

AGAR a Microbial Colony Dataset for Deep Learning Detection

Sylwia Majchrowska

NeuroSYS

Jaroslaw Pawlowski

NeuroSYS

Grzegorz Gula

Department of Pathogen Biology and Immunology, University of Wroclaw, Poland

Tomasz Bonus

NeuroSYS

Agata Hanas

NeuroSYS

Adam Loch

NeuroSYS

Agnieszka Pawlak

NeuroSYS

Justyna Roszkowiak

Department of Pathogen Biology and Immunology, University of Wroclaw, Poland

Tomasz Golan

NeuroSYS

Zuzanna Drulis-Kawa (✉ zuzanna.drulis-kawa@uwr.edu.pl)

Department of Pathogen Biology and Immunology, University of Wroclaw, Poland

Research Article

Keywords: AGAR, Microbial Colony Dataset, Deep Learning Detection

Posted Date: July 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-668667/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

AGAR a microbial colony dataset for deep learning detection

Sylwia Majchrowska^{1,2}, Jarosław Pawłowski^{1,2}, Grzegorz Guła^{1,3}, Tomasz Bonus¹, Agata Hanas¹, Adam Loch¹, Agnieszka Pawlak¹, Justyna Roszkowiak³, Tomasz Golan^{1,*}, and Zuzanna Drulis-Kawa^{1,3,+}

¹NeuroSYS, Rybacka 7, 53-656 Wrocław, Poland

²Faculty of Fundamental Problems of Technology, Wrocław University of Science and Technology, Wybrzeże S Wyspiańskiego 27, 50-372 Wrocław, Poland

³Department of Pathogen Biology and Immunology, Institute of Genetics and Microbiology, University of Wrocław, S. Przybyszewskiego 63, 51-148 Wrocław, Poland

*t.golan@neurosys.com

+zuzanna.drulis-kawa@uwr.edu.pl

ABSTRACT

The Annotated Germs for Automated Recognition (AGAR) dataset is an image database of microbial colonies cultured on agar plates. It contains 18 000 photos of five different microorganisms as single or mixed cultures, taken under diverse lighting conditions with two different cameras. All the images are classified into *countable*, *uncountable*, and *empty*, with the *countable* class labeled by microbiologists with colony location and species identification (336 442 colonies in total). This study describes the dataset itself and the process of its development. In the second part, the performance of selected deep neural network architectures for object detection, namely Faster R-CNN and Cascade R-CNN, was evaluated on the AGAR dataset. The results confirmed the great potential of deep learning methods to automate the process of microbe localization and classification based on Petri dish photos. Moreover, AGAR is the first publicly available dataset of this kind and size and will facilitate the future development of machine learning models. The data used in these studies can be found at <https://agar.neurosys.com/>.

1 Introduction

In recent years, data-driven artificial intelligence (AI) methods have dominated automated pattern search. In particular, deep learning (DL) approaches successfully reduce the need for feature engineering by leveraging a large amount of data to optimize the model's parameters and find the most important data features¹. DL is more and more broadly used in many scientific disciplines. Deep neural networks trained on huge datasets are irreplaceable in modern imaging diagnostics in medicine and allow for fast and precise clinical decisions and medical procedure implementations²⁻⁶. Another example is the application of DL in cancer research and diagnostics⁷⁻¹⁰, where it improves automation and opens new avenues for artificial intelligence-assisted precision oncology and medicine in general. Overall, the implementation of AI provides a new quality of medical tools and diagnostics enabling faster and more efficient healthcare procedures.

Another area where AI could be implemented is microbiology with regard to standard procedural requirements in the pharmaceutical, cosmetic, or food industries. The manufacturing processes are subjected to strict policies and regulations listed, for instance, in Pharmacopoeia¹¹⁻¹³ or European Medicines Agency (EMA) guidelines¹⁴, including extensive regulations governing microbiological purity in industrial areas^{15,16}. The above requirements obligate manufacturers to perform constant microbiological monitoring, which means thousands of samples analysed by experienced microbiologists. In most production plants, microbiological procedures

are entirely manual - starting from sample collection and culturing to the final plate evaluation with microbial colony count and identification. Large companies start implementing laboratory automation systems to speed up the operating process^{17,18}. Taking into account the current industry demand, we aim to develop an effective automatic analysis of microbiological samples based on DL algorithms, performing an efficient and precise microbial sample analysis.

The existing solutions for microbial colony analysis provide only some level of automation. Commonly, the process is limited to generating the photo of a microbial culture grown on a Petri dish, which is further investigated and manually curated by microbiologists. However, the full automation of microbial sample analysis might be accomplished by the implementation of deep neural networks providing robust and accurate algorithms for microbial colony localization, classification, and counting. DL-based solutions can generalize different acquisition setups and new microbe species easier than traditional computer vision algorithms, which strongly depend on hand-crafted feature selection¹⁹. As DL requires well-balanced and diverse data to optimally fit the model's parameters, there is a need to create huge publicly available datasets for neural network training and, consequently, achieving their best performance.

The main objective of our work was to develop a DL-based methodology to identify and count microbial colonies based on the photo of a standard agar plate culture. At first, we prepared the Annotated Germs for Automated Recogni-

tion (AGAR) dataset consisting of over 330k labeled microbial colonies (*Staphylococcus aureus*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Escherichia coli*, and *Candida albicans*) localized on 18k annotated Petri dish photos. Secondly, we evaluated the selected deep neural networks, such as Faster R-CNN²⁰ and Cascade R-CNN²¹ architectures for object detection, with several different backbones – ResNet²², ResNeXt²³, and HRNet²⁴ – to set benchmarks for the task. The whole process of collecting and annotating the data as well as the deep learning analysis is presented in Fig. 1.

