International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Multimodal Deep Learning Framework for Tuberculosis Antibiotic Resistance Prediction Using Chest X-Rays and Genomic Features

Suchethana H C, B.E., M.Tech., Assistant Professor,
Ullas S, Shrivatsa BS, Madhumitha E, Anushree HD

Department of Information Science and Engineering, Jawaharlal Nehru New College of Engineering, Shivamogga, Karnataka, India – 577201.

***

Abstract - Drug-resistant tuberculosis (TB) complicates treatment planning and prolongs transmission, while conventional drug susceptibility testing (DST) needs specialised laboratories and considerable time. This paper presents TB-OS V2.0, a multimodal research prototype that estimates the resistance of *Mycobacterium tuberculosis* to five antibiotics, namely rifampicin (RIF), isoniazid (INH), ethambutol (EMB), streptomycin (STR) and pyrazinamide (PZA), from a chest radiograph and a 36-feature genomic profile. A ResNet-18 network analyses the radiograph, an XGBoost model analyses the genomic vector, and an antibiotic-specific logistic-regression meta-learner combines the logit-transformed probabilities of the two branches through late fusion. In total, fifteen frozen models (three tiers for each antibiotic) are trained on antibiotic-specific cohorts from the TB Portals data with strict patient-level splits and zero patient overlap. On held-out test sets, the genomic branch reached ROC-AUC values between 0.9245 and 0.9739, whereas the chest X-ray branch reached 0.6376 to 0.7589. Late fusion produced small ROC-AUC gains for INH, EMB, STR and PZA (+0.0012 to +0.0108) but fell below the genomic model for RIF (0.9481 versus 0.9623). A Flask back end and a React dashboard expose the five independent probabilities, with a mean warm-request latency of 538.07 ms on a CPU-only setup. The results show that multimodal fusion is technically feasible but does not automatically outperform the strongest single modality. The system is a research prototype and not a clinically validated diagnostic tool.

Keywords: tuberculosis; antibiotic resistance; multimodal learning; late fusion; chest X-ray; genomic features

1. INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by *Mycobacterium tuberculosis*, a bacterium that mainly affects the lungs but can also involve other organs of the human body. Despite decades of progress in medicine, TB remains a significant public health challenge because of continued transmission, the complexity of diagnosis, and the emergence of drug-resistant strains [1, 2]. Successful treatment depends on identifying the disease correctly and on knowing whether the infecting bacteria are susceptible or resistant to the drugs that make up the treatment regimen.

Antibiotic resistance is one of the major obstacles in TB management. When *M. tuberculosis* develops resistance to one or more drugs, the effectiveness of standard regimens may fall, and delayed or inaccurate identification of resistance complicates treatment planning and increases the risk of onward transmission. Conventional approaches to resistance detection include phenotypic drug susceptibility testing and molecular diagnostic methods. These approaches are valuable, but they may need specialised laboratory facilities, trained personnel and additional processing time [3, 4]. These limitations motivate research into computational approaches that can analyse existing data and support resistance prediction.

Artificial intelligence (AI), and in particular machine learning (ML) and deep learning (DL), has opened new opportunities for analysing complex medical data [5]. Deep networks learn useful patterns directly from large image collections, while ML algorithms capture relationships among numerical and categorical features. In TB research, chest radiographs and genomic information are two different but potentially complementary sources of evidence. A chest X-ray (CXR) contains visual patterns associated with lung abnormalities, whereas genomic data describe bacterial lineage and mutations that may be associated with antibiotic resistance. A large body of work has examined genomic prediction of TB resistance [1, 3, 6, 7] and image-based analysis of bacterial susceptibility [4, 8, 9], yet combining a radiograph with a genomic profile for several drugs in one working system has received much less attention [10, 11].

This paper investigates that combination. We propose TB-OS V2.0, a multimodal framework that takes a chest X-ray and the corresponding processed genomic features and returns five *independent* resistance probabilities, one for each of RIF, INH, EMB, STR and PZA. Each antibiotic is treated as its own binary prediction task, so the five probabilities are marginal estimates that need not sum to one, and a sample may receive high probabilities for more than one drug. The image branch uses a ResNet-18 convolutional network [12], the genomic branch uses gradient-boosted trees (XGBoost) [13], and the two branches are combined by a trained logistic-regression meta-learner in a late-fusion step. Late fusion lets each modality be processed by a model suited to its data type before the predictions are merged, instead of mixing raw pixels with genomic values.

A second motivation is honest evaluation. Multimodal systems are often presented as superior to unimodal ones without a direct comparison. Here the CXR-only, genomic-only and fused models are all evaluated on the same held-out test partitions, with patient-level separation between training, validation and test subsets, so that the contribution of each modality can be measured instead of assumed.

1.1 Contributions

The main contributions of this work are summarised below.

- A multimodal framework that integrates chest X-ray images and a 36-feature genomic profile to predict resistance to five first- and second-line anti-TB drugs through five independent sigmoid-activated binary pipelines.
- An antibiotic-specific late-fusion mechanism based on an epsilon-clipped logit transform and a regularised logistic-regression meta-learner fitted on out-of-fold predictions, together with the fitted coefficients for every drug.
- A set of fifteen frozen models (CXR ResNet-18, genomic XGBoost and fusion meta-learner for each antibiotic) trained on antibiotic-specific cohorts with zero cross-split patient overlap.
- A held-out evaluation that compares the three model tiers for all five antibiotics and reports where fusion helps, where it does not, and why a fixed 0.50 threshold can be misleading.
- A deployable research prototype consisting of a Flask REST back end with cached singleton inference and a React/Vite dashboard that shows the branch probabilities, the fused probabilities and a relative ranking.

1.2 Scope and Intended Use

TB-OS V2.0 is an investigational research prototype developed for educational and scientific research. It has not been clinically validated, it has not received regulatory clearance, and it must not be used as a substitute for professional diagnosis, laboratory DST or treatment

decisions. Rifampicin resistance is never used as a proxy for resistance to other drugs, and the system does not generate MDR/XDR diagnoses, regimens or dosages. All probabilities are computed by the back-end inference engine; the front end performs no machine-learning calculation.

1.3 Organisation of the Paper

Section 2 reviews related work on genomic, image-based and multi-modal resistance prediction. Section 3 describes the data, preprocessing, model architecture and fusion mathematics. Section 4 presents the software implementation. Section 5 reports the experimental setup and results. Section 6 discusses the findings, and Section 7 lists limitations and future directions. Section 8 concludes the paper.

2. LITERATURE SURVEY

This section reviews research on resistance prediction in *M. tuberculosis* and in bacteria more generally. Resistance prediction has been approached from several directions: machine learning on whole-genome sequencing (WGS) data, deep learning on genomic mutations, image-based analysis of bacterial response to antibiotics, and epidemiological or experimental studies of resistance patterns. Since the reviewed papers use different data and methods, a structured classification was built to organise them. As shown in Fig. 1, the surveyed studies are grouped into four categories: genomic machine-learning methods, deep-learning methods on genomic mutations, image-based methods, and epidemiological and experimental studies, followed by the research gaps that motivate multimodal fusion.

2.1 Machine-Learning Methods on Genomic Data

Genomic prediction is the most mature direction. Xu et al. [1] presented a machine-learning framework that predicts antimicrobial resistance in *M. tuberculosis* from WGS data and identifies the single nucleotide polymorphisms (SNPs) associated with resistance. Genotype information and susceptibility-testing results were used to train several algorithms for rifampicin, isoniazid, pyrazinamide and ethambutol, and SHapley Additive exPlanations (SHAP) were applied to rank the SNPs that influence the predictions. The approach improves interpretability, but it needs high-quality genomic data and reliable susceptibility labels, it can be affected by limited or imbalanced training data, and it does not use any imaging modality.

Yurtseven et al. [7] combined machine learning with phylogenetic analysis to predict antibiotic resistance in *M. tuberculosis*, showing that evolutionary structure among isolates can be exploited alongside mutation features. Gröschel et al. [3] introduced GenTB, a user-friendly web platform that predicts resistance from genomic sequencing data. GenTB uses a Random Forest (RF) and a Wide and Deep Neural Network (WDNN) and was evaluated on a benchmark of 20,408 isolates with laboratory-confirmed susceptibility results. It was compared with Mykrobe and TB-Profiler and gave slightly higher sensitivity at the cost of a small reduction in specificity. The study also showed that low sequencing depth reduces accuracy for isoniazid and rifampicin, which emphasises the importance of sequencing quality control.

