, "Integrative analysis of multimodal patient data identifies efficacious drug regimens and personalized predictors of tuberculosis treatment prognosis."

[26] S. Madan, M. Lentzen, J. Brandt, D. Rueckert, M. Hofmann-Apitius, and H. Fröhlich, "Transformer models in biomedicine," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1. BioMed Central Ltd, Dec. 2024, doi: 10.1186/s12911-024-02600-5.

[27] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," *AI Open*, vol. 3, pp. 111–132, Jan. 2022, doi: 10.1016/J.AIOPEN.2022.10.001.

[28] S. Nerella *et al.*, "Transformers in Healthcare: A Survey."

[29] A. Vaswani *et al.*, "Attention Is All You Need," 2017.

[30] S. Yadav, N. Jeyaraman, M. Jeyaraman, and G. Rawal, "Artificial intelligence in tuberculosis diagnosis: Revolutionizing detection and treatment," *IP Indian J. Immunol. Respir. Med.*, vol. 9, no. 2, pp. 85–87, 2024, doi: 10.18231/j.ijirm.2024.017.

[31] "CASP Checklists - Critical Appraisal Skills Programme." .

[32] S. Santhanam, V. Ravindran, and C. Wincup, "Critical appraisal of an original research article," *J. R. Coll. Physicians Edinb.*, Sep. 2025, doi: 10.1177/14782715251369964.

[33] M. Nijiati *et al.*, "Deep learning and radiomics of longitudinal CT scans for early prediction of tuberculosis treatment outcomes," *Eur. J. Radiol.*, vol. 169, Dec. 2023, doi: 10.1016/j.ejrad.2023.111180.

[34] G. Mate Landry, R. Nsimba Malumba, F. C. Balanganayi Kabutakapua, and B. Boluma Mangata, "PERFORMANCE COMPARISON OF CLASSICAL ALGORITHMS AND DEEP NEURAL NETWORKS FOR TUBERCULOSIS PREDICTION," *J. Techno Nusa Mandiri*, vol. 21, no. 2, pp. 126–133, Sep. 2024, doi: 10.33480/techno.v21i2.5609.

[35] M. Zhan *et al.*, "A clinical indicator-based prognostic model predicting treatment outcomes of pulmonary tuberculosis: a prospective cohort study," *BMC Infect. Dis.*, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12879-023-08053-x.

[36] L. S. Peetluk *et al.*, "A clinical prediction model for unsuccessful pulmonary tuberculosis treatment outcomes Sterling 4* , and the Regional Prospective Observational Research in Tuberculosis (RePORT)-Brazil network," doi: 10.1093/cid/ciab598/6313211.

[37] A. D. Orjuela-Cañón, A. F. Romero-Gómez, A. L. Jutinico, C. E. Awad, E. Vergara, and M. A. Palencia, "Data Fusion of Medical Records and Clinical Data to Enhance Tuberculosis Diagnosis in Resource-Limited Settings," *Appl. Sci.*, vol. 15, no. 10, May 2025, doi: 10.3390/app15105423.

[38] H. Chefer, S. Gur, and L. Wolf, "Transformer Interpretability Beyond Attention Visualization."

[39] F. Zhang, F. Zhang, L. Li, and Y. Pang, "Clinical utilization of artificial intelligence in predicting therapeutic efficacy in pulmonary tuberculosis," *Journal of Infection and Public Health*, vol. 17, no. 4. Elsevier Ltd, pp. 632–641, Apr. 2024, doi: 10.1016/j.jiph.2024.02.012.

[40] O. Boursalie, R. Samavi, T. E. Doyle, and O. Boursalie, "Decoder Transformer for Temporally-Embedded Health Outcome Predictions," in *Proceedings - 20th IEEE International Conference on Machine Learning and Applications, ICMLA 2021*, 2021, pp. 1461–1467, doi: 10.1109/ICMLA52953.2021.00235.

[41] S. A. Fayaz *et al.*, "Machine learning algorithms to predict treatment success for patients with pulmonary tuberculosis," *PLoS One*, vol. 19, no. 10 October, Oct. 2024, doi: 10.1371/journal.pone.0309151.

