Check for updates

A Clinical Bacterial Dataset for Deep Learning in Microbiological Rapid On-Site Evaluation

Xiuli Wang[1,2], Yinghan Shi[1,2], Shasha Guo[3,4], Xuzhong Qu[3,4], Fei Xie[1], Zhimei Duan[1], Ye Hu[1], Han Fu[1], Xin Shi[1,2], Tingwei Quan[3,4], Kaifei Wang[1] ✉ & Lixin Xie[1] ✉

Microbiological Rapid On-Site Evaluation (M-ROSE) is based on smear staining and microscopic observation, providing critical references for the diagnosis and treatment of pulmonary infectious disease. Automatic identification of pathogens is the key to improving the quality and speed of M-ROSE. Recent advancements in deep learning have yielded numerous identification algorithms and datasets. However, most studies focus on artificially cultured bacteria and lack clinical data and algorithms. Therefore, we collected Gram-stained bacteria images from lower respiratory tract specimens of patients with lung infections in Chinese PLA General Hospital obtained by M-ROSE from 2018 to 2022 and desensitized images to produce 1705 images (4,912 × 3,684 pixels). A total of 4,833 cocci and 6,991 bacilli were manually labelled and differentiated into negative and positive. In addition, we applied the detection and segmentation networks for benchmark testing. Data and benchmark algorithms we provided that may benefit the study of automated bacterial identification in clinical specimens.

Background & Summary

Lower respiratory tract infections (LRTIs) as one of the global health threats, have high morbidity and lethality in different age groups[1]. In recent years, Hospital-Acquired Pneumonia (HAP) and Ventilator-Associated Pneumonia (VAP) have a high incidence, especially in intensive care units (ICU) and have become the most common nosocomial infection in China[2]. The proliferation of hospital infection usually results in the prevalence of multidrug-resistant and extensively drug-resistant bacteria, so early and accurate etiological diagnosis is crucial for the diagnosis and treatment of infectious diseases.

Bacteria morphology identification plays an essential role in specimen interpretation for rapid bacterial identification, and it is possible to reduce the cost and yield intuitive results by visualizing microbial features through microscopy techniques. This strategy has led to the emergence and widespread use of M-ROSE[3]. During tracheoscopy, the rapid on-site smear staining technique is performed on the specimens obtained (sputum, balf or puncture tissue), and a microscope observes to obtain hints about the infectious agents, which provide critical evidence for early antibacterial agent selection[4,5]. The GRACE-VAP (Gram Stain-Guided Antibiotics Choice for VAP)[6] program indicates that M-ROSE is not inferior to guideline-based therapy and significantly reduces the use of broad-spectrum antibiotics in ventilator-associated patients. Additionally, M-ROSE has comparable accuracy to bacterial cultures and is more than 1000 times speed than bacterial cultures[7]. A retrospective study in the ICU concluded that M-ROSE can significantly shorten the time for pathogen detection while maintaining accuracy comparable to that of clinical microbiology culture[8]. M-ROSE also plays an essential role in the initial diagnosis of cryptococcus[9]. Thus, M-ROSE has shown crucial applications in etiological diagnosis and guiding antimicrobial drug selection[10–12].

Throughout the M-ROSE process, the identification of bacterial types and positive/negative status still relies on manual discrimination by experts with experience in microbiology. This manual identification of bacteria

[1]College of Pulmonary and Critical Care Medicine, Chinese PLA General Hospital, Beijing, China. [2]Chinese PLA Medical School, Beijing, China. [3]Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics-Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China. [4]MoE Key Laboratory for Biomedical Photonics, Collaborative Innovation Center for Biomedical Engineering, School of Engineering Sciences, Huazhong University of Science and Technology, Wuhan, Hubei, 430074, China. ✉e-mail: wangkaifei@126.com; xielx301@126.com

in M-ROSE images is time-consuming and labor-intensive. The scale of bacterial morphology is typically at the micrometre level, and revealing their morphology on slides requires imaging systems with high resolution. Generally, a complete slide imaging area is approximately 1–3 cm^2, and after magnification by 100 times, numerous microscopic fields can be observed. Microbiologists need to transition gradually from low-power to high-power magnification, selecting areas of interest for the classification and counting of 200–500 cells and interpretation of microorganisms such as bacteria[13,14]. For higher accuracy, clinical physicians with a foundation in cytology and microbiology are recommended for these tasks. In practical clinical applications, it typically takes 2–3 hours for an experienced clinical physician to complete the counting, interpretation, and analysis of a patient's slide sample. As a result, there is a considerable workload on clinical microbiologists, and heavy clinical work can result in relatively delayed analysis of results. Therefore, the development of an automated method for identifying bacteria in Gram-stained smear slides of lower respiratory tract specimens is a crucial aspect in enhancing the speed and quality of M-ROSE.