Serajian et al. [14] analysed antibiotic resistance patterns in more than 27,000 *M. tuberculosis* genomes from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC), using the AI-based MTB++ tool to link mutations to resistance and to study how patterns vary across regions and time. Such large-scale studies are useful for surveillance but depend on computational predictions that need experimental confirmation, and the representation of genomes differs between regions.

2.2 Deep-Learning Methods on Genomic Mutations

Wang et al. [6] proposed TB-DROP, a deep-learning framework for predicting drug resistance from whole-genome mutations that does not rely only on predefined resistance-associated loci. The authors built a customised framework in TensorFlow 2, evaluated components of CNNs, denoising autoencoders and Wide and Deep networks, and finally designed a customised multilayer perceptron. A graphical interface lets users analyse sequencing data and obtain predictions. High-quality WGS data and considerable computing resources are required, and the predictions still need biological and clinical validation.

Akwaronwu et al. [2] reviewed machine-learning models for TB drug-resistance prediction, including Hierarchical Attentive Neural Networks (HANN), the MultiDrug Wide and Deep Neural Network (MD-WDNN), DeepAMR, logistic regression, Random Forest and TB-DROP, and compared them using accuracy, sensitivity, specificity and AUC. Performance was found to vary with the antibiotic, the dataset and the architecture, and prediction for ethambutol and pyrazinamide was described as particularly difficult because of complex mutation patterns and limited annotated data. The review identifies the integration of genomic, clinical and demographic information as a promising route to more complete systems, which is the direction taken in the present work. A broader survey by Liu et al. [5] discusses whether AI can help to address the antimicrobial resistance crisis, and Lin et al. [15] describe a machine-learning method for predicting the antimicrobial activity of molecules.

2.3 Image-Based Methods

Image-based work on resistance has mostly used microscopy rather than radiographs. Lei et al. [4] predicted antibiotic susceptibility from fluorescence microscopy images with deep learning, aiming to avoid lengthy conventional testing by detecting changes in bacterial characteristics after antibiotic exposure. Parra-Sánchez et al. [8] proposed an antibacterial susceptibility test from microscopic images using deep learning, and Farrar et al. [9] used citizen science to build an image-based predictor of ciprofloxacin resistance in *Escherichia coli*. Borowa et al. [16] used deep learning to identify bacterial species on microscopic polyculture images. These studies show that image analysis can capture phenotypic response, but their performance depends on the quality, quantity and diversity of the images, and they do not establish that TB resistance can be inferred from chest radiographs.

For TB itself, Khanna et al. [17] examined deep-learning models for tuberculosis prediction, and Lumamba et al. [18] systematically reviewed computer-vision applications for detecting or analysing TB from digitised human lung tissue images. Kumar et al. [10] described a transformer-enabled multimodal approach for TB classification. A chest X-ray, however, carries indirect visual information about lung disease and does not measure bacterial drug susceptibility directly, so a radiograph-only resistance predictor is expected to have limited discrimination.

2.4 Epidemiological and Experimental Studies

Epidemiological studies describe where resistance occurs and who is at risk. Dai et al. [19] analysed a ten-year record (2013 to 2022) from a designated TB hospital in Yunnan, China, using statistical analysis and ordered logistic regression. Resistance levels for most drug categories fell after 2017 and reached lower levels in 2021 before rising again, and male patients and those aged 30 to 60 years showed a higher risk of severe drug-resistant TB. The retrospective single-hospital design limits generalisation, and the study does not build a prediction model for individual patients.

Waller et al. [20] generated drug-resistant *M. tuberculosis* strains against 23 clinically relevant antibiotics and characterised them by whole-genome sequencing and susceptibility testing. They identified collateral sensitivity, where resistance to one drug increases sensitivity to another, and cross-resistance, which may support the design of drug combinations that delay further resistance. Experimental evolution is time-consuming, and the patterns may differ between strains and clin-

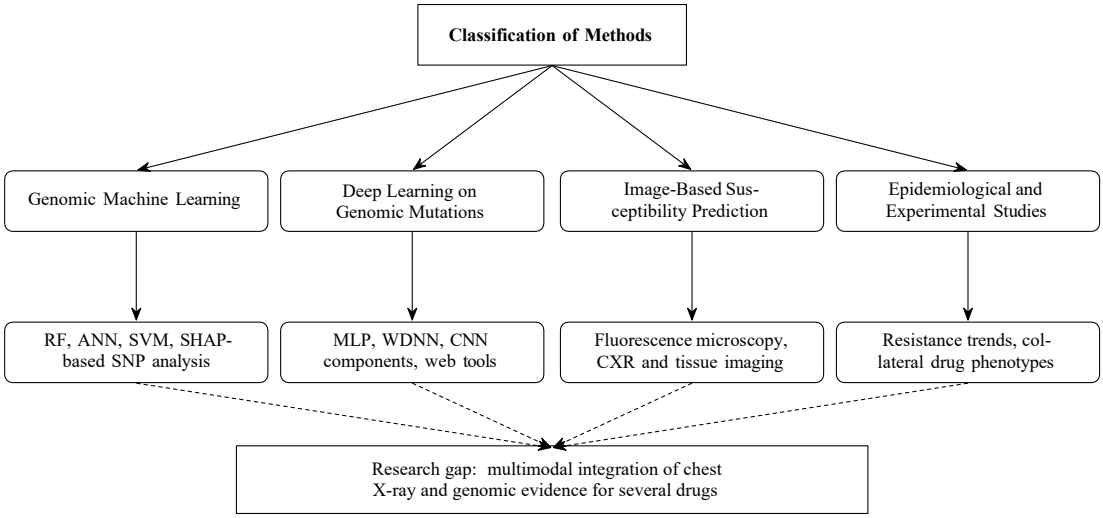


Fig. 1: Classification of Methods Used in Resistance Prediction

ical settings. Ballard et al. [11] review deep-learning approaches for multi-omics integration, which provides a methodological background for combining heterogeneous data sources.

2.5 Comparative Summary and Research Gap

Table 1 summarises the principal studies. Most genomic studies reach high discrimination but do not use imaging, image-based studies work on microscopy or tissue rather than radiographs and genomic profiles together, and epidemiological studies do not provide per-patient prediction. No reviewed study combines a chest radiograph and a genomic profile to predict resistance for several antibiotics in one deployable application, and none reports a direct unimodal-versus-fused comparison on the same held-out partitions.

3. PROPOSED METHODOLOGY

This section describes the data, preprocessing, individual prediction models and the late-fusion mechanism of TB-OS V2.0. The system accepts a paired input consisting of one chest X-ray and one 36-feature genomic profile, validates both, preprocesses them separately, produces image-based and genomic-based probabilities, combines them with a trained fusion model, and returns five independent antibiotic-specific probabilities.

3.1 System Overview

The architecture follows a sequential inference process with two parallel branches, as shown in Fig. 2. The user provides a radiograph and the corresponding genomic features. The back end validates the inputs and prepares each modality independently. The image branch (RGB conversion, resizing and normalisation, followed by ResNet-18) and the genomic branch (validation, ordering and scaling, followed by XGBoost) each generate a probability for every antibiotic. The late-fusion module applies a logit transformation and a trained logistic regression to produce a fused representation, and five independent outputs are displayed on the front-end dashboard. If an input or a required model component fails validation, the system returns an error instead of presenting an unsupported result.

The framework contains fifteen frozen models arranged as three tiers for each of the five antibiotics: Model A (CXR ResNet-18), Model B (genomic XGBoost) and Model C (late-fusion meta-learner). The front end only collects inputs and presents results. The back end validates requests, invokes the model components, computes the fusion probabilities and returns a structured response, which keeps presentation separate from prediction logic.

3.2 Dataset and Antibiotic-Specific Cohorts

The study uses linked chest radiographs, genomic features and phenotypic DST labels from the TB Portals consortium (March 2025 release). Each antibiotic has its own cohort because valid DST labels are not available for every drug in the same number of cases. The data preparation identified 3,079 linked conditions with CXR, genomic and DST information. After applying the CXR usability criterion, 2,423 conditions remained, of which 2,284 requested radiographs were located and read successfully to form the final rifampicin cohort of 2,284 conditions from 2,283 unique patients.

Table 2 lists the cohort and split sizes of the five antibiotic tasks. Counts differ between tasks and must not be added together as a number of unique patients, because the same patient may contribute to several tasks. Each drug label is derived from its own phenotypic field (for example `dst_pheno_rifanpicin`) and is never substituted for another drug.

