Computational and Structural Biotechnology Journal 27 (2025) 2208–2218

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research Article

Deep learning-based framework for Mycobacterium tuberculosis bacterial growth detection for antimicrobial susceptibility testing

Hoang-Anh T. Vo [a], Sang Nguyen [a], Ai-Quynh T. Tran [a], Han Nguyen [a], Hai Bich Ho [b,c], Philip W. Fowler [c,d,e], Timothy M. Walker [b,c], Thuy Thi Nguyen [a,*]

[a] School of Science, Engineering & Technology (SSET), RMIT University, Ho Chi Minh, Viet Nam
[b] Oxford University Clinical Research Unit, Ho Chi Minh, Viet Nam
[c] Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom
[d] Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance, University of Oxford, Oxford, United Kingdom
[e] National Institute of Health Research Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, United Kingdom

ABSTRACT

Tuberculosis (TB) kills more people annually than any other pathogen. Resistance is an ever-increasing global problem, not least because diagnostics remain challenging and access limited. 96-well broth microdilution plates offer one approach to high-throughput phenotypic testing, but they can be challenging to read. Automated Mycobacterial Growth Detection Algorithm (AMyGDA) is a software package that uses image processing techniques to read plates, but struggles with plates that exhibit low growth or images of low quality. We developed a new framework, TMAS (TB Microbial Analysis System), which leverages state-of-the-art deep learning models to detect *M. tuberculosis* growth in images of 96-well microtiter plates. TMAS is designed to measure Minimum Inhibitory Concentrations (MICs) rapidly and accurately while differentiating between true bacterial growth and artefacts. Using 4,018 plate images from the CRyPTIC (Comprehensive Resistance Prediction for Tuberculosis: An International Consortium) dataset to train models and refine the framework, TMAS achieved an essential agreement of 98.8%, significantly outperformed the 90% threshold established by the International Organization for Standardization (ISO). TMAS offers a reliable, automated and complementary evaluation to support expert interpretation, potentially improving accuracy and efficiency in tuberculosis drug susceptibility testing (DST).

1. Introduction

Tuberculosis still kills more people each year than any other single pathogen and drug resistance is an increasingly important problem [1]. Whilst the drug development pipeline is better stocked than for decades, the longevity of new drug regimens will depend heavily on our ability to minimise the evolution of resistance by treating patients with drugs to which their tuberculosis is susceptible [2,3]. Achieving this through the culture-based, phenotypic drug susceptibility testing (pDST) methods in routine use today would be costly and complex [4]. There has consequently been a drive from the WHO and others towards developing and implementing molecular assays, most recently targeted next generation sequencing (tNGS) [5].

The CRyPTIC (Comprehensive Resistance Prediction for Tuberculosis: an International Consortium) project was a global collaboration that set out to generate a near-comprehensive understanding of the relationship between *M. tuberculosis* genetic mutations and minimum inhibitory concentrations (MICs) to 14 anti-tuberculosis drugs. The project made a substantial contribution towards the WHO catalogue of mutations associated with resistance in *M. tuberculosis* [6,7].

CRyPTIC required a new phenotypic approach to generate MIC for up to 15,000 *M. tuberculosis* strains. The project therefore developed new 96-well broth microdilution plates together with ThermoFisher [8]. Such plates are now more commonly used and newer plates are in development that will generate MICs for many of the newer drugs. The experience of the CRyPTIC consortium however shows that plate reading can be challenging due to the presence of artefacts such as sediment, air bubbles, condensation, contamination, plate failures, and shadows, as well as variability in image quality. CRyPTIC addressed this by having plates read by laboratory scientists, by BashTheBug, a citizen science initiative that collects volunteer-based classifications, and by AMyGDA, a rule-based image analysis tool, and then seeking a consensus from these three methods [8–10]. In BashTheBug, the “wisdom of the crowd” method helps balance out individual mistakes, but volunteers often have a tendency to overcall as a cautious response to the risk of missing potential growth [9]. On the other hand, AMyGDA is highly sensitive to artifacts and often fails to detect bacterial growth when handling low-

* Corresponding author.
E-mail address: thuy.nguyen43@rmit.edu.vn (T.T. Nguyen).

https://doi.org/10.1016/j.csbj.2025.05.030
Received 25 February 2025; Received in revised form 20 May 2025; Accepted 20 May 2025
Available online 26 May 2025

Nomenclature

Abbreviations

AMyGDA	Automated Mycobacterial Growth Detection Algorithm
BRC	Biomedical Research Centre
CRyPTIC	Comprehensive Resistance Prediction for Tuberculosis: An International Consortium
CLAHE	Contrast-Limited Adaptive Histogram Equalization
DST	Drug Susceptibility Testing
HPRU	Health Protection Research Unit
ips	Images per second
mAP	Mean Average Precision
MGIT	Mycobacterial Growth Indicator Tube
MIC	Minimum Inhibitory Concentration
MSF	Mean Shift Filtering
NHS	National Health Service
PAS	Para-aminosalicylic acid
pDST	Phenotypic Drug Susceptibility Testing
TB	Tuberculosis
TMAS	Tuberculosis Microbial Analysis System
tNGS	Targeted next-generation sequencing

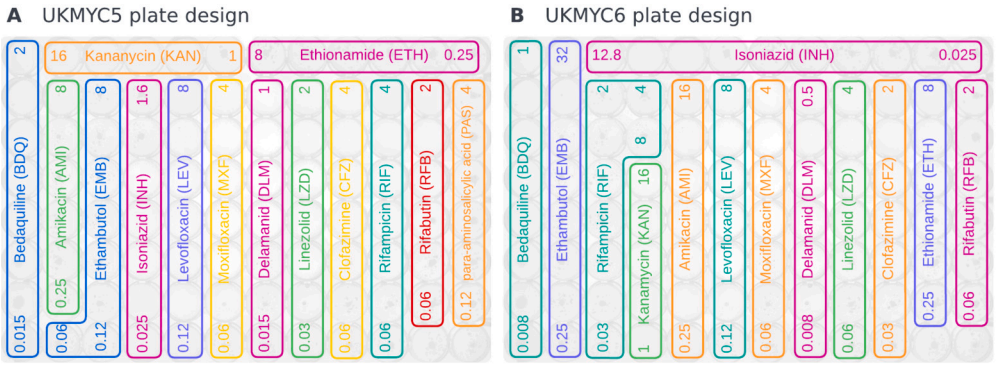


Fig. 1. Plate designs for (A) UKMYC5 and (B) UKMYC6. Text labels indicate drug names, while numerical values denote the corresponding drug concentration ranges [12]. Each well has double the drug concentration of the one before. The plates are visualized as background.

quality images [8]. Whilst it is likely that experts will continue to be needed for plate reading, a high-quality second reading will be key to ensuring accuracy and efficiency, especially when processing a large number of plates. Here we explore whether BashTheBug and AMyGDA can be improved upon.

Machine learning has demonstrated itself to be well-suited to the interpretation of images, although detecting bacterial growth on microtiter plates poses a challenge. Any method must distinguish between the growth of *M. tuberculosis* and plate errors such as contamination, condensation, sediment, or air bubbles, whilst not being misled by artefacts such as shadows and water droplets that form on the protective film due to humidity. In this study, TMAS, a deep learning-based framework with state-of-the-art models trained on the CRyPTIC dataset to automatically detect *M. tuberculosis* growth on 96-well plates, is developed and evaluated. It is benchmarked against AMyGDA and BashTheBug, which are established methods used in CRyPTIC, to determine whether it can provide more consistent and scalable interpretations, particularly in cases where traditional or semi-automated methods may be prone to misclassification. We harness a series of advances in deep learning-based visual object recognition, and object detection specifically [11], to assess its applicability to *M. tuberculosis* pDST.