The advancement of deep learning affords new possibilities for bacterial recognition in M-ROSE images. However, current related research lacks the necessary support of relevant clinical data and benchmark algorithms[15–18]. On one hand, numerous studies have focused on bacterial identification in culture dishes[15–18]. Some researchers found that the accuracy of automatically predicting Staphylococcus aureus (spherical or round) and Corynebacterium diphtheriae (rod-shaped) by deep learning methods exceeded 75%[19]. Hedie et al.[20] developed a CNN-based platform that utilized stereo microscope images to automatically identify slime moulds, achieving an accuracy of 77.24% for genus identification and 88.92% for suborder identification. On the other hand, it is hard to find publicly available data concerning bacterial identification in clinical, hindering the development of automated identification algorithms in M-ROSE.

Here we collect Gram-stained smeared bacterial images from respiratory specimens and create their annotations for deep learning recognition in M-ROSE. In addition, we present a deep learning-based benchmark algorithm for bacterial detection and segmentation. In object detection, the YOLOv5 algorithm is applied for the classification of positive/negative cocci and positive/negative bacilli. For segmentation, the U-Net algorithm is employed to detect positive/negative classifications of bacteria. Our work can advance the application of deep learning in M-ROSE. This application contributes to the diagnosis of clinical infections and colonization and provides valuable references for treatment plan selection.

Methods

Ethical approval.
The study was approved by the Chinese PLA General Hospital (Approval Number: 20220322001). All participants provided verbal and written informed consent. Specifically, before bronchoalveolar lavage, all patients underwent assessments to determine indications and exclude contraindications. Prior to conducting M-ROSE, patients received adequate information and provided verbal or written informed consent. Following data selection, discharged patients with M-ROSE image data were contacted by telephone to obtain verbal informed consent. In the case of patients still hospitalized, informed consent forms were signed either by the patients themselves or their legal representatives to authorize the sharing of image data obtained from bronchoalveolar lavage and M-ROSE procedures. The study conformed to the standard set by the Declaration of Helsinki, except for registration in a database.

Data collection.
We collected specimens (bronchoalveolar lavage fluid [BALF], and endotracheal aspirates) from patients undergoing bronchoalveolar lavage. Sample slides are prepared in three steps: specimen preparation, centrifugation, and smearing. After this, a Gram-stain was applied to stain the sample slides, which involved the following steps: initial staining with crystal violet, staining with iodine solution, decoloration with alcohol, and re-staining with a safranin staining solution. During ethanol decolorization, Gram-positive bacteria typically retain the purple color due to their thicker and more complex cell wall structure, which is composed of peptidoglycan, polysaccharides, phosphatic walls, and proteins, enabling them to readily absorb foreign substances. In contrast, Gram-negative bacteria have a relatively thin cell wall, and the outer lipid membrane is easily dissolved by ethanol, causing the bacteria to lose their purple color internally, and turn red during repeated staining. Occasionally, some of the dye from the initial stain remains on the periphery, giving it a dark purple color. Therefore, the Gram-stain method allows us to distinguish between Gram-positive and Gram-negative.

After preparing the sample slide, the physician will use an optical microscope (model: OLYMPUS CX31) with a 100x magnification objective lens to observe the slide and select the area of interest for imaging, instead of capturing the entire slide. A slide sample may contain several subimages, with each region having an image size of 4912 × 3684 pixels. The submitted data consisted of 1705 images from patients who underwent bedside bronchoalveolar lavage in the RICU of Chinese People's Liberation Army (PLA) General Hospital between January 2018 and March 2022.