3.2.1 Training, Validation and Test Splits

Records are divided into training, validation and test subsets at the patient level. Training data fit the models, validation data support model selection, and test data are reserved for the final assessment. Patient-level separation reduces the risk that records from one patient appear in more than one subset, and the validation records report zero patient overlap across the splits for every antibiotic cohort. This does not by itself establish external validity on other hospitals, countries or patient populations.

3.3 Antibiotics and Phenotypic Targets

The five targets represent different drug classes and different primary resistance genes (Table 3). Predictions are produced by independent sigmoid-activated pipelines. No softmax is used, so the drugs do not compete for probability mass, and no drug is used as a proxy for multi-drug resistance.

3.4 Chest X-ray Preprocessing

The image pipeline loads the submitted image, converts it to three-channel RGB, resizes it to 224 × 224 pixels, converts it to a tensor and applies ImageNet normalisation. For channel c, the normalised value is

$$x_c' = \frac{x_c - \mu_c}{\sigma_c}, \quad (1)$$

where x_c is the tensor pixel value, μ_c is the channel mean and σ_c the channel standard deviation, with $\mu = [0.485, 0.456, 0.406]$ and

Table 1: Summary of Related Work

Ref.	Data and method	Target	Limitation noted
[1]	WGS; multiple ML models; SHAP	RIF, INH, PZA, EMB	No imaging; needs quality labels
[3]	WGS; RF and WDNN; 20,408 isolates	First-line drugs	Depth-sensitive; no imaging
[6]	Whole-genome mutations; customised MLP	Multiple drugs	Needs validation; no imaging
[2]	Systematic review of ML/DL models	First- and second-line	No new model; EMB and PZA hard
[14]	27,000+ genomes; MTB++ tool	Resistance patterns	Computational; regional bias
[19]	Hospital records 2013–2022; ordered logistic regression	Severity of resistance	Single hospital; no prediction model
[20]	Experimental evolution; WGS; 23 drugs	Collateral phenotypes	Costly; strain-dependent
[4]	Fluorescence microscopy; deep learning	Susceptibility	Not TB radiographs or genomics
Ours	CXR (ResNet-18) + 36 genomic features (XGBoost); late fusion	RIF, INH, EMB, STR, PZA	Single data source; no external validation

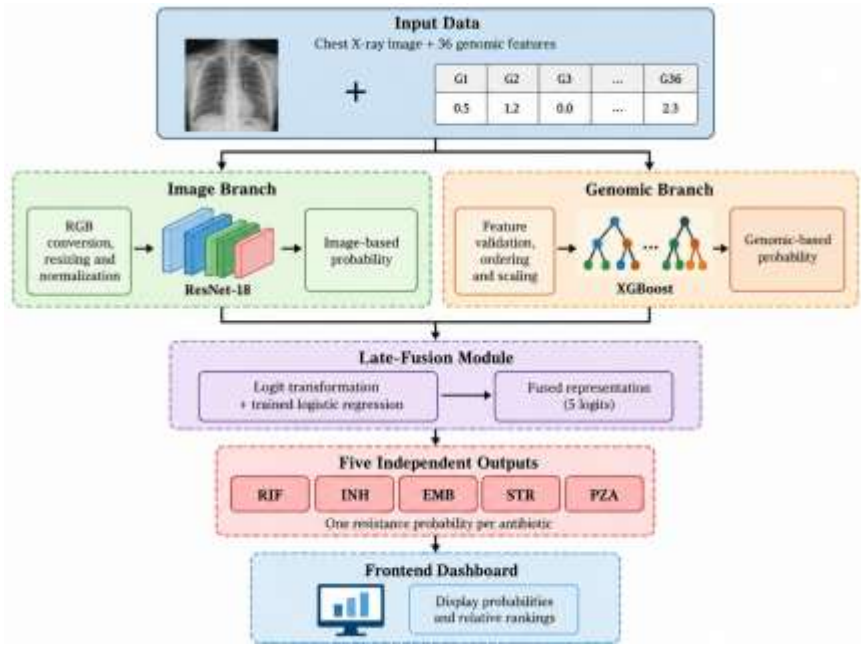


Fig. 2: Block Diagram of the Proposed Multimodal Framework

Table 2: Antibiotic-Specific Cohorts and Patient-Level Splits (Conditions / Patients)

Drug	Conditions	Patients	Train	Validation	Test	Overlap
RIF	2,284	2,283	1,598 / 1,597	229 / 229	457 / 457	0
INH	2,251	2,250	1,572 / 1,571	227 / 227	452 / 452	0
EMB	2,172	2,171	1,505 / 1,504	224 / 224	443 / 443	0
STR	970	969	680 / 679	110 / 110	180 / 180	0
PZA	906	906	644 / 644	94 / 94	168 / 168	0

Table 3: Supported Antibiotics and Primary Resistance Genes Sampled

Code	Antibiotic	Drug class	Genes sampled
RIF	Rifampicin	Rifamycins (first-line)	*rpoB*, *rpoC*
INH	Isoniazid	Hydrazides (first-line)	*katG*, *inhA*, *fabG1*, *ahpC*
EMB	Ethambutol	Arabinogalactan inhibitor	*embB*, *embA*
STR	Streptomycin	Aminoglycosides (second-line)	*rpsL*, *rrs*, *gid*
PZA	Pyrazinamide	Pyrazinoic acid prodrug	*pncA*

$\sigma = [0.229, 0.224, 0.225]$ for the red, green and blue channels. The same transformation is applied at training and inference time so that inputs match the representation expected by the network. Accepted file formats are PNG, JPG and JPEG.

3.5 Genomic Feature Representation and Scaling

Genomic data enter the model as a fixed-length structured vector, not as a raw sequence. The canonical vector has 36 features covering four groups: (i) lineage markers (for example *lin_lineage2*, *lin_lineage4*, *lin_lineage3* and *lin_La1*); (ii) binary mutation status for 24 high-frequency resistance-associated genes (for example *rpoB*, *katG*, *pncA*, *embB*, *gyrA* and *rpsL*); (iii) specimen and sequencing counts (*n_genomic_specimens* and *n_sequencing_runs*); and (iv) baseline demographics (*age_of_onset*, *sex_Male* and *sex_Female*).

Feature names, order, data types and scaling must match those used in training, because an incorrectly ordered vector can give misleading predictions even when its length is correct. A scikit-learn `StandardScaler` fitted on the training partition is saved for each antibiotic. For a standardised feature,

$$z_i = \frac{x_i - \mu_i}{\sigma_i}, \tag{2}$$

with μ_i and σ_i taken from the training data. The RIF scaler standardises three continuous features and leaves the other 33 unscaled, whereas the INH, EMB, STR and PZA scalers standardise all 36 features. Saved scalers are reused and never recomputed from a single request.

3.6 Input Validation

Validation checks that the image can be opened and that the genomic input holds the required features in an accepted CSV or JSON structure. Missing mandatory fields, invalid numeric values, unsupported formats and incompatible structures are rejected. Validation confirms that inputs *can* be processed; it does not prove that the image and the genomic profile belong to the same biological sample, which must be guaranteed by the submission workflow.

3.7 Image Model: ResNet-18

The image branch is a pretrained ResNet-18 backbone [12] adapted to a single-logit binary output. Convolutional layers and residual connections learn spatial representations from the preprocessed radiograph. Let I be the preprocessed image and f_θ the network with learned parameters θ. The representation and probability are

$$\begin{aligned} h_i &= f_\theta(I), \\ p_i^{(a)} &= P(y_a = 1 \mid I) = \sigma\left(w_a^\top h_i + b_a\right), \end{aligned} \tag{3}$$

where $y_a = 1$ is the resistance-positive class for antibiotic a as defined by the DST label. The learned patterns need not be specific to bacterial resistance, since they may reflect other characteristics of the training population, and a radiograph does not directly measure bacterial susceptibility. The image probability is therefore treated as one input to the fusion model and not as direct evidence of a resistance mechanism.

3.8 Genomic Model: XGBoost

The genomic branch uses gradient-boosted decision trees [13]. The trees learn nonlinear relationships and interactions among the structured features. For antibiotic a and the processed vector $\mathbf{x}_g = [x_1, \ldots, x_{36}]$, the genomic probability is

$$p_g^{(a)} = f_g^{(a)}(\mathbf{x}_g), \tag{4}$$

where $f_g^{(a)}$ is the trained XGBoost ensemble for that drug. Separate model artefacts are kept for RIF, INH, EMB, STR and PZA. The output is a statistical prediction, it is not a complete genomic interpretation, and it does not prove that a particular mutation causes resistance.