2. Materials and methods

2.1. Dataset

To develop and evaluate the TMAS framework, the publicly available dataset provided by the CRyPTIC Consortium was utilised, which contains extensive phenotypic data of *Mycobacterium tuberculosis* isolates [10]. This dataset comprises:

1. **Image Data:** A total of 15,209 images of 96-well broth microdilution plates captured after 14 (or in a few cases, 21) days of incubation using a *Thermo Fisher Vizion Instrument*. Each image shows an entire plate (referred to as a "plate image") and includes all 96 wells, each corresponding to different drug concentrations. The plates were set up in two designs:
   - **UKMYC5:** Contains 14 anti-TB drugs and two drug-free positive control wells.
   - **UKMYC6:** An updated design with 13 anti-TB drugs, excluding para-aminosalicylic acid (PAS), and extended concentration ranges for several drugs present in UKMYC5 (Fig. 1).
2. **Plate Readings and Ground Truth MICs:** The Minimum Inhibitory Concentration (MIC) values used in this study were derived from the *CRyPTIC* project through a robust consensus process. MIC readings were obtained from three complementary sources: (1) manual readings by expert laboratory technicians using the *Thermo Fisher Vizion* system, which served as a gold-standard reference; (2) automated analyses conducted with *AMyGDA*, a growth detection algorithm designed for reproducible plate image analysis; (3) crowd-sourced evaluations from the *BashTheBug* citizen science project, where non-expert participants analyzed the same images. These readings were integrated by CRyPTIC into a single, standardized consensus MIC value using a purpose-built algorithm [10]. This consensus value served as the ground truth for training and evaluating the performance of TMAS's growth detection model.

2.1.1. Image selection and partitioning

From the complete CRyPTIC dataset of 15,209 plate images, a subset of 4,018 images was randomly selected to train and evaluate TMAS. The selection of a subset of images was to ensure computational efficiency in training and testing the models. The images were chosen randomly to ensure a diverse representation of images, high-quality, low-quality, and challenging edge-case images.

The dataset of 4,018 images for experiments in this work was then partitioned into subsets for training, testing and evaluation purposes, in which the training and validation dataset is 80% and the test set is 20%.

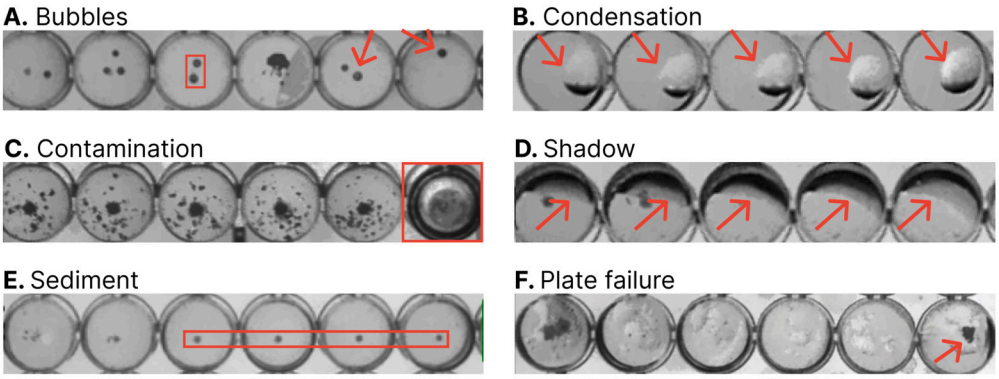


Fig. 2. Types of artefacts that can affect MIC determination and model performance, including **(A)** bubbles, **(B)** condensation, **(C)** contamination, **(D)** shadow, **(E)** sediment, and **(F)** plate failure.

1. **Training and Validation Set:** This set, consisting of 3,215 images (containing 41,850 MIC readings), was used for training and validating the deep learning models for growth detection.
   - **Training Subset:** 75% of the Training and Validation Set (2,411 images or 31,385 MIC readings)
   - **Validation Subset:** 25% of the Training and Validation Set (804 images or 10,465 MIC readings)

   To mitigate variability introduced by random sampling, the splitting and training process was repeated across five independent runs, with training and validation subsets reshuffled for each run. Model performance for growth detection was evaluated on the validation subsets, and the final results were reported as the mean and standard deviation across the five runs. This approach ensured a more reliable assessment of the growth detection model's generalizability while reducing potential biases introduced by specific data partitions.
2. **Test Set:** This comprehensive test set, consisting of 803 images (10,454 MIC readings), was used for evaluating TMAS's performance in predicting MICs.

2.1.2. Edge case dataset

In addition to the **Test Set**, a set of **100 edge-case images**, named the **Edge Case Dataset**, was included to assess TMAS's robustness under challenging conditions. These images were manually curated from the remaining images in the whole dataset of 15,209 plate images from CryPTIC, and they are characterized by severe artefacts and irregular growth patterns, contributed **1,324 MIC readings** to the overall dataset. Together, the Comprehensive Test Set and Edge Case Dataset were used to evaluate TMAS's MIC prediction performance across diverse conditions and scenarios.

2.2. Source of errors

In real-life environments, a range of artefacts can significantly impact the quality and clarity of images taken from 96-well plates, potentially affecting model performance. Systematic artefacts that may affect a region or an entire plate include shadows, sediment, and plate failures, which are challenging to detect and avoid. Randomly occurring artefacts, such as bubbles, condensation, or contamination, can lead to nonsensical growth patterns if erroneously detected. These include apparent bacterial proliferation at high antibiotic concentrations—where inhibition was expected—or unexpected absences of growth, causing non-continuous bacterial growth patterns [8]. The *TMAS framework* was specifically designed to mitigate the effects of these artefacts on MIC determination, representing a significant improvement over existing approaches like *AMyGDA*, which are more sensitive to such errors. See Fig. 2.

2.3. Software

TMAS was built based on advanced deep learning models, including Faster R-CNN [13], Mask R-CNN [14], Inception-ResNet [15], and YOLOv8 [16]. It was developed using object-oriented Python3 and implemented with Ultralytics YOLOv8 [17], Keras [18], TensorFlow [19], PyTorch [20] for bacterial growth detection. Image processing techniques were employed using the OpenCV2 Python API [21]. NumPy and Pandas were used to ensure efficient data management. TMAS reports MICs based on the specific plate design and offers the option of saving results in different formats (.csv, .json). TMAS can be installed directly from its PyPi package https://pypi.org/project/tmas using *pip install tmas* or via https://github.com/Oucru-Innovations/tmas.git.

2.4. Solution design

The TMAS framework includes image preprocessing, training and selecting the optimal deep learning model for bacterial growth detection, generating and evaluating outputs, and combining growth detection with plate design to report MICs for each drug. The overall architecture is presented in Fig. 3.

2.4.1. Image preprocessing

Image preprocessing was undertaken to deal with poor image quality, such as mitigating the effects of low contrast and uneven illumination in the acquired images. A mean shift filter [22] was employed to reduce image noise while preserving the integrity of object edges; the Contrast-Limited Adaptive Histogram Equalization (CLAHE) filter [23] was utilised for local contrast enhancement without excessive noise amplification; and pixel histogram normalisation [24] was implemented to optimise overall contrast. [8]. Each processed image is resized to 512×512 pixels and padded to preserve aspect ratio (Fig. 3A). This ensures uniform input to the deep learning model regardless of the original image size.

2.4.2. Growth detection using deep learning

To train the growth detection model, we have developed an annotation tool to retrieve consensus MIC values from the CRyPTIC dataset and link plate coordinates to areas of expected microbial growth within the images (see Fig. 4). This preparatory step allowed precise labelling of each image, and subsequent effective training in later stages of the analysis.

The performance of each candidate model was rigorously evaluated using standard metrics for bacterial growth detection, including Precision, Recall, F1 Score, and Mean Average Precision (mAP), along with computational efficiency indicators such as throughput. Based on these evaluations, the most optimal model was selected for integration into TMAS. To ensure the reliability of predictions, a confidence score threshold of 0.5 was applied during the evaluation phase; only detections with