AGAR dataset preparation

Designing a dataset requires several decisions to be made. In the case of data imaging for microbial colony counting and identification, one needs to select microorganism species to study, provide proper growth conditions (applied medium, incubation temperature and time, culture dilution factors, etc.), and choose an appropriate photo acquisition setup (light source location, external light handling, or camera type). Such decisions for the AGAR dataset were based on the goal of making a diverse dataset for automatic colony counting on agar plates which could be easily reproduced in any microbiological laboratory.

The selection of standard microorganisms for the AGAR dataset preparation was based on Pharmacopoeia^{11–13} regulations and ATCC guidelines²⁵. Five representatives from different bacterial groups (*S. aureus* subsp. *aureus* ATCC 6538, *B. subtilis* subsp. *spizizenii* ATCC 6633, *P. aeruginosa* ATCC 9027, *E. coli* ATCC 8739), and *C. albicans* ATCC 10231 as a yeast strain were chosen. A series of 10-fold dilutions of refreshed cultures were made and 100 ul were inoculated onto Trypticase Soy Agar (TSA) plates in five technical replicates. The microbial cultures were then incubated at 37°C for 18-24 hours. For data acquisition, single and mixed cultures (*S. aureus* & *P. aeruginosa*/*E. coli*/*C. albicans*; *P. aeruginosa* & *E. coli*/*C. albicans*; *E. coli* & *C. albicans*) were used. The images of colonies grown on agar plates were taken using one of two cameras. Thus, the whole photo dataset was divided into two major subsets: *higher-resolution* (4000 x 6000 px) and *lower-resolution* (2048 x 2048 px). Moreover, three different subgroups were distinguished in the former subset: *bright*, *dark*, and *vague* based on different illumination conditions in which the photos were taken. For visualization of plates captured in the various setups see Fig. 3, whereas details on microbial colony growth conditions and acquisition setups are provided in Supplementary Sections 1.1 and 1.2.

A web application was developed for microbiologists to upload and annotate agar plate culture photos. Each sample recorded in the database was marked as *countable*, *uncountable*, or *empty* (Fig. 2(d)). It was important to establish a countable range for manual counting, which is generally between 30 - 300 colonies for a standard 100 mm Petri dish^{26,27}. Every *countable* sample was manually labeled by microbiologists using a bounding box and species label per colony. Samples were classified as *uncountable* if more than 300

colonies were present, or if no discrete colonies were seen due to their agglomeration or blurred shape.

In total, 18 000 photos taken under diverse lighting conditions were collected in the AGAR database. Altogether, 336 442 colonies of 5 standard microbes were annotated. The summary of the AGAR dataset statistics is presented in Fig. 2(b). The AGAR dataset achieves a good balance in the number of instances for different microbes (see Fig. 2(a)), which is important for building a robust deep learning model. The interquartile range for all 12 271 countable samples is between 4 and 38 colonies, and around 84.4% of these samples have less than 50 colonies. The histograms of the number of annotated colonies per *countable* image are presented in Fig. 2(c). Additional information of provided annotations and bounding box size distribution among microbial classes can be found in the Supplementary Sections 1.2 and 1.3.

Deep Learning Analysis

To verify the suitability of the AGAR dataset for building deep learning models for image-based microorganism recognition, and to set a benchmark for the task, we evaluated the performance of the two neural network architectures for object detection: Faster R-CNN²⁰ and Cascade R-CNN²¹, with four different backbones: ResNet-50²², ResNet-101²², ResNeXt-101²³, and HRNet²⁴. The implementation from the MMDetection²⁸ toolbox was used. The backbones' weights were initialized using models pre-trained on ImageNet²⁹, available in the *torchvision* package of the PyTorch library. More technical details are provided in Supplementary Section 4.1.

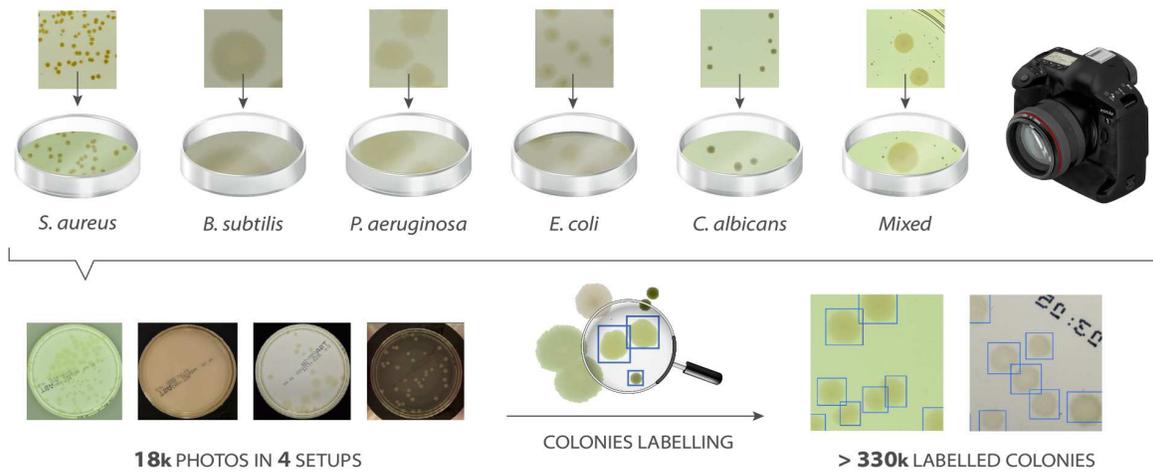
The quantitative results for each subset of the AGAR dataset analyzed separately are reported, as well as the results for models trained on the whole dataset. Each subset was split randomly into approximately 75% samples, which constituted the training set, and 25% for the validation and testing set. Due to the relatively high resolution of all images, the samples were divided into patches of 512x512 px to avoid scaling down, which would make the recognition of small colonies more challenging. If necessary, zero paddings were applied to get a proper aspect ratio. As the goal of this study was to establish a baseline for the microbe detection task, the neural network architecture was not modified, thus the patches were resized to match the default input layer size. More details on the data split and patch preparation can be found in Supplementary Section 2.1.

