Automating Antibiotic Susceptibility Testing with Machine Learning for Disk Diffusion Test Analysis

Jakub Lepik
Department of Telecommunications
Brno University of Technology
Brno, Czech Republic
239122@vut.cz

Michal Cicatka
Department of Telecommunications
Brno University of Technology
Brno, Czech Republic
0000-0002-8119-6679

***Abstract*—Rapid and reliable antibiotic susceptibility testing (AST) methods are imperative in response to the escalating challenges of antimicrobial resistance. This study focuses on enhancing disk diffusion testing, a cornerstone of AST, by integrating machine learning and automation. Leveraging state-of-the-art object detection models, including EfficientDet and Mask R-CNN and image-processing approaches, our methodology addresses the need for standardized evaluation processes across diverse laboratory equipment while enabling the integration of mobile devices into the workflow, democratizing AST, and enhancing its accessibility. We utilize a comprehensive disk diffusion dataset for object detection models captured by devices like mobile phones and professional solutions. Additionally, our experiments lay the groundwork for a web application adopting a device-agnostic approach, promising improved accessibility and efficiency in AST analysis.**

***Index Terms*—antibiotic sensitivity testing, disk diffusion test, machine learning, image processing**

I. INTRODUCTION

In our pursuit of understanding microbial responses to antibiotics, we turn to the power of algorithms and machine learning (ML). Evaluating antibiotic effectiveness against microbial cultures is a crucial aspect of modern medical research, particularly with the rise of antibiotic-resistant bacteria, which presents a significant challenge to global health. Antibiotic susceptibility testing (AST) remains a cornerstone in combating antimicrobial resistance, offering essential guidance for antibiotic prescription.

The need to develop rapid AST methods has become increasingly apparent considering recent advances, such as those explored in [1]. Automation methods and ML, in particular, have seen substantial progress across multiple domains, including clinical diagnostics and pharmaceutical research. However, their utilization in pivotal disk diffusion tests, unlike other methods of AST, remains relatively unexplored [2]. Further research and development in this domain are crucial for enhancing and complementing AST methodologies and effectively combating antibiotic resistance.

By integrating ML and other automation approaches with disk diffusion testing methodologies, we aim to enhance the accuracy, efficiency, and accessibility of this critical part of AST. The current reliance on highly qualified personnel poses significant challenges, as expertise is often scarce and evaluation results can be inconsistent. We seek to standardize the evaluation process across various laboratory equipment, from data captured with expensive professional-grade setups and equipment to phone-captured data, which is significantly less expensive as mobile phones are ubiquitous nowadays. This democratizes the process and enables more laboratories to automate while our methodology ensures reliability and consistency amongst such diverse conditions.

In earlier research, such as in [3], a semi-automated tool for analyzing images from disk diffusion tests was introduced. While demonstrating an agreement of 87 % with expert microbiologists, it focuses solely on using expensive professional-grade laboratory equipment. Similarly, in the studies conducted by [4] and [5], fully automated disk diffusion method evaluation systems were proposed. Despite their success, these methods lack adaptability to resource-limited settings. In 2021, an AI-based mobile application tailored for AST analysis was published, targeting resource-limited environments [6]. While achieving an overall accuracy of 90 %, this application operates exclusively on mobile platforms.

Conversely, we believe that restricting the solution to a single platform limits its overall accessibility and usability. Therefore, we aim to develop our own using a multi-platform approach, allowing for better accessibility of automated AST analysis across different acquisition devices and settings and enabling more laboratories to implement it.

Our methodology relies on an extensive dataset provided by Bruker Daltonics GmbH & Co. KG, comprising over 3200 images captured with different devices, including mobile phones and professional solutions such as the MBT Pathfinder®, colony-picking robot, developed by Bruker Daltonics GmbH & Co. KG. The dataset offers images capturing the disk diffusion approach to AST, specifically images of Petri dishes with agar medium, antibiotic sample disks, and cultured microbes, all ready for evaluation.

In the first experiment of this study, we utilize state-of-the-art object detection models such as EfficientDet and Mask R-CNN to identify antibiotic disks within microbial cultures. In the second experiment, we explore the assessment of inhibition zones of antibiotic samples using an image processing approach. Both experiments provide a foundation for integrating our methodology into a cross-platform web application that will adopt a device-agnostic approach.