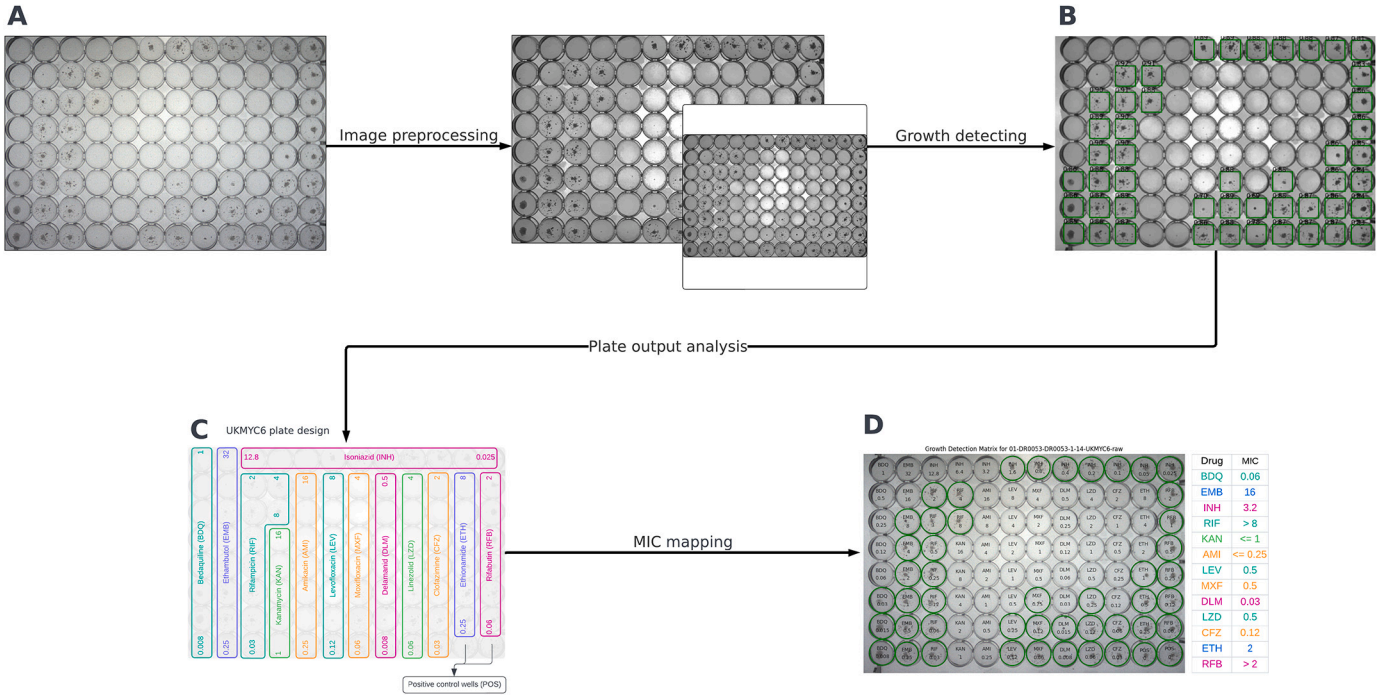
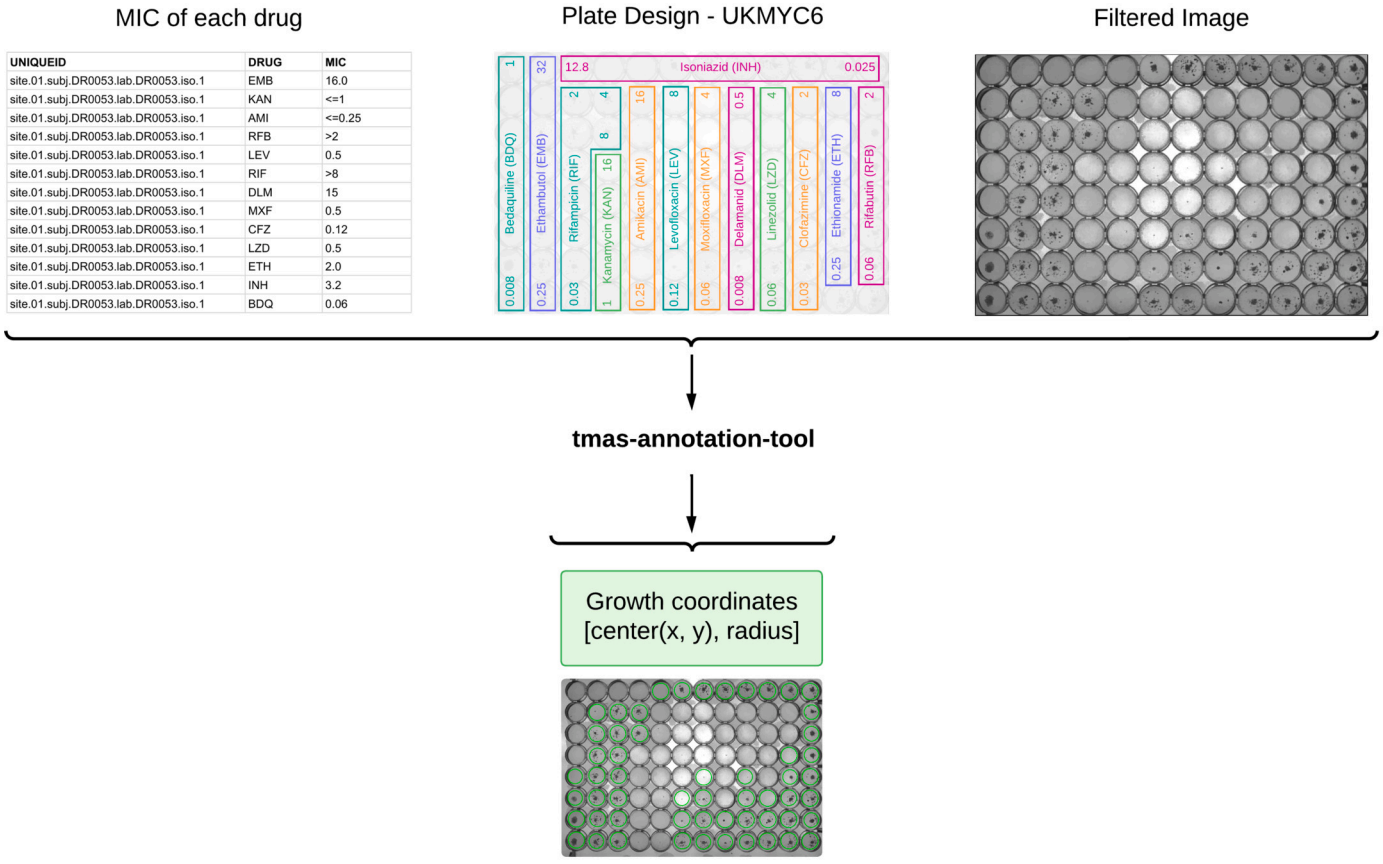


Fig. 3. The overall architecture of TMAS. The system consists of four steps to detect growth and determine the MICs of plate images. **(A)** Plate images are preprocessed by applying various image processing techniques. The images are then resized to 512×512 pixels and padded to fit the detection model's input. **(B)** A trained deep learning model is used to detect the bacterial growth in each well of the plates. **(C)** Drug-free, positive control wells are checked for the presence of growth. **(D)** MICs are reported based on growth detection and the corresponding plate design.



UNIQUEID	DRUG	MIC
site.01.subj.DR0053.lab.DR0053.iso.1	EMB	16.0
site.01.subj.DR0053.lab.DR0053.iso.1	KAN	<=1
site.01.subj.DR0053.lab.DR0053.iso.1	AMI	<=0.25
site.01.subj.DR0053.lab.DR0053.iso.1	RFB	>2
site.01.subj.DR0053.lab.DR0053.iso.1	LEV	0.5
site.01.subj.DR0053.lab.DR0053.iso.1	RIF	>8
site.01.subj.DR0053.lab.DR0053.iso.1	DLM	15
site.01.subj.DR0053.lab.DR0053.iso.1	MXF	0.5
site.01.subj.DR0053.lab.DR0053.iso.1	CFZ	0.12
site.01.subj.DR0053.lab.DR0053.iso.1	LZD	0.5
site.01.subj.DR0053.lab.DR0053.iso.1	ETH	2.0
site.01.subj.DR0053.lab.DR0053.iso.1	INH	3.2
site.01.subj.DR0053.lab.DR0053.iso.1	BDQ	0.06

Fig. 4. An illustration of annotation tool. The plate images were annotated using the developed tool by retrieving the growth information provided in the plate measurements to determine the growth coordinates in each image, which were then used as bounding boxes for model training.

Table 1
Performance comparison of different models, sorted by F1-score in descending order. Ips = images per second.

Models	Precision	Recall	F1-score	mAP	Throughput
YOLOv8	0.956 ± 0.005	0.954 ± 0.004	0.955 ± 0.004	0.982 ± 0.002	69 ips
Mask R-CNN	0.910 ± 0.030	0.969 ± 0.019	0.938 ± 0.011	0.952 ± 0.016	19 ips
Faster R-CNN	0.914 ± 0.029	0.961 ± 0.017	0.937 ± 0.008	0.970 ± 0.005	25 ips
Inception-ResNet	0.996 ± 0.001	0.707 ± 0.083	0.826 ± 0.058	0.794 ± 0.004	15 ips

Table 1 shows average performance across five runs. YOLOv8 achieved the highest F1-score, mAP, and throughput.

confidence scores ≥ 0.5 were considered valid, thereby ensuring a reasonable level of certainty in the results. Following detection, positive control wells outputs were verified (Fig. 3C). Images exhibiting discrepancies or errors were flagged for manual review to further ensure accuracy and robustness.