Several approaches have been undertaken for data augmentation, including the addition of Gaussian blur, salt and pepper noise, changing the image color space to LAB and HSV, preparing histogram equalization, rotating images in the [-45, 45] degree range, or cropping around the annotated bounding boxes, but always keeping a visible microbe object. However, the best results were obtained through splitting the image into patches and normalizing it using means and standard deviation per channel values, the same as for the Common Object in Context (COCO) dataset³⁰.

The evaluation metric for colony detection was based on the

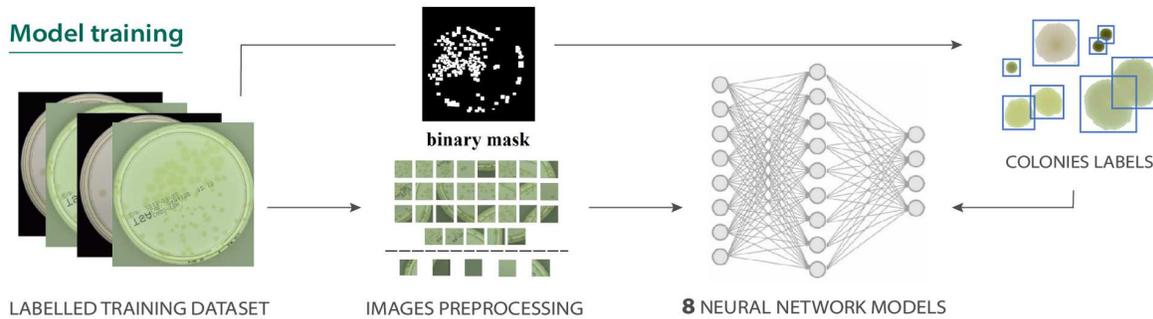
AGAR dataset preparation

5 MICROBIAL CULTURES ON TSA PLATES



Deep learning analysis

Model training



Model evaluating

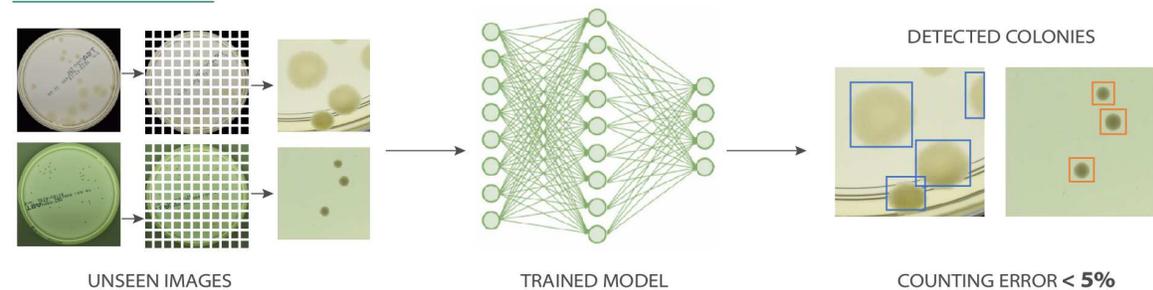


Figure 1. An overview of the processing pipeline from microbiological sample preparation to model performance evaluation. The *AGAR dataset preparation* (top panel) includes 5 microbial species as single and mixed cultures on agar plates from different liquid dilutions, followed by image acquisition with 4 different setups, finalized with colony labeling and identification by a microbiologist. The *deep learning analysis* (bottom panel) using 8 neural network architectures consists of two separate stages. Images randomly cut into patches constituted the network input in the *model training* stage, while overlapping patches of unseen images were used in the *model evaluating* stage. See technical details in Supplementary Section 2.1.