II. MATERIALS AND METHODS

A. *Object Detection Experiment*

Our approach primarily adopts an ML framework for its capacity to generalize effectively across diverse scenarios. That is crucial, especially since our goal is to utilize both mobile and professional platforms. Devices like mobile phones frequently generate lower-quality images, particularly due to overall poorer acquisition conditions. As mentioned earlier, the foundation of our methodology lies in the dataset provided by Bruker Daltonics GmbH & Co. KG, which includes more than 3200 high-resolution disk diffusion test images. This dataset consists of images with varying proportions taken from different capturing angles under various lighting and other acquisition conditions. Pairing the diversity of our dataset with the strength of ML empowers us to achieve high-precision detection across these greatly varying acquisition conditions met in real-world laboratory and clinical settings.

For the purpose of training the object detection models and further evaluation, each of the images in the dataset was manually annotated. This process involves labeling the positions of antibiotic sample disks by bounding boxes and then translating them to coordinates. Another step is measuring the radii of inhibition zones generated by the antibiotic samples in pixels. The output of this process can be seen in the Fig. 1. Although we avoid manual solutions by facilitating ML, this labor-intensive process remains essential. It significantly pays off as it contributes to the superior generalization capability of ML solutions compared to hard-coded approaches.

Fig. 1 further illustrates the conditions and challenges our approach must address by presenting ideally inferred example images. The top row showcases images captured by the aforementioned MBT Pathfinder® professional solution, while the bottom row features images captured with mobile phones. This section focuses on detecting sample disks highlighted by green bounding boxes. Antibiotic disks typically appear in oval shapes, as depicted in the figure. However, their overall appearance and color vary significantly across different acquisition settings, particularly in mobile-captured images, where the capturing conditions can often be inferior.

To achieve accurate detection, which is the foundation for our further experiments, we utilize TensorFlow framework [7] and pre-trained models from the TensorFlow Model Detection Zoo, such as EfficientDet D0 and D2 from the EfficientDet family [8], alongside Mask R-CNN with Inception Resnet v2 backbone [9] (referred to as just Mask R-CNN). Working with pre-trained models offers clear advantages over developing them from scratch. Trained on diverse datasets, they capture key features for object detection, improving performance, robustness, and resource efficiency.

We utilize transfer learning to adapt these pre-trained models, which is computationally intensive, especially for high-precision large image models. Training may span several days, necessitating access to powerful computing resources. To accelerate this, we leverage the computational capabilities of Bruker Daltonics GmbH & Co. KG server infrastructure equipped with NVIDIA RTX 4090 graphic cards.

B. *Antibiotic Effectivity Assessment Experiment*

Another part of our methodology incorporates an algorithm dedicated to assessing the effectiveness of antibiotics on microbial cultures. Its primary objective is determining the radii of circular inhibition zones surrounding antibiotic sample disks. It integrates multiple image processing techniques, including methods from the OpenCV library [10]. The desired outcomes of this experiment are further illustrated in Fig. 1, where inhibition zones, demarcated by a ground truth circle, are highlighted in cyan alongside their corresponding ground truth radii measured in pixels.

The effectiveness assessment process involves several image preprocessing steps. We employ a denoising technique and a Gaussian filter to refine the images for further processing.

As hinted in section II-A, this stage heavily depends on precise antibiotic sample disk detection, as our algorithm chooses several seed positions surrounding the closest proximity of each antibiotic disk. The average pixel brightness differences for each color channel of the given seeds are calculated. Moreover, a coefficient, determined based on the overall brightness of the image with less impact on the images with the lowest and highest values, is applied to the brightness differences. Subsequently, every seed and its adjusted brightness differences are sequentially fed into separate OpenCV flood-fill algorithms. The output of this process resembles a joint flood-filled area converted to a binary mask.

After employing the abovementioned methods, supplementary algorithms are essential to interpret flood-filled areas and convert them into specific radial measurements. To address this challenge, we utilize a morphological ellipsoid filter on the final joint flood-filled mask. This ensures that small holes or gaps are closed and the mask is smoothed before being put into our custom adaptive circle-enlarging algorithm.