2.4.3. MIC determination

After bacterial growth detection, Minimum Inhibitory Concentrations (MICs) are determined from 96-well plate images as follows:

- **Well Identification:** Wells are detected using Hough Circle Transform in OpenCV [21] and represented as circular regions in a standardized 8×12 grid corresponding to the 96-well layout. Each well's center coordinates and radii are stored for precise spatial referencing.
- **Mapping Growth Predictions to Wells:** The deep learning model outputs a set of bounding boxes, each associated with:
  - **Coordinates:** The (x_1, y_1, x_2, y_2) values defining the rectangular region of detected growth,
  - **Scores:** The model's confidence in the prediction, and
  - **Labels:** A class label indicating the presence of growth.

  To map these detections to the correct wells, TMAS checks whether the center of each well lies within any predicted bounding box labeled as growth. If so, the well is marked as positive for bacterial growth. This step results in a binary growth detection matrix (8×12) where 1 indicates detected bacterial growth and 0 represents no growth.
- **MIC Extraction from Plate Design:** Each well corresponds to a specific drug and dilution level as defined in the CRyPTIC plate design. Using the binary growth matrix and plate metadata, TMAS constructs a per-drug growth profile across increasing concentrations. MICs are then determined according to the following rules:
  - The MIC is defined as the lowest concentration with no observed growth.
  - If growth is present in all wells of a drug, the MIC is reported as greater than the highest concentration available on the plate.
  - If no growth is detected in any well, the MIC is reported as less than or equal to the lowest concentration.

  Special cases, including skipped wells (non-continuous growth indicating possible anomalies) and invalid positive controls (e.g., lack of growth in drug-free wells), are flagged. Such wells are excluded from downstream analysis and essential agreement calculations to ensure reliability. As part of the output, visualizations of the bacterial growth detection on plate images can be optionally provided alongside the MIC results for all drugs tested (Fig. 3D).

2.4.4. Performance evaluation

To evaluate TMAS's performance, we tested it on a comprehensive test set of 803 images, each paired with a corresponding ground truth image. For each drug in a test image, TMAS's predicted MIC was compared to the ground truth MIC for the same drug. The evaluation focused on the agreement between TMAS predictions and ground truth values based on dilution levels. Essential agreement was defined as the percentage of MIC predictions that were either identical to or within one doubling dilution of the ground truth MIC. Out of 10,454 MIC reading pairs from the 803 plate images, 176 pairs were excluded due to missing ground truth values, resulting in 10,278 MIC reading pairs being analyzed. These comparisons formed the basis for calculating TMAS's overall essential agreement.

2.5. Experiment setup

All experiments were conducted on Google Colab using an NVIDIA L4 GPU and local hardware including a B760 Edge Ti WiFi motherboard, a Core i5-12400 CPU, 16 GB DDR5 RAM, an MPG A850GL power supply unit, a MEG S280 cooler, an RTX 4080 SUPRIM graphics card, a M390 500 GB SSD, a GUNGIR 300R case, and peripherals GK30 and GM20. The environment was configured using Python 3.10.12 and PyTorch 2.3.1, utilising CUDA 12.2 (compiler version 12.2.140).

3. Results

3.1. Evaluation of TMAS' growth detection model

Four advanced Deep Learning models (Faster R-CNN, Mask R-CNN, Inception-ResNet, and YOLOv8) were trained and evaluated, focusing on key performance metrics: precision, recall, F1-score, mean average precision (mAP) [25], and throughput (the number of images processed per second, ips). The results, averaged over five runs, are summarized in Table 1.

Among the evaluated models, YOLOv8 consistently outperformed the others across all metrics, achieving the highest F1-score, mAP, and throughput. Mask R-CNN and Faster R-CNN also demonstrated strong performance, but their F1-scores and throughput were slightly lower than YOLOv8. Inception-ResNet achieved the highest precision but suffered from low recall, which significantly impacted its F1-score and mAP.

Furthermore, an analysis of the learning behaviors of the models, based on loss graphs (provided in the supplementary section), revealed that YOLOv8 exhibited faster convergence and more stable loss reduction during training, indicating superior learning efficiency compared to the other models. This advantage translated into higher detection accuracy and reliability, affirming YOLOv8 as the best-performing model in this study.

To further validate YOLOv8, we quantified its inference time on the same testing dataset. With a throughput of 69 images per second, YOLOv8 was the most computationally efficient model, significantly surpassing the throughput of Mask R-CNN (19 ips), Faster R-CNN (25 ips), and Inception-ResNet (15 ips).

Based on these evaluations, YOLOv8 was selected as the optimal model for growth detection and subsequently integrated into TMAS for Minimum Inhibitory Concentration (MIC) extraction. An example of the visualized detection results is shown in Fig. 5B, alongside a comparison to the ground truth in Fig. 5A.

3.2. TMAS's MIC reading result on test set

The TMAS framework was evaluated for its ability to determine Minimum Inhibitory Concentrations (MICs) by integrating output from the bacterial growth detection model applied to 96-well plate images.

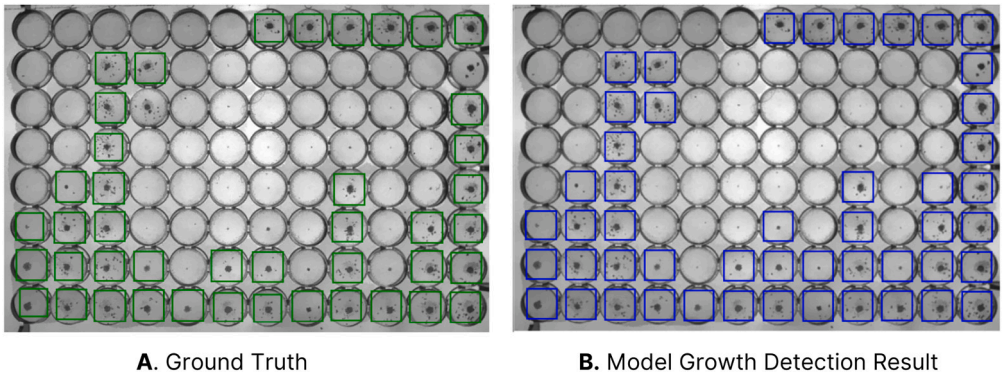


Fig. 5. Example of comparison of ground truth and YOLOv8 detection results on a plate image. **(A)** displays ground truth bacterial growth (green bounding boxes), while **(B)** illustrates YOLOv8 predictions (blue bounding boxes).