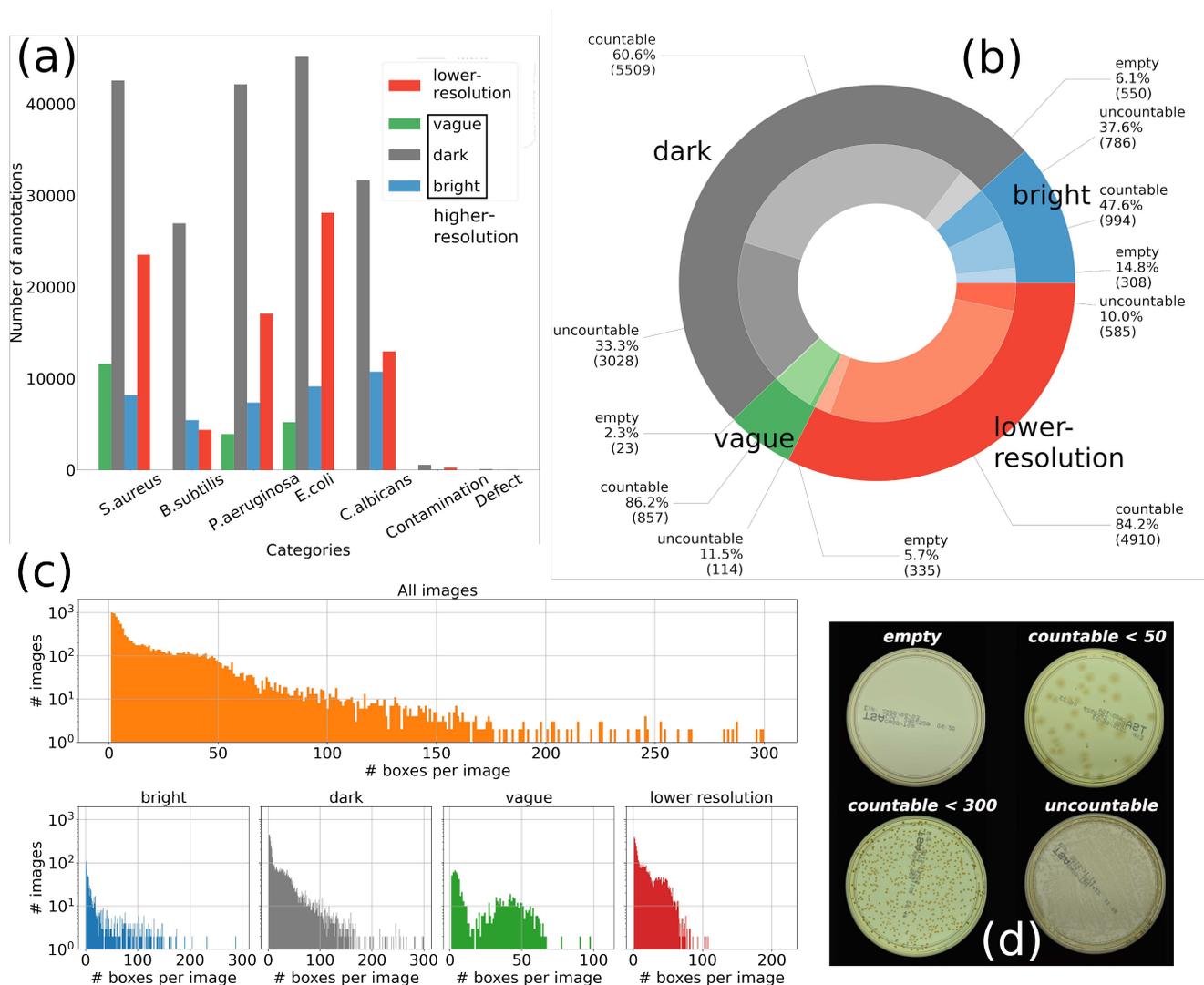


Figure 2. The AGAR dataset statistics summary. The number of samples distribution: (a) over different microbial species, and (b) for different acquisition setup subgroups. The combination of samples from *bright* and *dark* subgroups constitutes about half of the whole dataset and includes photos with the best quality in terms of the contrast between grown microorganisms and agar surface. (c) The histograms of the number of annotated instances per image calculated including *countable* samples for different subsets separately (blue, grey, green, red) and altogether (orange). The mean number of annotations per image equals 27.4. (d) Image examples of agar plates with different numbers of growing colonies, marked as *empty*, *countable*, and *uncountable*.

mean Average Precision (mAP) previously established for the COCO competition³⁰. The Average Precision (AP) is calculated as the average value of precision on the precision-recall (PR) curve for a given Intersection over Union (IoU) threshold. IoU describes the level of overlapping between truth and predicted bounding box. The mean value of AP for different IoU thresholds and/or classes gives the mAP. To measure the effectiveness of the colony counting, two separate metrics were provided, namely Mean Absolute Error (MAE), its cumulated variation (cMAE), and symmetric Mean Absolute Percentage Error (sMAPE), and evaluated for different models on every

data subset. Our post-processing of the predicted bounding boxes is described in detail in the Supplementary Section 2.2, while the calculations of all used metrics (mAP, MAE, cMAE, sMAPE) in the Supplementary Sections 3.1 and 3.2.

The analysis for each of the major subsets of the AGAR dataset (*higher-* and *lower-resolution*) was performed independently. The vague subset was excluded from the studies. It was later used as a different domain representation in the studies on the models' ability to generalize to new acquisition setups, see Supplementary Sections 4.7 and 4.8 for details. The results for the baseline model (Faster R-CNN with