Table 4: Antibiotic-Specific Fusion Parameters

Antibiotic	Intercept β_0	CXR coefficient β_i	Genomic coefficient β_g
RIF	−0.260561	0.356148	0.354061
INH	−0.007865	0.078151	0.449218
EMB	−0.453998	0.016043	0.066377
STR	0.263049	0.011007	0.328513
PZA	−0.108893	0.019691	0.293200

3.9 Late-Fusion Mechanism

Early fusion combines modalities before modality-specific models make predictions. Late fusion, used here, first obtains separate predictions from the image and genomic models and then combines them. This suits the data because a radiograph (spatial pixels) and a genomic vector (a fixed-length list of encoded features) have different structures, and it lets each branch be evaluated alone before fusion.

A simple arithmetic mean of the two probabilities would impose a fixed equal weighting. Instead, a regularised logistic-regression meta-learner is fitted on out-of-fold (OOF) cross-validation predictions of the training data, so the combination is learned and differs for each antibiotic.

3.9.1 Logit Transformation

To avoid numerical problems at zero and one, each probability is clipped and converted to log-odds:

$$\begin{aligned} \tilde{p} &= \min\left(\max(p, \varepsilon), 1 - \varepsilon\right), \quad \varepsilon = 10^{-6}, \\ \operatorname{logit}(\tilde{p}) &= \ln\frac{\tilde{p}}{1 - \tilde{p}}. \end{aligned} \tag{5}$$

3.9.2 Fusion Equation

For each antibiotic a, the fusion score and the final probability are

$$\begin{aligned} s^{(a)} &= \beta_0^{(a)} + \beta_i^{(a)} \operatorname{logit}\, \tilde{p}_i^{(a)} \\ &\quad + \beta_g^{(a)} \operatorname{logit}\, \tilde{p}_g^{(a)}, \end{aligned} \tag{6}$$

$$p_f^{(a)} = \sigma\left(s^{(a)}\right) = \frac{1}{1 + e^{-s^{(a)}}}, \tag{7}$$

where $\beta_0^{(a)}$ is the intercept, $\beta_i^{(a)}$ the image-branch coefficient and $\beta_g^{(a)}$ the genomic-branch coefficient. The coefficients are learned during training and stored in the fusion artefact of each antibiotic. Table 4 lists the fitted values.

These coefficients describe the fitted combination for each model. They should not be read as universal measures of biological importance or as proof that one modality is always superior. For EMB in particular, both coefficients are small, which keeps the fused score close to the intercept and compresses the output probabilities; this has consequences for threshold-based metrics, as discussed in Section 5.7.

3.9.3 Final Output and Provisional Threshold

The fusion module returns five separate probabilities

$$\begin{aligned} \mathbf{p}_f = \; & p_f^{(\mathrm{RIF})}, p_f^{(\mathrm{INH})}, p_f^{(\mathrm{EMB})}, \\ & p_f^{(\mathrm{STR})}, p_f^{(\mathrm{PZA})}. \end{aligned} \tag{8}$$

These are independent outputs and need not sum to one, because the task is not five-class classification. A provisional threshold of 0.50 converts a probability into a binary output when the evaluation or the interface requires it. This threshold is a research setting and not a clinically validated decision threshold.

3.10 Algorithms

The three core algorithms of the inference pipeline are given below. The image and genomic algorithms run in parallel after validation and both outputs are needed by the fusion algorithm.

Algorithm 1: Image preprocessing and prediction. *Input:* chest X-ray *I*. *Output:* image-based probability.

1. Receive the image and check that it can be opened.
2. Convert to RGB and resize to 224×224 pixels.
3. Convert to a tensor and apply ImageNet normalisation (Eq. 1).
4. Pass the processed image through the trained ResNet-18 of the relevant antibiotic.
5. Return $p_i^{(a)}$ to the fusion stage; on any failure, return an error instead of a prediction.

Algorithm 2: Genomic preprocessing and prediction. *Input:* structured genomic data with 36 features. *Output:* genomic-based probability.

1. Receive the data and validate the structure.
2. Map the values to the canonical 36-feature order.
3. Check for missing or invalid values.
4. Apply the saved scaler of the selected antibiotic (Eq. 2).
5. Pass the vector to the trained XGBoost model (Eq. 4).
6. Return $p_g^{(a)}$ to the fusion stage; report an error if validation or inference fails.

Algorithm 3: Late fusion. *Input:* $p_i^{(a)}$ and $p_g^{(a)}$. *Output:* $p_f^{(a)}$.

1. Receive the two probabilities for antibiotic *a*.
2. Clip each to $[10^{-6}, 1 - 10^{-6}]$ and convert to log-odds (Eq. 5).
3. Load the saved logistic-regression parameters of antibiotic *a*.
4. Compute the fusion score (Eq. 6).
5. Apply the sigmoid (Eq. 7) and return $p_f^{(a)}$.
6. Repeat with the corresponding parameters for all five antibiotics.

3.11 Complete Prediction Workflow

Fig. 3 shows the end-to-end workflow. The user submits a radiograph and genomic data through the front end; the front end sends the request to the back-end prediction endpoint; the back end validates the input and prepares each modality independently; ResNet-18 and XGBoost generate their probabilities; the trained fusion components combine them; the back end returns five independent probabilities in the response format expected by the front end; and the front end displays the results, reporting invalid inputs or inference errors appropriately.

4. SYSTEM IMPLEMENTATION

The application is split into a front end that manages user interaction and a back end that manages input processing and model inference. This separation keeps the presentation layer independent of the trained model parameters and centralises all machine-learning calculation in one place.

4.1 Back-End Implementation

The back end is a Flask REST API. A single prediction endpoint receives the paired inputs, validates them and coordinates the prediction components. The multi-antibiotic logic is implemented in a `NultiAntibioticEngine` class, which loads all fifteen frozen models once as a cached singleton and reuses them for every request, so models are neither retrained nor reloaded per request. Application routes and configuration are organised in `app. py`, `config. py` and `inference_config. py`, and a lightweight SQLAlchemy layer (SQLite in development) stores prediction records and telemetry. Additional utility modules in the repository cover image-quality assessment, modality-fusion gating, report generation and explainability helpers; the evaluated prediction path used in this paper is the three-tier pipeline described in Section 3.

4.2 Front-End Implementation

The front end is a single-page application built with React and Vite. It offers separate upload controls for the radiograph (PNG, JPG, JPEG) and the genomic file (CSV or JSON), optional patient and condition identifiers that are explicitly labelled as not affecting the prediction, and a results dashboard. A small API client validates the response schema before display, and a utility maps probabilities to a three-level visual band. The front end performs no probability calculation or normalisation; it only renders what the back end returns.

4.3 Model Artefact Management

The artefact collection contains ResNet-18 checkpoints, XGBoost genomic models, genomic scalers and logistic-regression fusion components, kept separately for each antibiotic. For every task, the inference engine must select the matching image model, genomic model, scaler and fusion component. Mixing artefacts between tasks could combine probabilities trained for different targets or apply a scaler with incompatible parameters. Feature order, preprocessing rules and output interpretation must stay identical to the evaluation set-up; any change to these components should be re-evaluated before the system is described as equivalent to the tested version. File integrity of the processed artefacts is tracked with a hash ledger.

4.4 Serial Demonstration Dataset

For demonstration and integration testing, a pool of 500 verified multimodal pairs is provided, named `CXR_XXXXX. png` and `GENDNIC_XXXXX. csv` with a shared five-digit serial number. Pairs 1 to 457 come from the frozen held-out RIF test partition and pairs 458 to 500 come from held-out validation encounters. The serial number is only a convenience alias; every pairing is linked through the condition identifier.

4.5 Software Environment

Table 5 lists the software used in the documented validation environment, which was CPU-based. This shows that the inference pipeline can run without a dedicated GPU in that environment.

4.6 Testing Strategy

Testing is organised in component, API, integration and model-evaluation layers. Component tests cover image preprocessing, genomic feature ordering, scaler use, artefact loading and fusion arithmetic. A 15-point automated back-end suite checks that the API accepts valid requests, rejects invalid inputs and returns a structured response, and a 26-point front-end suite checks the interface and the client. Integration tests confirm that the front end, back end and inference engine work together. Software tests and model evaluation answer different questions: a request can pass an API test even when the model has limited predictive power, and a model with good metrics can sit behind an application with a request-handling defect. Both are needed for a reliable research prototype, and neither establishes clinical validity.

5. EXPERIMENTAL RESULTS AND ANALYSIS

This section reports the experimental set-up, the evaluation metrics, the held-out results for all fifteen models, the user-interface results and the runtime measurements.