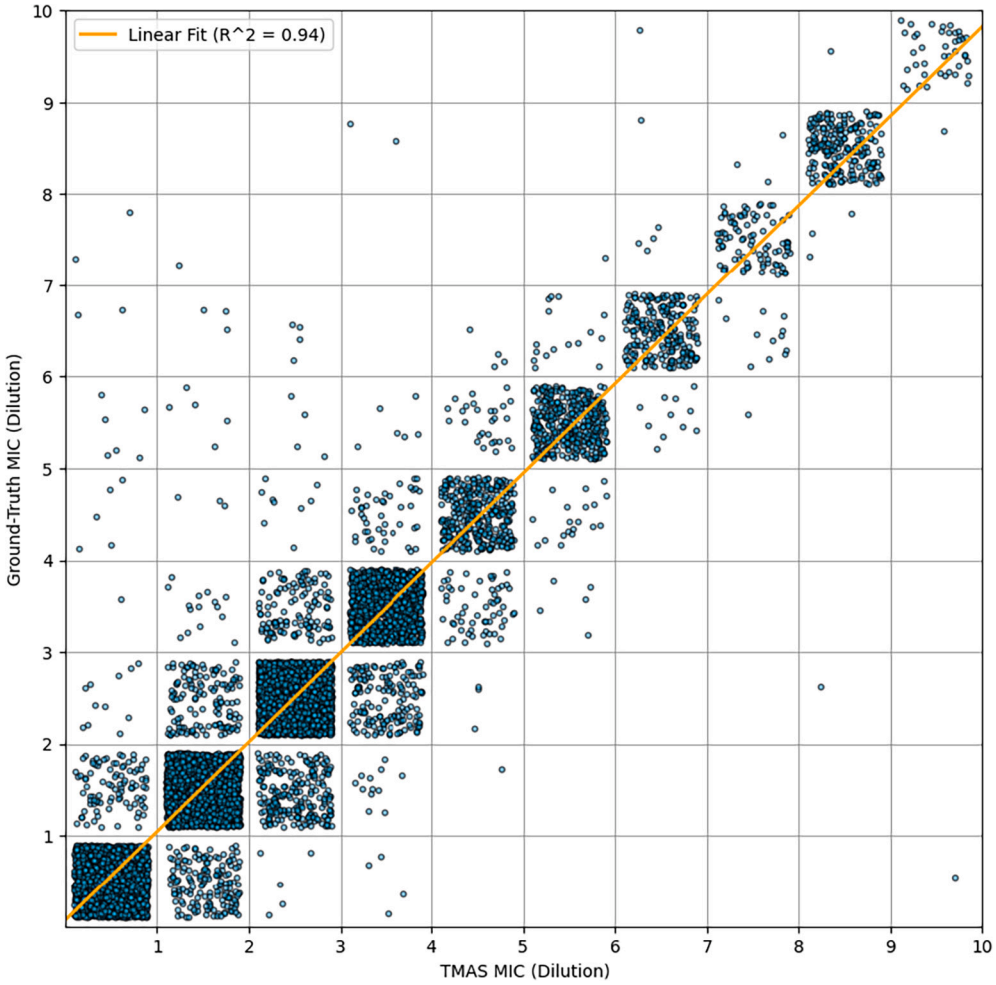


Fig. 6. Relationship between TMAS-predicted MIC (Dilution) and ground truth. Each point represents one of the 10,278 pairs of predicted and ground truth MIC dilutions, with a linear regression trend line ($R^2 = 0.94$) illustrating the strong agreement across all drugs. Variations in the number of wells for each drug may introduce minor biases in the distribution.

The concordance between TMAS's MIC predictions and the ground truth was high, as demonstrated by a strong linear relationship with an R^2 of 0.94 (Fig. 6). To ensure reliability in the essential agreement calculations, 63 additional MIC readings (0.61%) were excluded due to nonsensical growth patterns, such as the detection of bacterial growth in higher concentration wells but not in lower concentration wells. These cases, requiring manual review, were excluded to maintain the validity of the automated evaluation. This resulted in 10,215 MIC readings being used for essential agreement calculations.

TMAS achieved an essential agreement of 98.8% (10,096 out of 10,215 cases) with the ground truth. Of the 10,215 cases, TMAS predictions were identical to the ground truth for 9,009 cases (88.2%), one doubling dilution lower for 645 cases (6.31%), and one doubling dilution higher for 442 cases (4.33%). Predictions that differed by two or more doubling dilutions were classified as errors, accounting for 119 cases (1.16% error rate).

This performance significantly exceeds the 90% threshold established by the International Organization for Standardization (ISO) [26],

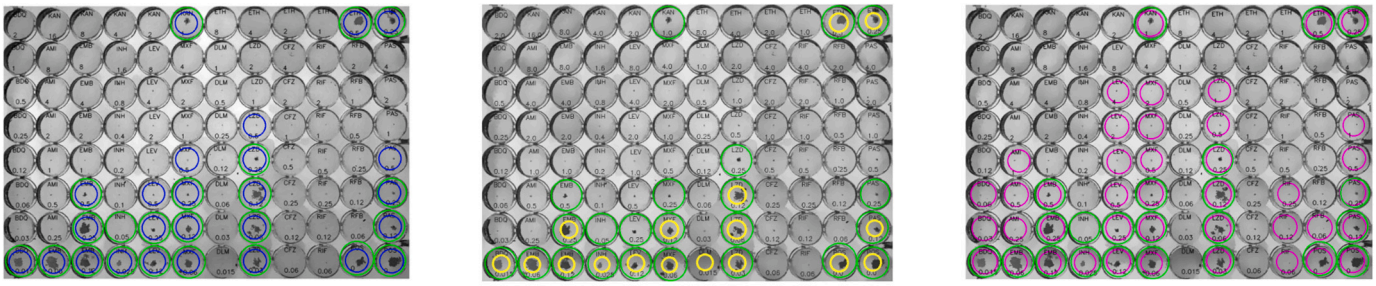


Fig. 7. Comparison of TMAS, AMyGDA, and BashTheBug growth detection results. The green circles indicate the Ground Truth data provided by the lab scientists. The left column **(A)** displays the TMAS's detection results, with bacterial growth marked by blue circles. The middle column **(B)** displays the AMyGDA detection results, with bacterial growth marked by yellow circles. The last column **(C)** represents the BashTheBug detection results, with bacterial growth marked by pink circles.

Table 2
Comparison of MIC values predicted by TMAS, AMyGDA, and BashTheBug against the MIC measured by laboratory scientists on the comprehensive test set.

Description	Method	Number of Cases	Proportion of Cases
Unable to determine an MIC	TMAS	0	0%
	AMyGDA	725	6.94%
	BashTheBug	605	5.79%
Identical to Lab-Scientist Reading	TMAS	8,738	83.6%
	AMyGDA	6,327	60.5%
	BashTheBug	4,692	44.9%
One doubling dilution lower than Lab-Scientist Reading	TMAS	670	6.41%
	AMyGDA	273	2.61%
	BashTheBug	1,726	16.5%
One doubling dilution higher than Lab-Scientist Reading	TMAS	838	8.02%
	AMyGDA	2,507	24.0%
	BashTheBug	118	1.13%
Two or more doubling dilutions different from Lab-Scientist Reading	TMAS	208	1.99%
	AMyGDA	622	5.95%
	BashTheBug	3,313	31.7%

Compared to 87.1% (9,107 out of 10,454 cases) of AMyGDA and 62.5% (6,536 out of 10,454 cases) of BashTheBug, TMAS achieved the highest score of 98.0% (9,881 out of 10,454 cases) agreement with Lab-Scientist readings, with MIC values within one doubling dilution of each other.
Note: *The sum of the "Proportion of Cases" for TMAS & BashTheBug is 100.02% due to rounding to two decimal places.*

which is commonly used to assess and validate new drug susceptibility testing (DST) methods. These results demonstrate the reliability and accuracy of TMAS in determining MICs under diverse testing conditions.

3.3. TMAS, AMyGDA, and BashTheBug performance on test set compared to Lab-scientist readings

To compare TMAS with AMyGDA and BashTheBug, MICs determined by the laboratory-based expert reader were used as the benchmark. This approach provides an independent reference point since the ground truth contains results from AMyGDA and BashTheBug, making it unsuitable for objective evaluation. Each method's ability to detect bacterial growth was assessed across two test sets: the comprehensive test set and the edge case dataset with plate artefacts.

As observed in Fig. 7, the results from TMAS and AMyGDA are relatively similar. Both methods successfully detected limited growth in wells containing EMB, LEV, MXF, LZD, PAS, and KAN. However, the bacterial growth detected by TMAS was consistently reported at least one doubling dilution higher than that observed by laboratory scientists. AMyGDA occasionally failed to detect the same level of growth, resulting in MICs that were one dilution lower than those determined by laboratory scientists. This led to discrepancies in MICs compared to the Lab-Scientist Reading, although results were still within essential agreement. The consensus of the BashTheBug volunteers was less accurate at detecting bacterial growth. In some cases, it failed to identify growth detected by TMAS and AMyGDA, leading to a higher discrepancy in its reported MICs. See Table 2.

TMAS achieved the greatest concordance with 83.6% of its readings exactly matching those of lab scientists, compared to 60.5% of AMyGDA's readings and 44.9% of BashTheBug's. Moreover, TMAS's discrepancy rates for readings one doubling dilution off from Lab-Scientist readings stood at 14.43% (6.41% one dilution lower + 8.02% one dilution higher), being lower than AMYGDA's 26.61% (2.61% + 24.0%) and BashTheBug's 17.63% (16.5% + 1.13%). AMYGDA's performance suffers from its reliance on structured scenarios, and it cannot adapt to handle low-quality or distorted images that result from plate artefacts. Only 1.99% of TMAS readings differed by two or more doubling dilutions, compared to 5.95% for AMYGDA and 31.7% for BashTheBug.

3.4. TMAS, AMYGDA, BashTheBug performance on edge case dataset compared to Lab-scientist reading

Next we compared the performance of TMAS, AMyGDA, and BashTheBug on hard-to-read plates containing artefacts. Lab-scientist readings again served as the ground truth benchmark for this comparison. 13 out of 1,324 readings from 100 images were excluded due to the lack of ground truth, resulting in 1,311 in total. See Table 3.

As an example in (Fig. 8), TMAS (Fig. 8A) and BashTheBug (Fig. 8C) performed well in detecting bacterial growth among these edge cases.

Table 3
Comparison of MIC values measured by TMAS, AMyGDA, and BashTheBug against the MIC measured by laboratory scientists on the edge case dataset.