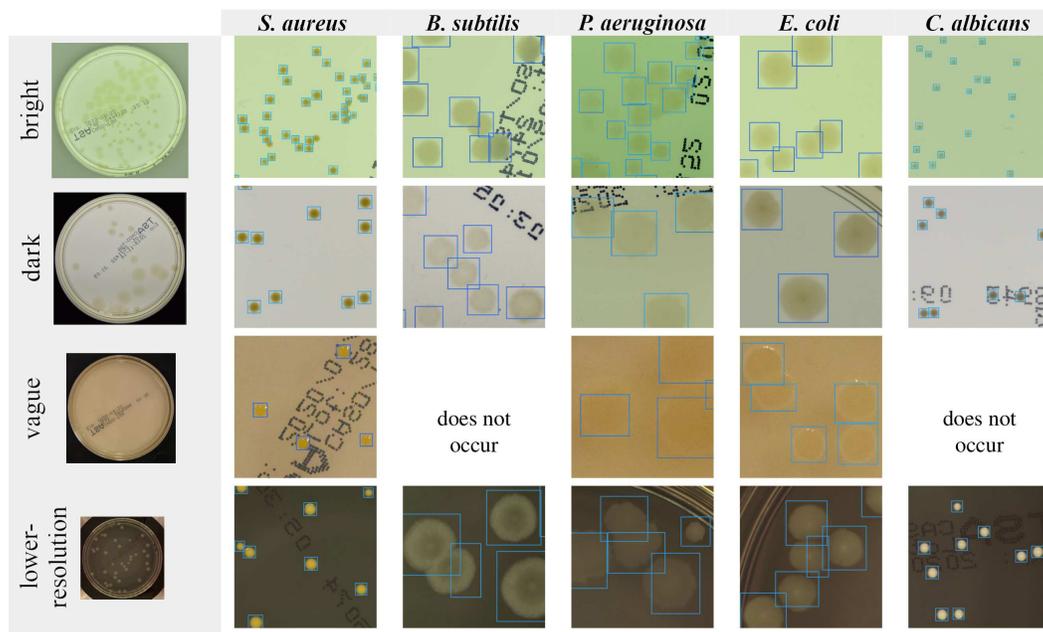


Figure 3. Examples of Cascade R-CNN with HRNet predictions for different types of microorganisms grown on TSA (Petri plates) captured with different image acquisition setups. Entire plates preserving the photoscale are presented in the first column. The other columns represent 20×20 mm fragments of the plate. Well suited bounding boxes were found for organisms forming smaller, dense colonies, like *S. aureus* and *C. albicans*. In the case of *P. aeruginosa* growing in large translucent colonies, some were omitted, especially in the *vague* subgroup, where the edges of colonies could not be clearly defined. The *vague* subgroup does not include *B. subtilis* and *C. albicans* at all.

ResNet-50), along with more complex, multi-stage detectors (Cascade R-CNN with HRNet), are presented in Fig. 4. The visualisation of predictions made by Cascade R-CNN with HRNet backbone is shown in Fig. 3. Detailed descriptions of results achieved by other tested detectors are placed in Supplementary Sections 4.2 and 4.3.

Our main outcomes in terms of mAP are presented in Fig. 4 (a). In general, Cascade R-CNN performs slightly better with mAP scores equal to 52.0% for the *higher-* and 59.4% for the *lower-resolution* subset, compared to Faster R-CNN, which achieved 49.3% and 56.0%, respectively. A comprehensive study showed that the increase of backbone complexity gives just slightly better results (see Supplementary Section 4.2), which suggests that, in this class of architectures, the model’s capability is reached and extending backbone networks further might not lead to much better performance.

The average precision calculated per microbe is higher for microbes forming smaller colonies, namely *S. aureus* and *C. albicans*. The rest of the studied microorganisms tend to aggregate or overlap, especially in low dilution samples (more colonies on the plate surface). Moreover, larger colonies tend to have blurred edges with lower contrast with regard to agar substrate. This all leads to less accurate microbe detection in terms of mAP. The mAP score measures the quality of detection using IoU, which is related to true and predicted bounding boxes overlapping. For the colony counting, the

recognition of microbes is crucial, but the exactness of their location is of secondary importance. Therefore, the sMAPE and MAE metrics were used for tuning the algorithms applied in the post-processing stage to merge predictions for individual patches into the whole test image. The values of two parameters, namely classification probability threshold and NMS threshold for the modified soft non-maximum suppression algorithm³¹, were established on the training dataset and then used to eliminate the excess of bounding boxes (e.g. appearing twice through the edges of neighbouring overlapped patches) for the test data.

The results of microbe counting are presented in Fig. 4(b). Overall, Cascade R-CNN with HRNet predictions are more accurate with sMAPE equal 4.86% for *higher-* and 3.81% for *lower-resolution* subsets. It is worth noting that the most lightweight of the considered models (Faster R-CNN with ResNet-50) is also very effective with 5.32% and 4.68% sMAPE for the same subsets. As these metrics do not include the error coming from microorganism misclassification, cMAE being a sum of MAE calculated for every microbe separately is also provided (in brackets in *altogether* row). For cMAE, only properly recognized microbes are taken as a correct count. The small difference between MAE and cMAE (e.g. 1.57 vs 1.76 in case of Cascade with HRNet for *lower-resolution* subset) proves the great ability of the models to distinguish between different species.