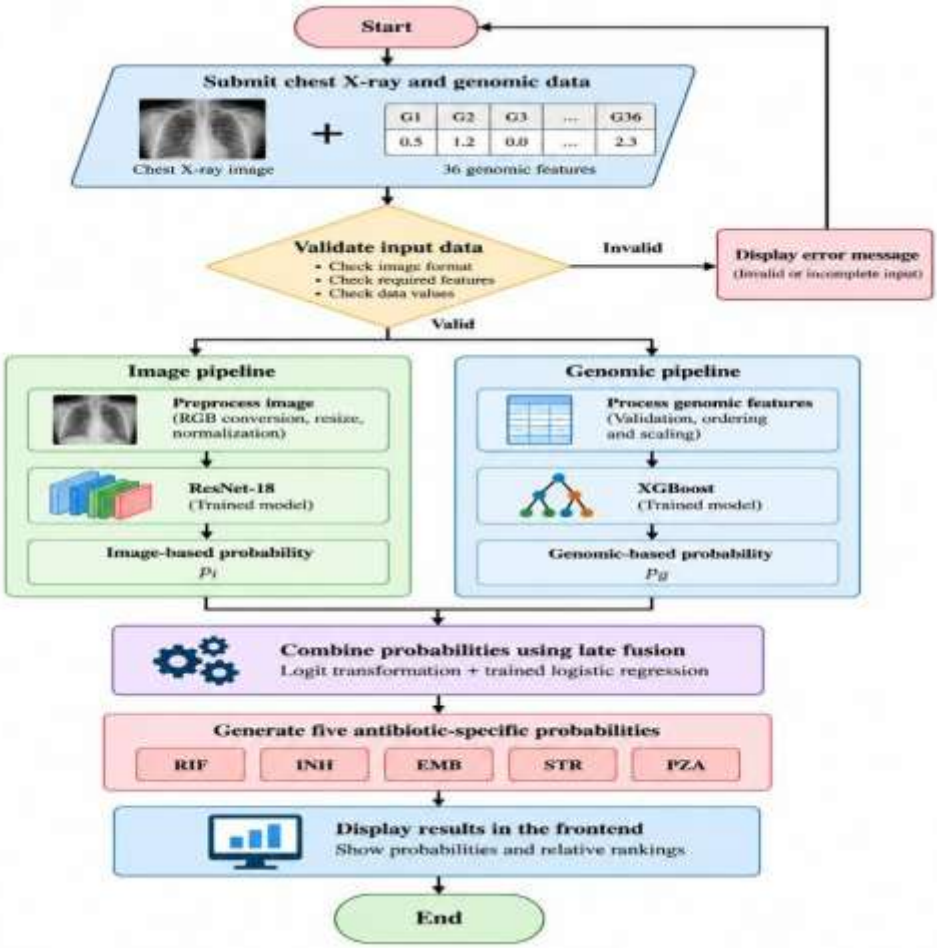


Fig. 3: Flowchart of the Prediction Process

Table 5: Software Environment of the Prototype

Component	Technology (version)	Role
Back-end language	Python 3.12	Inference orchestration and data processing
Web back end	Flask 2.3.3	HTTP API and request handling
Deep learning	PyTorch 2.5.1 (CPU)	ResNet-18 inference
Image utilities	Torchvision 0.20.1 (CPU); Pillow 11.0.0	Image loading and transformation
Genomic model	XGBoost 3.2.0	Structured genomic prediction
Preprocessing and fusion	Scikit-learn 1.5.2	Scaling and logistic-regression fusion
Front end	React 18.3.1	User interface and results dashboard
Build tools	Node.js 24.19.0; Vite 8.0.10	Development and production builds
Visual components	Three.js 0.184.0; Framer Motion 12.38.0; Lucide React 1.14.0	Visual effects, animation and icons

5.1 Experimental Set-Up

All metrics come from the untouched held-out test partition of each antibiotic, evaluated with the frozen model checkpoints. Training data fit the models, validation data supported model selection, and the test data were used only for the final assessment. Each antibiotic has its own test-set size: 457 (RIF), 452 (INH), 443 (EMB), 180 (STR) and 168 (PZA). The inference environment was CPU-based, and execution time varies with the processor, memory, input size and model-loading behaviour.

5.2 Evaluation Metrics

Metrics are computed for each antibiotic and for each model branch (Table 6). Accuracy can be misleading when classes are imbalanced, so ROC-AUC and PR-AUC are reported together with sensitivity, specificity and the F1-score. Threshold-dependent metrics use the provisional cutoff $p \geq 0.50$.

Table 6: Classification Metrics Used in the Evaluation

Metric	Definition
Accuracy	(TP + TN) / (TP + TN + FP + FN)
Precision	TP / (TP + FP)
Sensitivity (recall)	TP / (TP + FN)
Specificity	TN / (TN + FP)
F1-score	2 × Precision × Recall / (Precision + Recall)
ROC-AUC	Area under the receiver operating characteristic curve (ranking across thresholds)
PR-AUC	Area under the precision–recall curve

5.3 Ranking Performance of the Three Model Tiers

Table 7 and Fig. 4 compare the ROC-AUC and PR-AUC of the CXR, genomic and fused models for all five antibiotics. The genomic model clearly dominates the image model for every drug. The fused model is

slightly lower than the genomic model for RIF and slightly higher for INH, EMB, STR and PZA.

5.4 Threshold-Based Performance

Table 8 gives accuracy, sensitivity, specificity and F1-score at the provisional 0.50 cutoff. The genomic model is strong for all five drugs. The image model shows an unbalanced error profile: for STR its sensitivity is 0.8585 but its specificity only 0.2297, and for PZA the opposite holds, with a sensitivity of 0.0128 and a specificity of 0.9778, so it labels almost every PZA sample as susceptible at this threshold. The fused model matches the genomic model for INH, is close to it for STR and PZA, and is lower for RIF.

5.5 Effect of Fusion Relative to the Genomic Model

Because the genomic branch is the strongest single modality, the fairest question is whether adding the image probability helps. Table 9 reports the difference between the fused and genomic ranking metrics. The change is negative for RIF (a drop of 0.0142 in ROC-AUC), essentially neutral for INH, STR and PZA (gains of 0.0012 to 0.0019 in ROC-AUC), and largest for EMB (+0.0108 in ROC-AUC and +0.0409 in PR-AUC). These differences are small, were measured on one split per drug and were not tested for statistical significance, so they should be read as indicative and not as proof that fusion improves performance.

5.6 Rifampicin Analysis and WGS Benchmark

For rifampicin the genomic model reached a ROC-AUC of 0.9623 and a PR-AUC of 0.9470, the image model 0.7589 and 0.7371, and the fused model 0.9481 and 0.9394. A recorded WGS benchmark has a ROC-AUC of 0.9607 and a PR-AUC of 0.9456, so the genomic model is higher by 0.0016 and 0.0014 respectively. These differences are too small to be called an improvement unless the test records, target definitions and evaluation procedures of the benchmark are confirmed to be comparable, and the exact provenance of the benchmark should be stated whenever the value is quoted. The image model's sensitivity (0.7661) exceeds its specificity (0.5649), which shows that its false-positive and false-negative errors are unbalanced at the evaluated threshold.

5.7 Ethambutol: Ranking versus Fixed Threshold

The EMB fused model illustrates the difference between ranking quality and a fixed decision threshold. Its ROC-AUC (0.9353) and PR-AUC (0.8372) are the best of the three tiers, but the fused probabilities on the EMB test set span only 0.0384 to 0.4990. Since every value lies below 0.50, the sensitivity and F1-score evaluate to 0.0000 at that cutoff and the specificity to 1.0000. The model therefore ranks positive and negative samples well but never crosses the provisional cutoff. This is consistent with the very small fusion coefficients of Table 4, which compress the score range. The case shows why 0.50 is a provisional research threshold and not a validated clinical one, and it argues for probability calibration and drug-specific operating points in future work.

5.8 User-Interface Results

The prototype dashboard (TB-OS V2.0) offers a unified workflow for the two modalities. Fig. 5 shows the initial screen with separate upload areas for the radiograph and the genomic CSV or JSON file. The optional patient and condition identifiers are labelled as tracking-only fields that do not affect the prediction, and the screen states that both modalities are required.

Fig. 6 shows the state after a radiograph and a genomic file containing 36 approved features have been selected. The selected files are displayed before the request is submitted, which gives early feedback on the input; whether the two files truly belong to the same sample must still be ensured by the data-submission process.