Description	Method	Number of Cases	Proportion of Cases
Unable to determine an MIC	TMAS	0	0%
	AMyGDA	267	20.4%
	BashTheBug	278	21.2%
Identical to Lab-Scientist Reading	TMAS	816	62.2%
	AMyGDA	574	43.8%
	BashTheBug	483	36.8%
One doubling dilution lower than Lab-Scientist Reading	TMAS	236	18.0%
	AMyGDA	74	5.64%
	BashTheBug	148	11.3%
One doubling dilution higher than Lab-Scientist Reading	TMAS	160	12.2%
	AMyGDA	269	20.5%
	BashTheBug	18	1.37%
Two or more doubling dilutions different from Lab-Scientist Reading	TMAS	99	7.55%
	AMyGDA	127	9.69%
	BashTheBug	384	29.3%

TMAS achieved the highest agreement at 92.4% (1,212 out of 1,311 cases), compared to 69.9% (917 out of 1,311 cases) for AMyGDA and 49.5% (649 out of 1,311 cases) for BashTheBug, with agreement defined as MIC values that are identical to or within one doubling dilution of the Lab-Scientist readings.
Note: *Due to rounding to two decimal places, the total proportion of cases sums to 99.95% for TMAS, 99.97% for BashTheBug, and 100.03% for AMyGDA.*

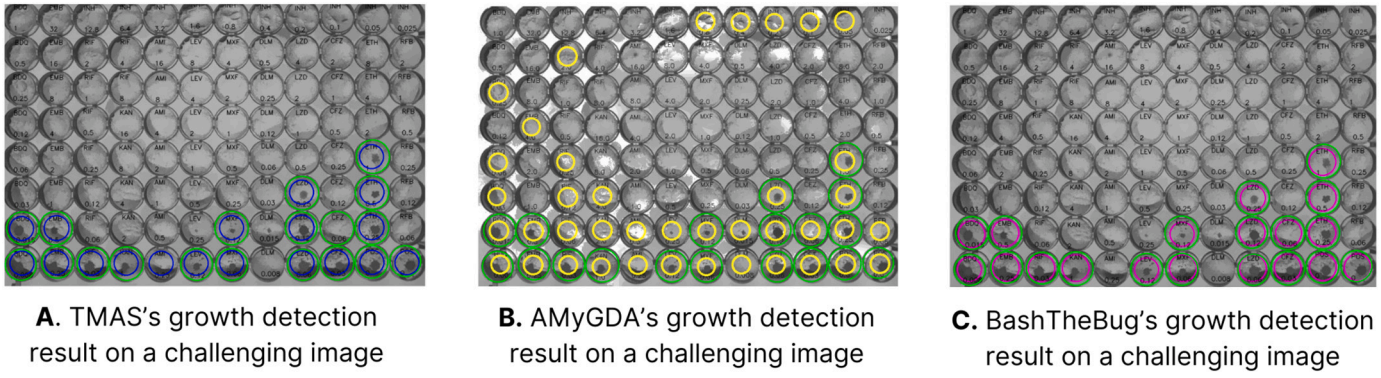


Fig. 8. Comparison of TMAS and AMyGDA growth detection results on challenging images. The Ground Truth data provided by the lab scientists is indicated by the green circles. The left-hand plate image A) displays the TMAS's detection results, with bacterial growth marked by blue circles. The middle plate image B) displays the AMyGDA detection results, with bacterial growth marked by yellow circles. The right-hand plate image C) represents the BashTheBug detection results, with bacterial growth marked by pink circles.

While BashTheBug perfectly aligned with lab-scientist readings in this case, TMAS exhibited one false positive reading for AMI and one false negative reading for CFZ. AMyGDA (Fig. 8B) frequently misclassified condensation as bacterial growth, leading to a high rate of incorrect MIC readings. As it relies on crowd-sourced inputs, BashTheBug is particularly susceptible to errors arising from misinterpretations of visual artefacts, such as mistaking sediment for microbial growth or condensation patterns, which increases the risk of inaccuracies in MIC determination.

From the comparison of TMAS, AMyGDA, and BashTheBug on the edge case test set, we can see that TMAS consistently demonstrated the closest correlation with the Lab-Scientist Readings, achieving a match in 62.2% of cases, compared to 43.8% of AMyGDA's readings and 36.8% of BashTheBug's. It also demonstrated the least discrepancies, with only 7.55% of cases differing by two or more dilutions, compared to 9.69% for AMyGDA and 29.3% for BashTheBug. Additionally, TMAS achieved the lowest error rate, with no MICs undetermined, compared to error rates of 20.4% for AMyGDA and 21.2% for BashTheBug. TMAS is therefore more accurate, reliable, and robust than existing methods, even when presented with challenging edge cases.

4. Discussion

Various methods have been employed for *Mycobacterium tuberculosis* (MTB) growth detection and susceptibility testing, each with unique trade-offs in cost, turnaround time, and scalability. The Mycobacterium Growth Detector Tubes (MGIT) 960 system offers full automation and high sensitivity using fluorescence-based oxygen consumption detection, delivering results within 7 to 14 days [27] and achieving up to 99.5% agreement with reference methods. This is the most widely used method but constitutes an expensive approach to generating MICs for multiple drugs as once. The Agar Proportion Method (APM), regarded as a gold standard, relies on visible colony formation on solid media but is labor-intensive and has a long turnaround time of up to 21 days [28]. The Resazurin Microtiter Assay (REMA) is a colorimetric approach that detects bacterial growth through a redox-induced color change, typically within 8 days. It is affordable, can generate MICs, is suited to high-throughput applications, and is capable of detecting resistant subpopulations at levels as low as 1% to 4% [21]. Microscopic-observation drug-susceptibility (MODS) assays similarly lend themselves to measuring MICs and have a rapid turn-around time [29]. However, some biosafety concerns around these methods still persist and the reading of

all these plates remains subjective. An automated approach is therefore preferred. Imaging-based systems can mitigate some of the subjectivity that accompanies human reading of these assays. They also have the potential for earlier detection, with studies reporting a 4.4-fold reduction in time to detection using high-resolution imaging compared to standard protocols, while maintaining comparable accuracy [30]. Moreover, such systems support high-throughput screening and automated MIC interpretation across large datasets [8].

Inspired by these advancements, we developed the Tuberculosis Microbial Analysis System (TMAS), an automated framework based on deep learning for detecting bacterial growth from images of 96-well microtiter plates and determining Minimum Inhibitory Concentrations (MICs). TMAS addresses a critical need in TB diagnostics by providing a consistent, and automated method that reduces reliance on manual reading and mitigates variability associated with human and semi-automated approaches. Microtitre plates are an efficient means of simultaneously measuring MICs for multiple drugs, and cost less than traditional DST methods, such as the Mycobacterial Growth Indicator Tube (MGIT), would if used to assay MICs for this many drugs. TMAS leverages advanced deep learning algorithms to accurately read 96-well broth microdilution plates. On the comprehensive test set, TMAS achieved an essential agreement with the ground truth of 98.8%, significantly exceeding the 90% minimum required for drug susceptibility testing methods by ISO 20776-2:2021 [26]. Furthermore, TMAS was able to interpret challenging images, maintaining high agreement rates despite the presence of artefacts and low quality images. It also achieved a trade-off between computational efficiency and accuracy, with an average inference time of 14.3 milliseconds and a throughput of 69 images per second, ensuring its practical applicability.

TMAS offers more than just technical improvements. By delivering fast and accurate MIC readings, it has the potential to directly support improved clinical decision-making, enabling earlier and more accurate identification of effective drug regimens. This could reduce instances of false resistance or delayed treatment, which are key obstacles in the management of drug-resistant tuberculosis. In research and surveillance contexts, TMAS could streamline the extraction of phenotypic data from large image datasets, enabling more efficient analysis of emerging resistance patterns.

TMAS addresses several critical limitations of earlier systems such as AMyGDA and BashTheBug, offering notable improvements in accuracy, robustness, and reliability. Our evaluations demonstrated that TMAS consistently outperformed AMyGDA and BashTheBug across multiple metrics. On the comprehensive test set, TMAS achieved 83.6% agreement with lab-scientist readings, compared to 60.5% for AMyGDA and 44.9% for BashTheBug. Additionally, TMAS reported discrepancies (one doubling dilution off) in only 14.4% of cases, lower than AMyGDA's 26.6% and BashTheBug's 17.6%. TMAS also exhibited the lowest error rate (1.99%) for readings that were two or more dilutions different from the ground truth, compared to 5.95% for AMyGDA and 31.7% for BashTheBug.