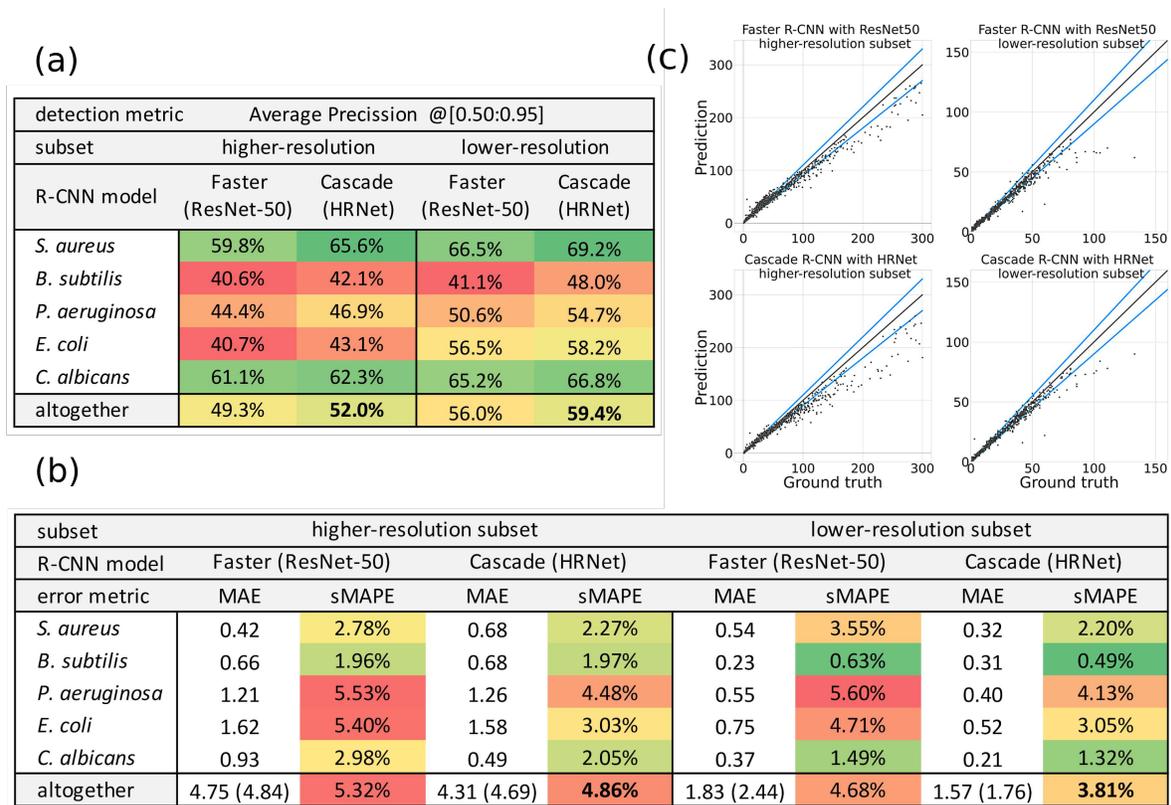


Figure 4. Overall results for colony detecting and counting tasks. (a) Average precision for the two different models obtained on the *higher-* and *lower-resolution* subsets: the Cascade R-CNN model with HRNet backbone gives the highest mAP. (b) MAE and sMAPE metrics for Faster R-CNN and Cascade R-CNN calculated separately for each class of microbes and jointly (including cMAE in brackets). (c) Microbial colony counting performance – the number of predicted colonies versus ground truth (each point represents the result for a single plate).

The selected models' predictions for microbial counting on test AGAR subsets are presented in Fig. 4(c). The straight black lines represent perfect results, while the blue ones indicate a 10% error to highlight the acceptable error range. The detectors tend to underestimate the exact number of colonies for more populated samples. This is largely due to the nature of the sMAPE metric, which weights errors inversely to the number of instances. Moreover, the AGAR dataset is dominated by plates with less than 50 colonies (see Fig. 2(c)), which impacts the values of the probability and NMS thresholds for post-processing. The extension for higher populated samples using double thresholding approach is described in Supplementary Section 4.6.

Although *empty* and *uncountable* plates come with no colony annotations, qualitative studies were carried out to evaluate the performance of the trained models. Experiments on uncountable samples showed that many plates labeled by microbiologists as having more than 300 colonies were, in fact, within the countable range, as recognized by the detectors. As discussed above, the models are more accurate for samples with less than 50 colonies. However, they give very good estimates for plates with hundreds or even thousands

of instances, correctly identifying single colonies in highly populated samples, with the maximum number of predicted colonies on one plate equal to 2782. It is worth noting that it takes seconds for the deep learning system, while it could take up to an hour in the case of manual counting.

The analysis of samples marked by microbiologists as *empty* presented the great capability of deep learning methods for microbiological quality control. The detectors were able to recognize colonies difficult to see and missed by human specialists. Overlooked microbes were usually small and located either at the edge of a Petri dish or on a plate factory marking. This is a strong indication that human-AI interaction can lead to a higher quality of microbiological analyses.

Further discussion on empty and uncountable samples can be found in Supplementary Section 4.5. The complementary analysis of the impact of the models' initialization parameters are provided in Supplementary Section 4.4. The detectors' ability to generalize to different AGAR subsets with transfer learning technique is described in Supplementary Section 4.7. The models' performance on every subset when trained on the whole dataset is provided in Supplementary Section 4.8.

Discussion and Conclusions

There is a growing interest in applying computer vision algorithms to microbiological analyses, especially in industry and pharmaceutical branches, but many existing approaches are based on traditional image processing techniques^{32–36}. Although deep learning methods have recently become more commonly used^{17,37–41}, the proposed procedures are usually not end-to-end solutions and cannot be used independently to process an image of an entire Petri plate. The lack of effective DL-based approaches for microbiological purposes possibly results from the poor availability of huge datasets needed for deep neural network training. To fill this gap, the AGAR dataset is introduced with 18 000 photos of agar plate cultures (Petri dish) and 336 442 annotated colonies of five microorganisms (staphylococci, Gram-positive and negative bacilli, and yeast) most commonly used according to the Pharmacopoeia guidelines.

An earlier study by Ferrari et al.^{17,40,41} presented the creation of the MicrobIA, a smaller dataset composed of about 29 000 labeled segments (fragments of the Petri dish with single agglomerates) of bacterial colonies grown on blood and chromogenic agar plates. The segments were extracted with the WASPlab system from photos of whole Petri plates and manually assigned to one of seven classes (segments containing from 1 to 6 colonies and outliers). The MicrobIA dataset was then used to train a convolutional neural network to classify images into one of the six colony counts or the outliers category (bubbles, dust or dirt on the agar). Authors reported the per-colony error of 28% measured on the individual segments¹⁷. Another study using the MicrobIA dataset⁴² explored the possibility of image generation for agar plates' segmentation and investigated the domain shift problem. It was proved that synthetic data can be applied to train a deep neural network for agar plate image segmentation.