The results dashboard in Fig. 7 displays the five independent outputs with the CXR, genomic and fused probabilities of each drug and a relative ranking. In the displayed example the fused probabilities are 1.89% (RIF), 19.79% (INH), 32.22% (EMB), 41.70% (STR) and 24.02% (PZA). Streptomycin has the highest numerical value, followed by EMB and PZA, while RIF and INH occupy the bottom two positions. The top three values are marked green and the bottom two red; this is a display convention that makes numerical comparison easier and is not a clinical classification. A result can be among the top three and still lie below the provisional 50% threshold, as in this example. Fig. 8 shows the input section and the results in one view.

5.9 Runtime and Software Validation

The recorded validation run measured warm API request times of 672.18 ms, 507.25 ms and 434.77 ms, with a mean of 538.07 ms on the CPU-based set-up. These values describe one validation run and are not a universal latency guarantee, because runtime depends on the machine, model initialisation, file size and system load. The back-end and front-end test suites and the model-artefact loading checks passed for the tested code version. Passing software tests shows that the selected workflows operate as intended; it does not demonstrate clinical utility, external validity or regulatory approval.

6. DISCUSSION

6.1 Why the Genomic Branch Dominates

Across all five antibiotics the genomic model reached ROC-AUC values between 0.9245 and 0.9739, against 0.6376 to 0.7589 for the image model. A plausible explanation is that the 36-feature vector contains binary mutation indicators for genes that are directly linked to the mechanism of resistance of each drug (for example *rpoB* for rifampicin, *katG* and *inhA* for isoniazid, *embB* for ethambutol and *pncA* for pyrazinamide), together with lineage markers. A radiograph, in contrast, records the macroscopic appearance of the lungs and carries only indirect information about bacterial drug susceptibility. This is an interpretation of the observed metrics and not a proven causal explanation. It agrees with earlier genomic studies, which report strong discrimination from sequence-derived features [1, 3, 6], and with the observation that no study has shown that resistance can be read directly from a chest radiograph.

6.2 What the Chest X-ray Contributes

The CXR model ranks samples better than chance for every drug, which indicates that the radiograph does carry some resistance-related signal, possibly through correlated characteristics such as disease extent, treatment history or patient subgroup. These patterns are not necessarily specific to resistance, and they may reflect properties of the training population. The threshold-based metrics show that the signal is poorly calibrated: for PZA the model almost never exceeds 0.50, and for STR it rarely falls below it. This is the main reason why the image probability is used as one input to a learned fusion model and not as a standalone answer.

6.3 Behaviour of Late Fusion

Fusion did not behave uniformly. For RIF, where the genomic model is already very strong (ROC-AUC 0.9623) and both fusion coefficients are of similar size (0.356 and 0.354), adding the weaker image logit lowered the ranking metrics slightly. For INH, STR and PZA, the fusion coefficient of the image branch is small compared with that of the genomic branch, and the fused model matched the genomic model within about 0.002 ROC-AUC. EMB showed the largest gain in ranking metrics, but also the most extreme threshold effect. Taken together,

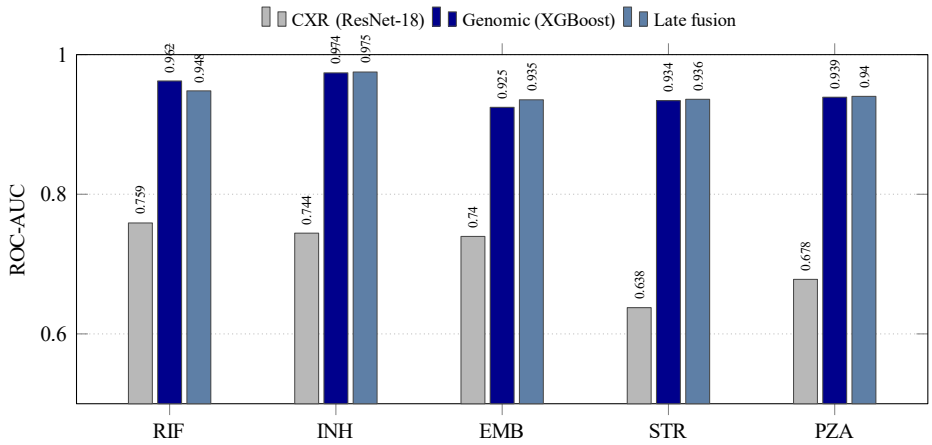


Fig. 4: Held-Out ROC-AUC of the CXR, Genomic and Late-Fusion Models for the Five Antibiotics

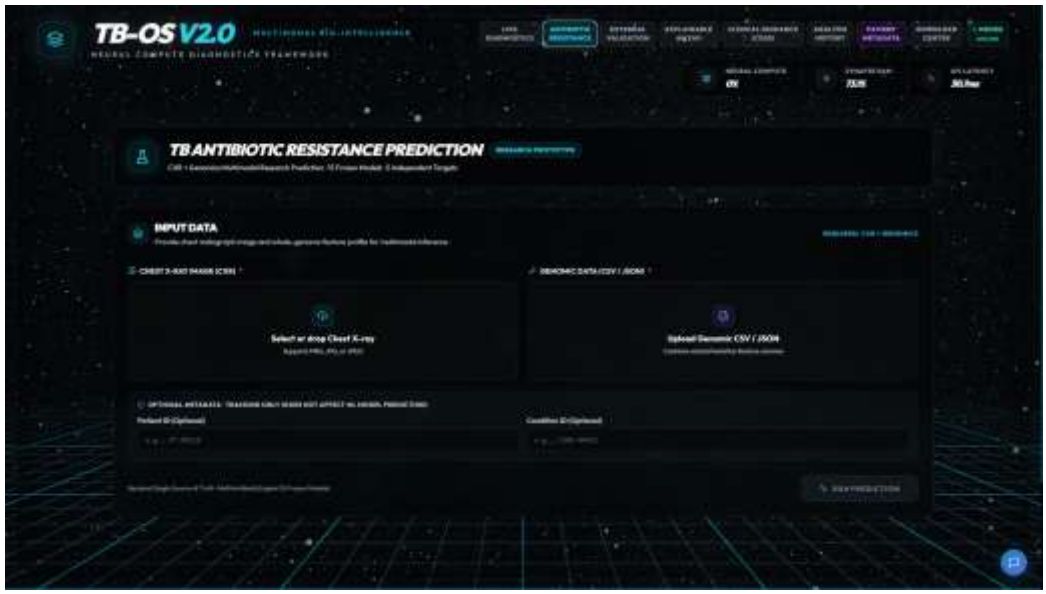


Fig. 5: Initial Input Interface

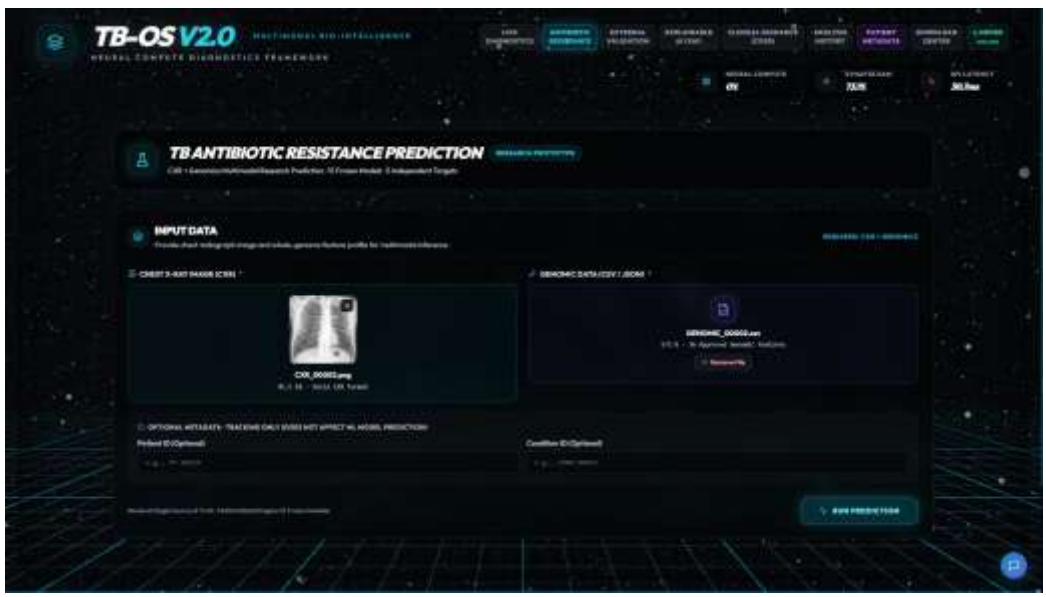


Fig. 6: Input Files Selected Before Prediction

Table 7: Held-Out ROC-AUC and PR-AUC of the Three Model Tiers

		ROC-AUC			PR-AUC		
Drug	*N*	**CXR**	**Genomic**	**Fusion**	**CXR**	**Genomic**	**Fusion**
RIF	457	0.7589	**0.9623**	0.9481	0.7371	**0.9470**	0.9394
INH	452	0.7443	0.9739	**0.9751**	0.7731	**0.9780**	0.9778
EMB	443	0.7396	0.9245	**0.9353**	0.5564	0.7963	**0.8372**
STR	180	0.6376	0.9341	**0.9360**	0.7280	0.9624	**0.9631**
PZA	168	0.6781	0.9390	**0.9402**	0.6042	0.9263	**0.9279**