The adaptive circle-enlarging algorithm scans the binary mask, gradually expanding itself and the inhibition zone circle until reaching a brightness deviation threshold. Suppose this threshold is not met within a set number of iterations. In that case, it suggests the image is fully flood-filled, implying the selected antibiotic sample likely did not generate any inhibition zone, which is vital for preventing false results.

III. RESULTS

The efficiency of our trained object detection models was assessed using the Mean Average Precision (mAP) metric defined in [11], with a script rewritten in Python specifically developed for [12]. Along with mAP, we also measured the average inference time of each model. (Average time to process one image by the model.) The mAP metric offers a comprehensive evaluation, considering average precision across various Intersections over Union (IoU) thresholds, ranging from 0.5 to 0.95, with gradual increments of 0.05.

Table I evaluates trained object detection models using the abovementioned metrics. Models were tested on our dataset's

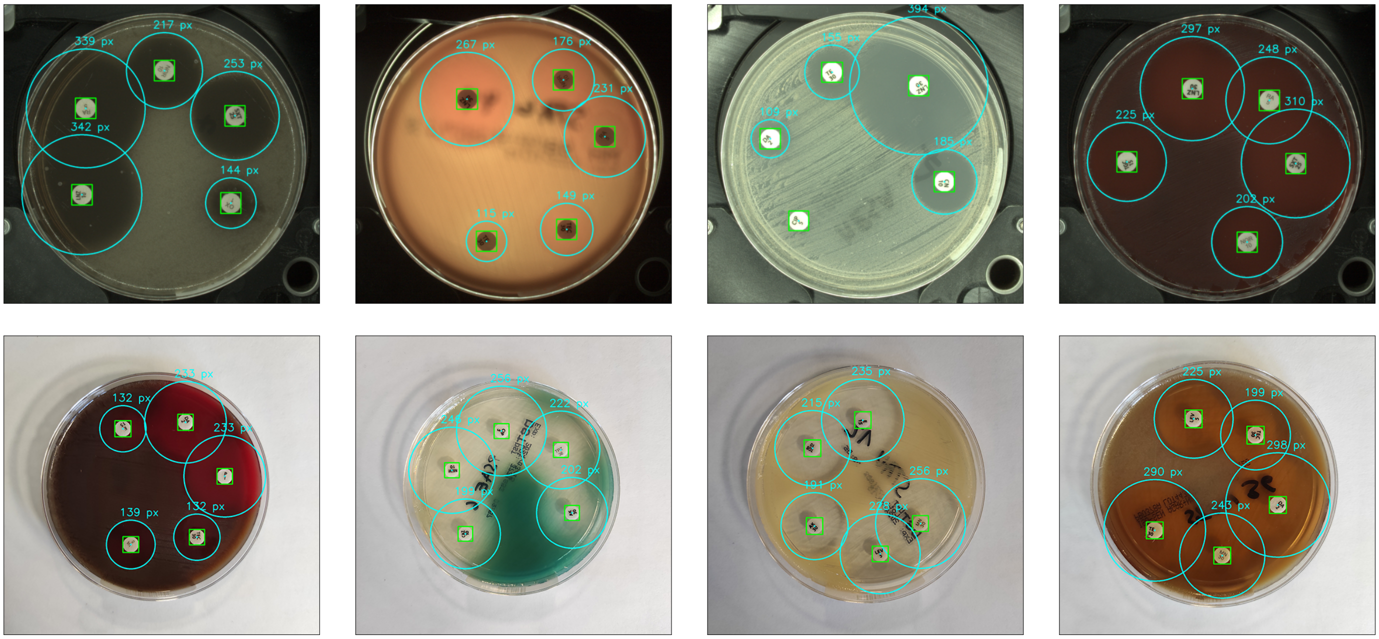

Fig. 1. Examples of ideally processed images: top images taken with professional lab equipment (MBT Pathfinder®), bottom images with mobile phones.

test subset, consisting of 10 % of the total number of images. We assessed each model's performance based on its mAP score and the inference time using a laptop equipped with an AMD Ryzen 5800H CPU.

From Table I, it is apparent that the Mask R-CNN model achieved the highest mAP score of 94.37 %, indicating its superior performance in accurately detecting various antibiotic disks. However, this improved accuracy comes at the cost of increased inference time, with Mask R-CNN requiring 3896 ms per inference on the specified hardware.