BashTheBug, a crowd-sourced citizen science platform, exhibited variability in performance due to its reliance on non-expert contributors. While the "wisdom of the crowd" approach helps average out individual errors, the project's high Gini coefficient indicates an uneven distribution of classifications, with a small subset of volunteers contributing disproportionately more than others [9]. However, as each volunteer could only contribute a single measurement per MIC, the overall impact of this imbalance was limited. A more critical limitation of BashTheBug was that volunteers tended to be conservative in their classifications, often overcalling bacterial growth as a precautionary measure. This tendency also increased the likelihood of misclassifying contamination or other artefacts as bacterial growth, leading to inflated MIC values.

Similarly, AMyGDA, an automated image-processing-based tool, exhibited some limitations due to its reliance on simple techniques [8]. AMyGDA primarily detects bacterial growth by counting dark pixels in the center of wells, rendering it highly susceptible to artefacts. Unlike deep learning-based approaches, which can learn image patterns to distinguish between true bacterial growth and artefacts, AMyGDA requires manual parameter tuning to mitigate false positives. This often results in undercalling MICs compared to laboratory scientists' readings, as a conservative threshold is applied to minimize erroneous classifications. Although AMyGDA employs mean shift filtering (MSF) and contrast-limited adaptive histogram equalization (CLAHE)—techniques also utilized in TMAS for preprocessing—these methods alone are insufficient to accurately detect bacterial growth in low-quality images. By contrast, TMAS's deep learning-based approach leverages learned feature representations to robustly differentiate growth from artefacts across diverse imaging conditions, enhancing both accuracy and reliability.

TMAS demonstrated its robustness when evaluated on an edge case dataset characterized by severe artefacts and irregular growth patterns. TMAS achieved 62.2% agreement with lab-scientist readings, outperforming AMyGDA (43.8%) and BashTheBug (36.8%). Additionally, TMAS reported discrepancies (one doubling dilution off) in 31.6% of cases, compared to AMyGDA's 28.6% and BashTheBug's 22.7%. TMAS also exhibited the lowest rate of major discrepancies, with only 7.55% of cases differing by two or more dilutions, compared to 9.69% for AMyGDA and 29.3% for BashTheBug. These results emphasize TMAS's ability to minimize prediction errors and maintain high performance under various conditions, making it a reliable tool for applications in resource-limited settings.

From a deployment perspective, TMAS could operate as a digital assistant to laboratory scientists, automating initial plate interpretation while minimizing errors and maintaining expert oversight for clinical validation. In more advanced implementations, TMAS could be integrated into real-time diagnostic workflows, where images of plates are captured and uploaded directly to participant records for immediate analysis. This would not only streamline high-throughput testing pipelines but also enable automated flagging of discrepancies and ensure traceable decision-making.

While TMAS could, in principle, support dynamic growth monitoring through time-series imaging during incubation, such protocols present practical challenges, including increased labor for plate handling, contamination risks, and reliance on expensive hardware such as the Thermo Fisher Vizion system, factors that may restrict accessibility in resource-limited settings. To address these challenges, a more integrated setup could be envisioned, wherein plates remain stationary under a mounted camera within a controlled environment. This design would enable automated, scheduled image capture without requiring manual plate movement, further minimizing labor demands and reducing contamination risk. Alternatively, mobile-based solutions could eliminate the need for specialized hardware altogether. However, such an approach would require substantial model optimization, the development of a dedicated app interface, and potentially the deployment of cloud-based infrastructure to support inference if on-device processing proves infeasible.

While the primary focus of this study is the development of a high-accuracy deep-learning based framework for MIC determination, we recognize the value of a user-friendly web-based interface for broader accessibility. A full-featured web platform could be developed in the future to facilitate wider adoption of TMAS in clinical and research environments. This work would require additional engineering efforts and could form a future project.

Several aspects to consider around developing and using TMAS. First, while TMAS was developed for UKMYC5 and UKMYC6 plate layouts, it could potentially function on modified plates with additional wells or altered drug ranges because the main task remains to detect growth. However, achieving optimal accuracy on substantially altered designs may require adjusting the annotation process, updating the plate design map within TMAS, and potentially retraining certain aspects of the model. Second, TMAS currently produces a binary (growth vs. no growth) classification within each well based on bounding boxes for the growth detection phase. Although this approach effectively localizes

bacterial growth, implementing additional post-processing to quantify partial or percentage growth could offer further granularity for research or clinical purposes. Finally, adapting TMAS to novel platforms—such as mobile devices or alternative imaging setups—would require further development to ensure consistent model performance, sufficient computational capacity, and stable lighting conditions.

A notable strength of TMAS is its demonstrated ability to generalize across diverse conditions, as evidenced by its performance on both a broad comprehensive test set and a curated "edge-case" dataset. The edge-case dataset, characterized by severe artefacts and irregular growth patterns, further challenged the model beyond routine scenarios. Future developments could further enhance TMAS's capabilities. These include gathering images from multiple imaging systems aside from the Thermo Fisher Vizion instrument, refining the detection pipeline to account for partial or percentage growth, and exploring integrated incubator–camera workflows that automate imaging. Additionally, customizing the model's loss function to accommodate varying phenotype quality (e.g., high, medium, or low) could improve its predictive accuracy in heterogeneous datasets.

5. Conclusion

In this study, we developed the Tuberculosis Microbial Analysis System (TMAS), a machine learning-based tool designed to automate MIC determination from plate images. Built on advanced deep learning models trained on a subset of 4,018 images from the CRyPTIC dataset, TMAS can learn to detect bacterial growth with high precision, distinguishing true growth from artefacts such as shadows, bubbles, sediment, condensation and contamination. TMAS achieved an essential agreement of 98.8% with the ground truth, significantly exceeding the 90% minimum required for drug susceptibility testing. It also outperformed existing automated methods in accuracy, reliability, and efficiency, demonstrating robust performance even with challenging edge-case datasets. By reducing reliance on manual interpretation, TMAS streamlines TB diagnostics, improves laboratory efficiency, and enhances access to high-quality DST. Its high accuracy and scalability position TMAS as a valuable tool for clinical microbiology laboratories and research institutions addressing multidrug-resistant tuberculosis.

CRediT authorship contribution statement

Hoang-Anh T. Vo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Sang Nguyen:** Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis. **Ai-Quynh T. Tran:** Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Funding acquisition. **Han Nguyen:** Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis. **Hai Bich Ho:** Supervision, Conceptualization. **Philip W. Fowler:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Timothy M. Walker:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Funding acquisition, Conceptualization. **Thuy Thi Nguyen:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: PWF receives consultancy fees from the Ellison Institute of Technology, Oxford.

Acknowledgements

We would like to express our gratitude to NVIDIA and MSI for providing powerful components that greatly assisted our experiments with various state-of-the-art deep learning models. We extend our thanks to RMIT University Vietnam for providing the budget for cloud services for the experiments. TMW is a Wellcome Trust Clinical Career Development Fellow (214560/Z/18/Z). Funded by Wellcome (106680/B/14/Z). PWF would like to acknowledge funding from the National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Healthcare Associated Infections and Antimicrobial Resistance at Oxford University in partnership with UK Health Security Agency [HPRU-2012-10041], the National Institute for Health Research (NIHR) and Oxford Biomedical Research Centre (BRC). The findings and conclusions in this report are solely the responsibility of the authors and do not necessarily represent the official views of the NHS, the NIHR or the Department of Health and Social Care.

A preliminary version of this work was previously posted as a preprint on bioRxiv (doi: 10.1101/2025.02.14.638231).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csbj.2025.05.030.

Fig. S1. Loss graph for different proposed models. Analysis of the training and validation loss graphs for each state-of-the-art model, including YOLOv8, FasterRCNN, MaskRCNN, Inception-ResNet, represented by blue and orange lines, respectively.