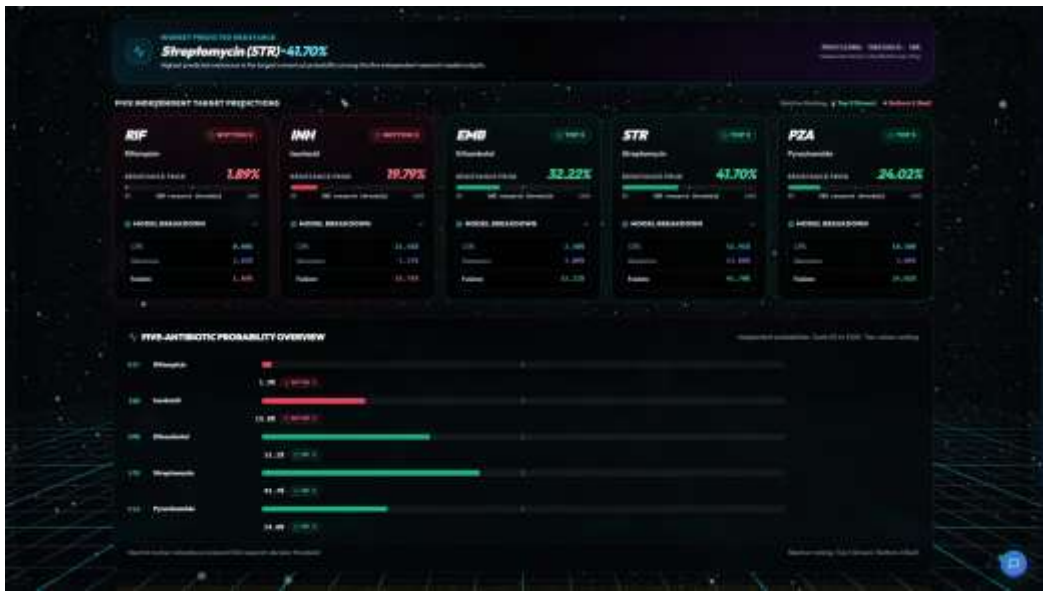


Fig. 7: Antibiotic Prediction Results Dashboard

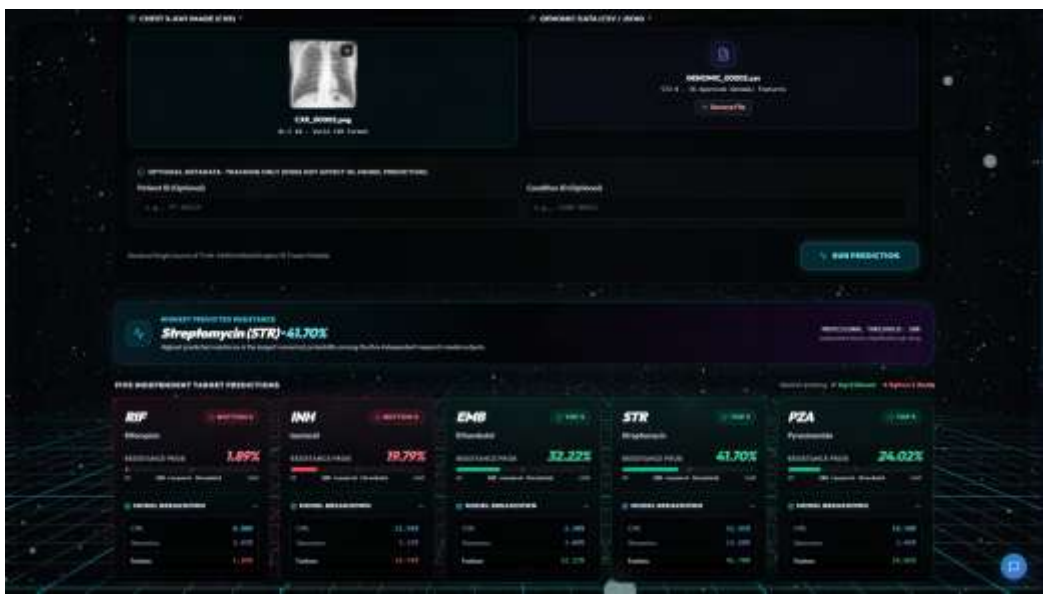


Fig. 8: Integrated Input and Results View

Table 8: Held-Out Threshold-Based Metrics at $p \geq 0.50$

Drug	Model	Accuracy	Sensitivity	Specificity	F1-score
RIF	CXR	0.6608	0.7661	0.5649	0.6830
	Genomic	0.9628	0.9495	0.9749	0.9606
	Fusion	0.8928	0.8991	0.8870	0.8889
INH	CXR	0.6814	0.5992	0.7721	0.6636
	Genomic	0.9602	0.9409	0.9814	0.9612
	Fusion	0.9602	0.9409	0.9814	0.9612
EMB	CXR	0.7246	0.4604	0.8454	0.5120
	Genomic	0.8691	0.9281	0.8421	0.8165
	Fusion	0.6862†	0.0000†	1.0000	0.0000†
STR	CXR	0.6000	0.8585	0.2297	0.7165
	Genomic	0.8889	0.9057	0.8649	0.9057
	Fusion	0.8889	0.9434	0.8108	0.9091
PZA	CXR	0.5298	0.0128	0.9778	0.0247
	Genomic	0.9226	0.8974	0.9444	0.9150
	Fusion	0.9167	0.8718	0.9556	0.9067

† All fused EMB test probabilities fall below 0.50 (see Section 5.7).

Table 9: Fusion Minus Genomic: Change in Ranking Metrics

Drug	Δ ROC-AUC	Δ PR-AUC	Direction
RIF	−0.0142	−0.0076	Lower
INH	+0.0012	−0.0002	Neutral
EMB	+0.0108	+0.0409	Higher
STR	+0.0019	+0.0007	Neutral
PZA	+0.0012	+0.0016	Neutral

these results suggest that the image branch adds little complementary information beyond the genomic features in this cohort, or that its errors are too correlated with the genomic errors to improve the fusion. Both explanations remain hypotheses, and testing them would need ablations and error analysis, which are listed as future work. The honest conclusion is that multimodal fusion is technically feasible in a single application but that it did not systematically outperform the strongest individual branch.

6.4 Drug-Specific Difficulty

Performance differed between antibiotics. The precision–recall behaviour of the EMB models (genomic PR-AUC 0.7963, fused 0.8372) is clearly lower than that of INH (about 0.978) or STR (about 0.963), which is in line with the review by Akwaronwu et al. [2], who describe ethambutol and pyrazinamide as hard targets because of complex mutation patterns and limited annotated data. STR and PZA also have the smallest cohorts (970 and 906 conditions, with 180 and 168 test samples), so their estimates carry more uncertainty, and the reported figures should be accompanied by confidence intervals in future work.

6.5 Interpretation and Clinical Safety

Probabilities are model estimates and not laboratory results. The provisional 0.50 threshold, the green and red rank encoding and the highest-probability label in the dashboard are conveniences for inspection. They are not treatment recommendations, and no mutation should be described as the proven cause of resistance on the basis of model output. Incorrect predictions could have serious consequences if treated as definitive, so the system must not replace laboratory DST or professional clinical assessment.

6.6 Comparison with Existing Work

The results support a direct comparison among the CXR-only, genomic-only and fused branches within this project and, for rifampicin, against a recorded WGS benchmark whose values are close to those of the genomic model. They do not establish superiority over published methods, because those studies use different populations, label definitions, splits and metrics. The main difference of the present system is the integration of a radiograph and a genomic profile into one web-based workflow with five antibiotic-specific outputs and visible branch breakdowns, together with an open report of the cases in which fusion does not help.