TABLE I
COMPARISON OF ANTIBIOTIC SAMPLE DETECTION PERFORMANCE USING DIFFERENT TRAINED MODEL ARCHITECTURES

Architecture	**Inference time [ms]**	**mAP [%]**
EfficientDet D0	237	87.43
EfficientDet D2	454	88.62
Mask R-CNN	3896	94.37

Conversely, the EfficientDet models, while exhibiting slightly lower mAP scores compared to Mask R-CNN, showcased significantly faster inference times. EfficientDet D0 and D2 achieved mAP scores of 87.43 %, and 88.62 % respectively, with inference times of 236 ms and 453 ms.

Our dataset images are large, so they must be downsampled before inferencing. We would suspect that models with greater input sizes would yield more precise results. However, in our case, the EfficientDet D2, despite having a greater input size of 768 × 768 pixels, gains minimal improvements to D0, which operates on 512 × 512 pixel images.

The radii measuring algorithm was designed to operate on images captured by professional-grade equipment and non-professional solutions such as mobile phones. Despite this, the variability in data quality between images captured by professional-grade equipment and those obtained from mobile phones necessitated a nuanced approach to evaluation. In this part, we present the results based on two distinct datasets: images captured by professional-grade equipment, specifically MBT Pathfinder®, and images obtained from mobile phones. Each dataset contains 60 varying images, with more than 250 antibiotic samples in each of them. As the evaluation of our radii measuring algorithm is completely separate from the evaluation of the object detection experiment, we utilize the ground truth locations of the disks instead.

Given the inherent differences in image capture conditions and data quality between professional-grade equipment and mobile phones, we recognize the need to evaluate algorithm performance separately for each dataset. This allows for accurately assessing the algorithm's capabilities under diverse capture conditions and ensures the results' reliability.

Table II summarizes the comparison of radius detection performance across the two types of datasets. Given the significant variation in the radii of inhibition zones in pixels between datasets captured on mobile devices and MBT Pathfinder®, with average values of 399.78 pixels and 270.74 pixels, respectively, the metrics used are relative to the average ground truth radii sizes in pixels. Specifically, we utilize Relative Mean Absolute Error (RMAE), Relative Root Mean Square Error (RRMSE), and Relative Standard Deviation (RSTD).

As we delve into the results, notable differences in algorithm performance between the two datasets are observed. While inferring the one captured by the professional solution, the algorithm demonstrates fairly decent performance, as evidenced by lower error metrics: RRMAE (12.05 vs. 18.49 %), RRMSE (16.45 vs. 26.29 %), and RSTD (14.13 vs. 25.53 %),

TABLE II
COMPARISON OF RADIUS DETECTION PERFORMANCE ACROSS DATASETS ACQUIRED IN VARYING ACQUISITION CONDITIONS

	Acquisition device	
Metric	**MBT Pathfinder®**	**Mobile**
RMAE [%]	12.05	18.49
RRMSE [%]	16.45	26.29
RSTD [%]	14.13	25.53

indicating closer alignment between predicted and ground truth inhibition zone sizes. Conversely, the algorithm performs worse on the dataset acquired from mobile phones. Higher error metrics indicate less accurate detection of inhibition zone sizes than on the professional-grade equipment dataset. For example, the RSTD for inferring on the mobile dataset is more than 80 % higher than on the professional-captured dataset (25.53 vs. 14.13 %), indicating a greater degree of inconsistency in the algorithm's performance between each of the predictions. This further highlights the algorithm's struggle to generalize across datasets with more varying capture conditions.

IV. CONCLUSION

In this study, we endeavored to automate the evaluation of antibiotic effectiveness in the disk diffusion method across a wide range of laboratory equipment. Human evaluation of antibiotic effectiveness requires specialized training and expertise, which may not always be readily available. The standardized approach we propose enhances reliability while enabling laboratory personnel to focus on more critical aspects of their work. With the help of ML and its great generalization capabilities, we capitalize on the widespread availability and ubiquity of image-capturing devices like mobile phones. Utilizing both professional and mobile-capturing devices, we democratize the evaluation process and make it accessible to a broader range of laboratories.