[42] S. You *et al.*, "Predicting resistance to fluoroquinolones among patients with rifampicin-resistant tuberculosis using machine learning methods," *PLOS Digit. Heal.*, vol. 1, no. 6 June, Jun. 2022, doi: 10.1371/journal.pdig.0000059.

[43] L. Shichman *et al.*, "Predictive Modeling and Machine Learning Insights into Multi-Drug Resistant Tuberculosis Treatment Outcomes," in *2025 IEEE Systems and Information Engineering Design Symposium, SIEDS 2025*, 2025, pp. 185–190, doi: 10.1109/SIEDS65500.2025.11021160.

[44] S. Chishakwe, N. Moyo, B. M. Ndlovu, and S. Dube, "Intrusion Detection System for IoT environments using Machine Learning Techniques," *2022 1st Zimbabwe Conf. Inf. Commun. Technol. ZCICT 2022*, pp. 1–7, 2022, doi: 10.1109/ZCICT55726.2022.10045992.

[45] B. Ndlovu, F. J. Kiwa, M. Muduva, and C. T. Chipfumbu, "Developing a Logistic Regression Machine Learning Model that Predicts Viral Load Outcomes for Children Living with HIV in Gutu District , Zimbabwe," *Indones. J. Innov. Appl. Sci.*, pp. 277–304, 2025, doi: 10.47540/ijias.v5i3.2275.

[46] H. Chamboko and B. Ndlovu, "Twitter (X) Sentiment Analysis on Monkeypox : A Systematic Literature Review," vol. 14, no. 2, pp. 629–639, 2025, doi: 10.14421/ijid.2025.5196.

[47] N. W.C Mukura and B. Ndlovu, "Performance Evaluation of Artificial Intelligence in Decision Support System for Heart Disease Risk Prediction," no. Who 2018, pp. 83–93, 2023, doi: 10.46254/ap04.20230043.

[48] M. Nijiati *et al.*, "Deep learning on longitudinal CT scans: automated prediction of treatment outcomes in hospitalized tuberculosis patients," *iScience*, vol. 26, no. 11, Nov. 2023, doi: 10.1016/j.isci.2023.108326.

[49] A. K. Mondal, A. Bhattacharjee, P. Singla, and A. P. Prathosh, "XViTCOS: Explainable Vision Transformer Based COVID-19 Screening Using Radiography," *IEEE J. Transl. Eng. Heal. Med.*, vol. 10, 2022, doi: 10.1109/JTEHM.2021.3134096.

[50] B. Ndlovu, K. Maguraushe, and O. Mabikwa, "A Comparative Analysis of Machine Learning Techniques and Explainable AI on Voice Biomarkers for Effective Parkinson's Disease Prediction," *J. Inf. Syst. Informatics*, vol. 7, no. 3, pp. 2196–2228, Sep. 2025, doi: 10.51519/JOURNALISI.V7I3.1172.

[51] T. Ngwazi and B. Ndlovu, "Early Detection of Diabetic Retinopathy Through Explainable AI Models : A Systematic Review," vol. 14, no. 2, pp. 616–628, 2025, doi: 10.14421/ijid.2025.5200.

[52] B. Ndlovu, K. Maguraushe, and O. Mabikwa, "Machine Learning and Explainable AI for Parkinson's Disease Prediction: A Systematic Review," *Indones. J. Comput. Sci.*, vol. 14, no. 2, 2025, doi: https://doi.org/10.33022/ijcs.v14i2.4837.

[53] S. Sathitratanacheewin, P. Sunanta, and K. Pongpirul, "Deep learning for automated classification of tuberculosis-related chest X-Ray: dataset distribution shift limits diagnostic performance generalizability," *Heliyon*, vol. 6, no. 8, Aug. 2020, doi: 10.1016/j.heliyon.2020.e04614.

[54] D. Saraswat *et al.*, "Explainable AI for Healthcare 5.0: Opportunities and Challenges," *IEEE Access*, vol. 10. Institute of Electrical and Electronics Engineers Inc., pp. 84486–84517, 2022, doi: 10.1109/ACCESS.2022.3197671.