6.7 Societal and Environmental Considerations

The framework may support research into whether image-derived and genomic-derived signals are complementary, and a single interface for two modalities makes experimental workflows easier to inspect. Several responsible-use points apply. Chest X-rays and genomic records are sensitive data, and their collection, storage, transfer and display need access controls, consent and institutional approval. Performance may differ across patient groups, sites, imaging equipment and genomic lineages, and uneven performance could widen disparities if it is not examined. Reusing frozen artefacts avoids retraining for every request, and CPU-only inference lowers the hardware requirements of the tested set-up, although the actual energy savings were not measured. No public-health benefit has been demonstrated so far.

7. LIMITATIONS AND FUTURE WORK

7.1 Limitations

The study has several limitations. First, all data come from a single consortium release (TB Portals) and no external validation on other hospitals, countries or imaging equipment was performed. Second, each result comes from one patient-level split per antibiotic without confidence intervals or significance tests, and the STR and PZA test sets are small. Third, the genomic input is a processed 36-feature vector and not a raw sequence, so genomic information outside these features is not used. Fourth, the 0.50 threshold is provisional and the fused probabilities are not calibrated, as the EMB case shows. Fifth, the radiograph and the genomic profile are paired through record identifiers, and the application cannot verify that two uploaded files belong to the same patient. Finally, the system has not been evaluated in a prospective or clinical setting.

7.2 Future Work

Future work should evaluate the models on independent datasets from different institutions and populations, report confidence intervals and statistical tests, and add probability calibration and drug-specific operating points. Alternative image architectures and image pre-training on radiograph data, richer genomic representations and other fusion techniques (for example gated or attention-based fusion) can be compared with the present logistic-regression fusion, together with ablations that isolate the contribution of each modality. Explainable AI methods can show which image regions and genomic features drive the predictions. The application can be extended with stronger input validation and sample-identity checks, secure data management, detailed reports and prediction history. After technical, ethical and clinical validation, the framework could be examined as a source of supplementary information for resistance research and surveillance.

8. CONCLUSION

This paper presented TB-OS V2.0, a multimodal research prototype that combines a ResNet-18 chest X-ray model and an XGBoost genomic model through an antibiotic-specific logistic-regression late-fusion layer to estimate resistance of *M. tuberculosis* to rifampicin, isoniazid, ethambutol, streptomycin and pyrazinamide. Fifteen frozen models were trained on antibiotic-specific cohorts with zero patient overlap across splits and exposed through a Flask API and a React dashboard that returns five independent probabilities.

On the held-out test sets, the genomic model was the strongest single modality for every drug (ROC-AUC 0.9245 to 0.9739), the chest X-ray model had limited discrimination (0.6376 to 0.7589), and late fusion gave only small changes: a decrease for rifampicin (0.9481 against

0.9623) and gains of at most 0.0108 ROC-AUC for the other four drugs. The ethambutol case also showed that a fixed 0.50 threshold can hide good ranking performance. The work therefore demonstrates that multimodal integration is practical, while showing that it does not automatically beat the best single modality. The system remains a research prototype, and external validation, calibration and clinical evaluation are required before any practical use can be considered.

REFERENCES

[1] Y. Xu, Y. Mao, X. Hua, et al., "Machine learning-based prediction of antimicrobial resistance and identification of AMR-related SNPs in *Mycobacterium tuberculosis*," *BMC Genom. Data*, vol. 26, Art. no. 48, 2025, doi: 10.1186/s12863-025-01338-x.

[2] B. G. Akwaronwu, F. Y. Ayankoya, S. O. Kuyoro, I. U. Akwaronwu, and O. J. Adeniyi, "Machine learning for predicting drug resistance in tuberculosis: a systematic review of model performance and clinical applications," *Big Data Min. Anal.*, vol. 9, no. 1, pp. 314–337, 2026, doi: 10.26599/BDMA.2025.9020063.

[3] M. I. Gröschel, M. Owens, L. Freschi, et al., "GenTB: a user-friendly genome-based predictor for tuberculosis resistance powered by machine learning," *Genome Med.*, vol. 13, Art. no. 138, 2021, doi: 10.1186/s13073-021-00953-4.

[4] B. Lei, T. Zhang, J. Li, J. Qu, W. Xiong, K. Yu, and L. Yang, "Rapidly antibiotic susceptibility prediction via deep learning from bacterial fluorescence microscopy images," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 21469–21481, 2025, doi: 10.1109/TASE.2025.3607484.

[5] G. Y. Liu, D. Yu, M. M. Fan, et al., "Antimicrobial resistance crisis: could artificial intelligence be the solution?," *Mil. Med. Res.*, vol. 11, Art. no. 7, 2024, doi: 10.1186/s40779-024-00510-1.

[6] Y. Wang, Z. Jiang, P. Liang, et al., "TB-DROP: deep learning-based drug resistance prediction of *Mycobacterium tuberculosis* utilizing whole genome mutations," *BMC Genomics*, vol. 25, Art. no. 167, 2024, doi: 10.1186/s12864-024-10066-y.

[7] A. Yurtseven, S. Buyanova, A. Agrawal, et al., "Machine learning and phylogenetic analysis allow for predicting antibiotic resistance in *M. tuberculosis*," *BMC Microbiol.*, vol. 23, Art. no. 404, 2023, doi: 10.1186/s12866-023-03147-7.

[8] L. F. Parra-Sánchez, S. I. Rivera-Manrique, A. González-Vega, and L. C. Padierna, "A novel antibacterial susceptibility test from microscopic images by deep learning," *IEEE Access*, vol. 13, pp. 166670–166685, 2025, doi: 10.1109/ACCESS.2025.3612431.

[9] A. Farrar, C. Feehily, P. Turner, et al., "Infection Inspection: using the power of citizen science for image-based prediction of antibiotic resistance in *Escherichia coli* treated with ciprofloxacin," *Sci. Rep.*, vol. 14, Art. no. 19543, 2024, doi: 10.1038/s41598-024-69341-3.

[10] S. Kumar, S. Sharma, and K. T. Megra, "Transformer enabled multi-modal medical diagnosis for tuberculosis classification," *J. Big Data*, vol. 12, Art. no. 5, 2025, doi: 10.1186/s40537-024-01054-w.

[11] J. Ballard, Z. Wang, W. Li, et al., "Deep learning-based approaches for multi-omics data integration and analysis," *BioData Min.*, vol. 17, Art. no. 38, 2024, doi: 10.1186/s13040-024-00391-z.

[12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[13] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[14] M. Serajian, C. Testagrose, M. Prosperi, et al., "A comparative study of antibiotic resistance patterns in *Mycobacterium tuberculosis*," *Sci. Rep.*, vol. 15, Art. no. 5104, 2025, doi: 10.1038/s41598-025-89087-w.

[15] B. Lin, S. Yan, and B. Zhen, "A machine learning method for predicting molecular antimicrobial activity," *Sci. Rep.*, vol. 15, Art. no. 6559, 2025, doi: 10.1038/s41598-025-91190-x.

[16] A. Borowa, D. Rymarczyk, D. Ochon´ska, A. Sroka-Oleksiak, M. Brzychczy-Włoch, and B. Zielin´ski, "Identifying bacteria species on microscopic polyculture images using deep learning," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 1, pp. 121–130, 2023, doi: 10.1109/JBHI.2022.3209551.

[17] D. Khanna, N. Garg, and P. Sharma, "Prediction analysis of *Mycobacterium tuberculosis* by using deep learning model," in *2023 Int. Conf. Res. Methodol. Knowl. Manag., Artif. Intell. Telecommun. Eng. (RMK-MATE)*, 2023, pp. 1–4, doi: 10.1109/RMKMATE59243.2023.10369599.

[18] K. D. Lumamba, G. Wells, D. Naicker, et al., "Computer vision applications for the detection or analysis of tuberculosis using digitised human lung tissue images: a systematic review," *BMC Med. Imaging*, vol. 24, Art. no. 298, 2024, doi: 10.1186/s12880-024-01443-w.

[19] J. Dai, Q. Li, J. Chen, et al., "Drug resistance patterns among tuberculosis patients at a tuberculosis designated hospital in Yunnan: a retrospective analysis of the 2013–2022 period," *BMC Infect. Dis.*, vol. 26, Art. no. 33, 2026, doi: 10.1186/s12879-025-12251-0.

[20] N. J. E. Waller, C. Y. Cheung, G. M. Cook, et al., "The evolution of antibiotic resistance is associated with collateral drug phenotypes in *Mycobacterium tuberculosis*," *Nat. Commun.*, vol. 14, Art. no. 1517, 2023, doi: 10.1038/s41467-023-37184-7.