The Object Detection Experiment involved training machine-learning object detection models to detect antibiotic disks. After reviewing the results, the models trained on our dataset exhibit robust performance in detection accuracy. During this process, we also demonstrated that different models vary in computation efficiency and accuracy, emphasizing the importance of selecting the most suitable model for future work. The results of this experiment underscore the effectiveness of the ML approach as the ability of the models to generalize on data of different acquisition conditions and devices is high. While the experiments provided promising results, there is room for improvement. Exploring additional state-of-the-art object detection models could lead to better balances between efficiency and accuracy.

Conversely, while yielding satisfactory results, the Antibiotic Effectivity Assessment Experiment revealed challenges in generalization across mobile phone data. This disparity underscores the limitations of hand-coded algorithms in accommodating varied acquisition conditions. In future experiments, we may explore additional machine-learning approaches to address these challenges and enhance the adaptability and robustness of our methodology. Additionally, the varying conditions under which the photos were taken highlight the importance of standardization in the image acquisition process. Especially for non-professional solutions, improving lighting and stabilization conditions could dramatically enhance the quality and consistency of images.

The next step in our methodology involves integrating the experiments into a valuable application for use in laboratories that could run on a broad range of architectures. This endeavor will provide more insights into the effectiveness of object detection algorithms versus their accuracy and how much resources can be allocated accordingly.

ACKNOWLEDGMENT

This article was made in cooperation with Bruker Daltonics GmbH & Co. KG. The company generously provided a disk diffusion dataset and other valuable resources, including access to their server. We sincerely appreciate their support and collaboration in this research endeavor.

REFERENCES

[1] D. Song, Y. Lei, Mini-review: Recent advances in imaging-based rapid antibiotic susceptibility testing, Sensors and Actuators Reports 3 (2021) 100053. doi:10.1016/j.snr.2021.100053.

[2] O. Dauwalder, F. Vandenesch, Disc diffusion AST automation: one of the last pieces missing for full microbiology laboratory automation, Clinical Microbiology and Infection 26 (5) (2020) 539–541. doi:10.1016/j.cmi.2020.01.021.

[3] C. Alonso, C. Domínguez, J. Heras, E. Mata, V. Pascual, C. Torres, M. Zarazaga, Antibiogramj: A tool for analysing images from disk diffusion tests, Computer Methods and Programs in Biomedicine 143 (2017) 159–169. doi:10.1016/j.cmpb.2017.03.010.

[4] M. Hombach, M. Jetter, N. Blöchliger, N. Kolesnik-Goldmann, E. C. Böttger, Fully automated disc diffusion for rapid antibiotic susceptibility test results: a proof-of-principle study, Journal of Antimicrobial Chemotherapy 72 (6) (2017) 1659–1668. doi:10.1093/jac/dkx026.

[5] A. Cherkaoui, G. Renzi, N. Vuilleumier, J. Schrenzel, Performance of Fully Automated Antimicrobial Disk Diffusion Susceptibility Testing Using Copan WASP Colibri Coupled to the Radian In-Line Carousel and Expert System, Journal of Clinical Microbiology 59 (9) (Aug. 2021). doi:10.1128/jcm.00777-21.

[6] M. Pascucci, G. Royer, J. Adamek, M. A. Asmar, D. Aristizabal, L. Blanche, A. Bezzarga, G. Boniface-Chang, A. Brunner, C. Curel, G. Dulac-Arnold, R. M. Fakhri, N. Malou, C. Nordon, V. Runge, F. Samson, E. Sebastian, D. Soukieh, J.-P. Vert, C. Ambroise, M.-A. Madoui, AI-based mobile application to fight antibiotic resistance, Nature Communications 12 (1) (Feb. 2021). doi:10.1038/s41467-021-21187-3.

[7] T. Developers, TensorFlow (2024). doi:10.5281/ZENODO.10798587.

[8] M. Tan, R. Pang, Q. V. Le, EfficientDet: Scalable and Efficient Object Detection (2019). doi:10.48550/ARXIV.1911.09070.

[9] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN (2017). doi:10.48550/ARXIV.1703.06870.

[10] G. Bradski, The OpenCV Library, Dr. Dobb's Journal of Software Tools (2000).

[11] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, PASCAL VOC2012 Challenge Results, http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html.

[12] J. Cartucho, R. Ventura, M. Veloso, Robust Object Recognition Through Symbiotic Deep Learning In Mobile Robots, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 2336–2341.