References

[1] World Health Organization. Global tuberculosis report 2023. Geneva: World Health Organization. ISBN 9789240083851, 2023. Available from: https://www.who.int/publications/i/item/9789240083851.
[2] Dartois Véronique A, Rubin Eric J. Anti-tuberculosis treatment strategies and drug development: challenges and priorities. Nat Rev, Microbiol Nov. 2022;20(11):685–701. https://doi.org/10.1038/s41579-022-00731-y. ISSN: 1740-1526, 1740-1534.
[3] Van Rie Annelies, et al. Balancing access to BPaLM regimens and risk of resistance. Lancet Infect Dis Oct. 2022;22(10):1411–2. https://doi.org/10.1016/S1473-3099(22)00543-6.
[4] Rancoita Paola MV, et al. Validating a 14-drug microtiter plate containing bedaquiline and delamanid for large-scale research susceptibility testing of Mycobacterium tuberculosis. Antimicrob Agents Chemother Sept. 2018;62(9):e00344. https://doi.org/10.1128/AAC.00344-18. ISSN: 0066-4804, 1098-6596.
[5] WHO Global Tuberculosis Programme (GTB). Use of targeted next-generation sequencing to detect drug-resistant tuberculosis: rapid communication. Available from: https://www.who.int/publications/i/item/9789240076372, July 2023.
[6] Friedman Jerome, Hastie Trevor, Tibshirani Robert. Statistical challenges in the analysis of DNA methylation data. PLoS Biol 2021;19(5):e3001721. https://doi.org/10.1371/journal.pbio.3001721. Available from: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001721.
[7] Walker Timothy M, et al. The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: a genotypic analysis. Lancet Microbe Apr. 2022;3(4):e265–73. https://doi.org/10.1016/S2666-5247(21)00301-3.
[8] Fowler Philip W, et al. Automated detection of bacterial growth on 96-well plates for high-throughput drug susceptibility testing of Mycobacterium tuberculosis. Microbiology Dec. 2018;164(12):1522–30. https://doi.org/10.1099/mic.0.000733. ISSN: 1350-0872, 1465-2080.
[9] Fowler Philip W, et al. A crowd of BashTheBug volunteers reproducibly and accurately measure the minimum inhibitory concentrations of 13 antitubercular drugs from photographs of 96-well broth microdilution plates. eLife May 2022;11:e75046. https://doi.org/10.7554/eLife.75046.
[10] The CRyPTIC Consortium. A data compendium associating the genomes of 12,289 Mycobacterium tuberculosis isolates with quantitative resistance phenotypes to 13 antibiotics. PLoS Biol Aug. 2022;20(8):e3001721. https://doi.org/10.1371/journal.pbio.3001721. Ed. by Jason Ladner.
[11] Sarker Iqbal H. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci Nov. 2021;2(6):420. https://doi.org/10.1007/s42979-021-00815-1. ISSN: 2662-995X, 2661-8907.
[12] The CRyPTIC Consortium, et al. Epidemiological cutoff values for a 96-well broth microdilution plate for high-throughput research antibiotic susceptibility testing of M.

tuberculosis. Available from: https://doi.org/10.1101/2021.02.24.21252386, Mar. 2021.
[13] Ren Shaoqing, et al. Faster R-CNN: towards real-time object detection with region proposal networks. Version number: 3. Available from: https://doi.org/10.48550/ARXIV.1506.01497, 2015.
[14] He Kaiming, et al. Mask R-CNN. Version number: 3. Available from: https://doi.org/10.48550/ARXIV.1703.06870, 2017.
[15] Szegedy Christian, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning. Version number: 2. Available from: https://doi.org/10.48550/ARXIV.1602.07261, 2016.
[16] Redmon Joseph, et al. You only look once: unified, real-time object detection. Available from: arXiv:1506.02640 [cs], May 2016. https://doi.org/10.48550/arXiv.1506.02640.
[17] Ultralytics: YOLO vision. Available from: https://github.com/ultralytics/ultralytics, 2024 (visited on 06/01/2024).
[18] Chollet François. Deep learning with Python. second edition. New York: Manning Publications Co. LLC. ISBN 978-1-61729-686-4, 2021.
[19] Abadi Martín, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. Available from: arXiv:1603.04467 [cs], Mar. 2016. https://doi.org/10.48550/arXiv.1603.04467.
[20] Paszke Adam, et al. PyTorch: an imperative style, high-performance deep learning library. Available from: arXiv:1912.01703 [cs], Dec. 2019. https://doi.org/10.48550/arXiv.1912.01703.
[21] Zelinsky Alex. Learning OpenCV—computer vision with the OpenCV library (Bradski, G.R. et al.; 2008) [on the shelf]. IEEE Robot Autom Mag Sept. 2009;16(3):100. https://doi.org/10.1109/MRA.2009.933612.
[22] Comaniciu D, Meer P. Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell May 2002;24(5):603–19. https://doi.org/10.1109/34.1000236.
[23] Zuiderveld Karel J, et al. Contrast limited adaptive histogram equalization. In: Graphics gems 4.1. Boston, MA, USA: Publisher: Academic Press; 1994. p. 474–85.
[24] Gonzalez Rafael, Faisal Zahraa. Digital image processing. second edition. Prentice Hall; June 2019.
[25] Szeliski Richard. Computer vision: algorithms and applications. Springer Nature; 2022.
[26] International Organization for Standardization. ISO 20776-2:2021 clinical laboratory testing and in vitro diagnostic test systems — susceptibility testing of infectious agents and evaluation of performance of antimicrobial susceptibility test devices part 2: evaluation of performance of antimicrobial susceptibility test devices against reference broth micro-dilution. Tech. Rep. ISO 20776-2:2021. Geneva, Switzerland: ISO; 2021.
[27] Hanna Bruce A, et al. Multicenter evaluation of the BACTEC MGIT 960 system for recovery of Mycobacteria. J Clin Microbiol Mar. 1999;37(3):748–52. https://doi.org/10.1128/JCM.37.3.748-752.1999. ISSN: 0095-1137, 1098-660X. Available from: https://journals.asm.org/doi/10.1128/JCM.37.3.748-752.1999 (visited on 04/28/2025).
[28] Woods Gail L. Susceptibility testing of mycobacteria, nocardiae, and other aerobic actinomycetes: approved standard. CLSI document volume 31, number 5. second edition. Wayne, PA: Clinical and Laboratory Standards Institute. ISBN 978-1-56238-746-4, 2011.
[29] Moore David AJ, et al. Microscopic-observation drug-susceptibility assay for the diagnosis of TB. N Engl J Med Oct. 2006;355(15):1539–50. https://doi.org/10.1056/NEJMoa055524. ISSN: 0028-4793, 1533-4406. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa055524 (visited on 04/28/2025).
[30] Ghodbane Ramzi, et al. Rapid diagnosis of tuberculosis by real-time high-resolution imaging of Mycobacterium tuberculosis colonies. J Clin Microbiol Aug. 2015;53(8):2693–6. https://doi.org/10.1128/JCM.00684-15. Ed. by G.A. Land, ISSN: 0095-1137, 1098-660X. Available from: https://journals.asm.org/doi/10.1128/JCM.00684-15 (visited on 04/28/2025).

Computational and Structural Biotechnology Journal

Deep learning-based framework for Mycobacterium tuberculosis bacterial growth detection for antimicrobial susceptibility testing

Hoang-Anh T. Vo, Sang Nguyen, Ai-Quynh T. Tran, Han Nguyen, Hai Bich Ho, Philip W. Fowler, Timothy M. Walker, and Thuy Thi Nguyen

Citation: Vo H, Nguyen S, Tran A, Nguyen H, Ho H, Fowler P, Walker T, Nguyen T. Deep learning-based framework for Mycobacterium tuberculosis bacterial growth detection for antimicrobial susceptibility testing. *Comput Struct Biotechnol J.* 2025;**27**:2208-2218. DOI: 10.1016/j.csbj.2025.05.030

Abstract

Tuberculosis (TB) kills more people annually than any other pathogen. Resistance is an ever-increasing global problem, not least because diagnostics remain challenging and access limited. 96-well broth microdilution plates offer one approach to high-throughput phenotypic testing, but they can be challenging to read. Automated Mycobacterial Growth Detection Algorithm (AMyGDA) is a software package that uses image processing techniques to read plates, but struggles with plates that exhibit low growth or images of low quality. We developed a new framework, TMAS (TB Microbial Analysis System), which leverages state-of-the-art deep learning models to detect *M. tuberculosis* growth in images of 96-well microtiter plates. TMAS is designed to measure Minimum Inhibitory Concentrations (MICs) rapidly and accurately while differentiating between true bacterial growth and artefacts. Using 4,018 plate images from the CRyPTIC (Comprehensive Resistance Prediction for Tuberculosis: An International Consortium) dataset to train models and refine the framework, TMAS achieved an essential agreement of 98.8%, significantly outperformed the 90% threshold established by the International Organization for Standardization (ISO). TMAS offers a reliable, automated and complementary evaluation to support expert interpretation, potentially improving accuracy and efficiency in tuberculosis drug susceptibility testing (DST).

Graphical abstract