In contrast, our study demonstrates the possibility of exploiting deep learning-based detectors to build end-to-end solutions for multiclass microbial colony recognition and counting. The provided benchmarks of the chosen DL models prove that the AGAR dataset can be used to build robust models, which adjust well to real data achieved in various acquisition setups, resulting in different illumination and resolution of a Petri plate photo. The exhaustive analysis of eight different deep neural networks for object detection was performed on two AGAR subsets referred to as *higher-* and *lower-resolution*. The best performing model used in our study (Cascade R-CNN with HRNet backbone) achieved the low counting error of 4.92% and 3.81%, respectively on both subsets. This is a significant improvement over previous reports describing the microbial colony counting with computer vision algorithms¹⁷. In the case of detection, the mAP scores within the range from 49.3% to 59.4% were achieved for different detectors, which is an excellent result compared to other reports (44.6% for Cascade R-CNN and 36.7% for Faster R-CNN [16]) done with the same architectures on the COCO dataset.

In summary, the selected R-CNN models perform very well

in detecting microbial colonies. This is likely because the detected instances have similar shapes and all species of microbes are well represented in the training data. Moreover, the results obtained with base Faster R-CNN and more complex Cascade R-CNN do not differ much. For that reason, exploring different state of the art AI approaches, such as Transformers⁴³, seems to be the right future direction for the implementation of deep learning in microbiology.

There is a visibly increasing demand for artificial intelligence in numerous human activities, including social interactions, industry, agriculture, and medicine. The AGAR dataset, which compiles a huge variety of Petri dish culture photos, is a great data source for further exploration. The use of generative models on AGAR might, for example, improve the performance of colony detection on a wider variety of images from other domains, as well as with a larger number of microbial species. Additionally, the large variety in the distribution of colony sizes shows the potential of the AGAR dataset for detailed studies of colony growth dynamics, which can, for example, help us better understand antibiotic resistance of bacteria.

References

1. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
2. Georgios A. Kaissis, Marcus R. Makowski, Daniel Rückert, and Rickmer F. Braren. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 8:305–311, 2020.
3. Ge Wang, Jong Chul Ye, and Bruno De Man. Deep learning for tomographic image reconstruction. *Nature Machine Intelligence*, 8:737–748, 2020.
4. Edward Korot, Zeyu Guan, Daniel Ferraz, Siegfried K. Wagner, Gongyu Zhang, Xiaoxuan Liu, Livia Faes, Nikolas Pontikos, Samuel G. Finlayson, Hagar Khalid, Gabriella Moraes, Konstantinos Balaskas, Alastair K. Denniston, and Pearse A. Keane. Code-free deep learning for multi-modality medical image classification. *Nature Machine Intelligence*, 3:288–298, 2021.
5. Brendon Lutnick, Brandon Ginley, Darshana Govind, Sean D. McGarry, Peter S. LaViolette, Rabi Yacoub, Sanjay Jain, John E. Tomaszewski, Kuang-Yu Jen, and Pinaki Sarder. An integrated iterative annotation technique for easing neural network training in medical image analysis. *Nature Machine Intelligence*, 1:112–119, 2019.
6. Maya Varma, Mandy Lu, Rachel Gardner, Jared Dunmon, Nishith Khandwala, Pranav Rajpurkar, Christopher Beaulieu Jin Long, Katie Shpanskaya, Li Fei-Fei, Matthew P. Lungren, and Bhavik N. Patel. Automated abnormality detection in lower extremity radiographs using deep learning. *Nature Machine Intelligence*, 1:578–583, 2019.

7. Kevin Faust, Sudarshan Bala, Randy van Ommeren, Alessia Portante, Raniah Al Qawahmed, Ugljesa Djuric, and Phedias Diamandis. Intelligent feature engineering and ontological mapping of brain tumour histomorphologies by deep learning. Nature Machine Intelligence, 1:316–321, 2019.
8. Leonie L. Zeune, Yoeri E. Boink, Afroditi Nanou Guus van Dalum, Sanne de Wit, Kiki C. Andree, Joost F. Swennenhuis, Leon W.M.M. Terstappen Stephan A. van Gils, and Christoph Brune. Deep learning of circulating tumour cells. Nature Machine Intelligence, 2:124–133, 2020.
9. Roman Schulte-Sasse, Stefan Budach, Denes Hnisz, and Annalisa Marsico. Integration of multiomics data with graph convolutional networks to identify new cancer genes and their associated molecular mechanisms. Nature Machine Intelligence, pages 124–133, 2021.
10. Alexander Binder, Michael Bockmayr, Miriam Hägele, Stephan Wienert, Daniel Heim, Katharina Hellweg, Masaru Ishii, Albrecht Stenzinger, Carsten Denkert Andreas Hocke, Klaus-Robert Müller, and Frederick Klauschen. Morphological and molecular breast cancer profiling through explainable machine learning. Nature Machine Intelligence, 3:355–366, 2021.
11. European Pharmacopoeia, chapter 2.6.1 Sterility, page 155–158. Council of Europe, 6 edition, 2007. Ph. Eur. 6.0, 01/2008:20601.
12. <62> microbiological examination of non-sterile products: Tests for specified microorganisms. In United States Pharmacopoeia and National Formulary (USP 31-NF 26), volume 31. United States Pharmacopoeia, Washington, USA, 2007.
13. The Pharmacopoeia of Japan. Society of Japanese Pharmacopoeia, Tokyo, Japan, 17 edition, 2016.
14. ICH Topic Q4B Annex 4B Microbiological Examination of Non-Sterile Products: Tests for Specified Microorganisms General Chapter, European Medicines Agency, 2017. EMEA/CHMP/ICH/308817/2008.
15. Microbiological Examination of Non-Sterile Products: Tests for Specified Microorganisms, World Health Organization. https://www.who.int/medicines/publications/pharmacopoeia/SpecifiedMicroorgsTests_QAS11-408_FINALMarch2012.pdf?ua=1, 2012. QAS/11.408FINAL.
16. Q4B Evaluation and Recommendation of Pharmacopoeial Texts for Use in the ICH Regions, Annex 4B(R1) Microbiological Examination of Nonsterile Products: Tests for Specified Microorganisms General Chapter, Guidance for Industry, U.S. Food and Drug Administration [FDA]. <https://www.fda.gov/media/71267/download>, 2017.
17. Alessandro Ferrari, Stefano Lombardi, and Alberto Signoroni. Bacterial colony counting with convolutional neural networks in digital microbiology imaging. Pattern Recognition, 61:629–640, 2017.
18. Matthew L. Faron, Blake W. Buchan, Christopher Coon, Theo Liebrechts, Anita van Bree, Arjan R. Jansz, Genevieve Soucy, John Korver, and Nathan A. Ledebor. Automatic digital analysis of chromogenic media for vancomycin-resistant-enterococcus screens using copan wasplab. Journal of Clinical Microbiology, 54(10):2464–2469, 2016.
19. Niall O’Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. Deep learning vs. traditional computer vision. In Kohei Arai and Supriya Kapoor, editors, Advances in Computer Vision, pages 128–144, Cham, 2020. Springer International Publishing.
20. Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, NIPS’15, page 91–99, Cambridge, MA, USA, 2015. MIT Press.
21. Z. Cai and N. Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6154–6162, 2018.
22. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016.
23. S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5987–5995, 2017.
24. Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, and Yadong Mu. Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, April 2020.
25. Strains specified by official microbial assays, American Type Culture Collection [ATCC]. <https://www.atcc.org/~media/PDFs/Marketing%20Material/Microbiology/Pharmaceutical%20Microbiology.ashx>, 2017.
26. Diane M Tomaszewicz, Donald K Hotchkiss, Ralston B Read Jr George W Reinbold, and Paul A Hartman. The most suitable number of colonies on plates for counting. Journal of Food Protection, 43(4), 1980.
27. Scott Sutton. Accuracy of plate counts. Journal of validation technology, 17(3):42–46, 2011.

28. Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.
29. J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.
30. Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing.
31. Navaneeth Bodla, Bharat Singh, Rama Chellappa, and Larry S. Davis. Improving object detection with one line of code. *CoRR*, abs/1704.04503, 2017.
32. C. Zhang, W. Chen, W. Liu, and C. Chen. An automated bacterial colony counting system. In *2008 IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing (SUTC 2008)*, pages 233–240, 2008.
33. Silvio D. Brugger, Christian Baumberger, Marcel Jost, Werner Jenni, Urs Brugger, , and Kathrin Mühlemann. Automated counting of bacterial colony forming units on agar plates. *PloS One*, 7(3), 2012.
34. Guozhen Zhu, Bin Yan, Mengting Xing, and Chunna Tian. Automated counting of bacterial colonies on agar plates based on images captured at near-infrared light. *Journal of Microbiological Methods*, 153:66 – 73, 2018.
35. Arif ul Maula Khan, Angelo Torelli, Ivo Wolf, and Norbert Gretz. Autocellseg: robust automatic colony forming unit (cfu)/cell analysis using adaptive image segmentation and easy-to-use post-editing techniques. *Scientific Reports*, 8(7302), 2018.
36. Mattia Siragusa, Stefano Dall’Olio, Pil M. Fredericia, Mikael Jensen, and Torsten Groesser. Cell colony counter called coconut. *PloS One*, 13(7), 2018.
37. Dong Nie, Elizabeth A. Shank, and Vladimir Jojic. A deep framework for bacterial image segmentation and classification. In *BCB ’15: Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 306–314, 2015.
38. Bartosz Zieliński, Anna Plichta, Krzysztof Misztal, Przemysław Spurek, Monika Brzywczy-Włoch, and Dorota Ochońska. Deep learning approach to bacterial colony classification. *PloS One*, 12(9), 2017.
39. Muhammed Talo. An automated deep learning approach for bacterial image classification. *arXiv preprint arXiv:1912.08765*, 2019.
40. Alessandro Ferrari and Signoroni Lombardi. Bacterial colony counting by convolutional neural networks. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, pages 7458 – 7461, 2015.
41. Mattia Savardi, Alessandro Ferrari, and Alberto Signoroni. Automatic hemolysis identification on aligned dual-lighting images of cultured blood agar plates. *Computer Methods and Programs in Biomedicine*, 156:13–24, 2018.
42. Paolo Andreini, Simone Bonechi, Monica Bianchini, Alessandro Mecocci, and Franco Scarselli. Image generation by gan and style transfer for agar plate image segmentation. *Computer Methods and Programs in Biomedicine*, 184:105268, 2020.
43. Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable DETR: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020.

Acknowledgements

Project “Development of a new method for detection and identifying bacterial colonies using artificial neural networks and machine learning algorithms” is co-financed from European Union funds under the European Regional Development Funds as part of the Smart Growth Operational Program. Project implemented as part of the National Centre for Research and Development: Fast Track (grant no. POIR.01.01.01-00-0040/18).

Author contributions statement

SM, JP, TB, AH and AP contributed to the DL analysis code. GG and JR prepared microbiological samples, captured photos of standard agar plates culture and annotated them. AL and TG prepared an application to annotate images. SM analyzed and interpreted the AGAR datasets statistics. SM and JP planned the experiments and developed DL-based methodology to identify and count microbial colonies. JP and SM performed calculations and analysed the results of the experiments. ZDK led the microbiologist team, while TG led the DL team during the whole project. TG and ZDK supervised the project.

All authors provided critical feedback and helped shape the research, the analysis and the manuscript.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryinformationforAGARmicrobialcolonydatasetfordeeplearningdetection.pdf](#)