The dominant bacteria in these images include Gram-negative bacilli such as Acinetobacter baumannii, Klebsiella pneumoniae, Pseudomonas aeruginosa, Gram-positive bacilli such as Corynebacterium, Gram-negative cocci such as Cryptococcus, and Gram-positive cocci such as Staphylococcus aureus. Bacilli and cocci have a relatively clear morphology differentiation by shape. Gram-positive and Gram-negative organisms can be distinguished by color characteristics, with positive organisms typically bluish-violet and negative organisms typically pinkish-red, likely with slight variations due to deep staining or incomplete colorisation, and are relatively easy to identify by algorithm. The submitted data focuses only on positive/negative bacilli and cocci. Although these identifications lack subclass identification, they still contribute to guiding antibiotic treatment regimens, can reduce the use of broad-spectrum antibiotics, dynamically monitor disease changes, and assess drug efficacy.
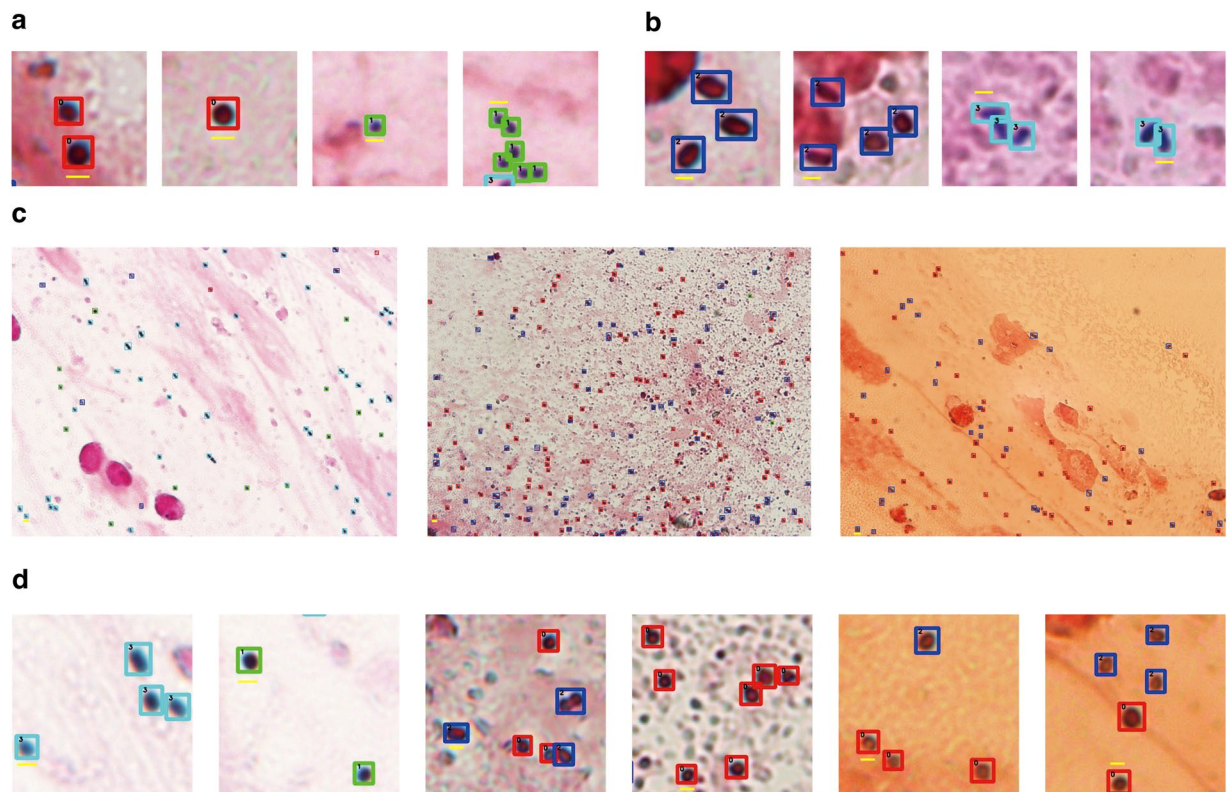
Fig. 1 Variation analysis of different types of bacteria images in data annotation. (**a**) Example of cocci with distinctions between negative cocci (red box) and positive cocci (green box). (**b**) Example of bacilli, with divisions into positive-negative bacilli (blue box) and positive bacilli (cyan box). (**c**) The annotation results from different styles of images. (**d**) Local zoom of labelled bacteria. The scale bars (yellow line) represent 1 μm.

Data annotation.　Data annotation mainly includes target detection annotation and segmentation annotation. In the target detection task, LabelImg[21] software was used to label the images, and bacteria were divided into Gram-positive/negative coccus and Gram-positive/negative bacillus. We use a rectangular box to mark the target bacteria. We asked the rectangular box to accurately frame the boundaries of the target bacteria without excessive background components, select targets with obvious bacterial morphology, and exclude targets with missing edges or irregular morphology. For the bacterial segmentation labelling task, we used LableMe software to draw the outline of the bacteria in a manually labelled box. Judge the corresponding categories by color. Gram-positive bacteria are classified as G+ and Gram-negative bacteria are classified as G-. There are obvious differences in morphology between coccus and bacillus. Microscopically, bacteria typically have smooth edges and rounded ends, while cocci are rounded and bacilli are short rod-shaped.

Fig. 1a,b respectively show typical Gram-positive/negative cocci and Gram-positive/negative bacilli. The data we collected is Gram-stained slide, for Gram-negative bacteria, Gram staining typically appears red. Due to their resilient outer membrane, ethanol faces difficulty penetrating the peptidoglycan layer to dissolve the crystal violet complex, occasionally resulting in staining as purple or developing a purple ring around the periphery (the first two panels in Fig. 1a,b). For Gram-positive bacteria, its appearance is uniformly purple in images. Because their cell walls are thick, with a higher number of layers in the peptidoglycan network, and tight cross-linking. During the decolorization process with alcohol or acetone, losing water causes the mesh to contract, additionally, the absence of lipids prevents the formation of gaps during ethanol treatment. As a result, the crystal violet and iodine complex remain firmly embedded in the cell wall, maintaining the purple color (the last two panels in Fig. 1a,b).

Differences in the shape and color of bacteria are the main basis on which we can distinguish bacteria. Fig. 1c shows images of different styles with artificial recognition; The variation in image style is mainly due to the amount of time the operator spends on staining, as well as the characteristics of the specimen. However, the features used to determine the bacteria types in Fig. 1a,b still hold in different styles of images. In other words, cocci and bacilli are distinguished primarily on the basis of morphology, while positive and negative bacteria are distinguished primarily on the basis of colour. We show in Fig. 1d a zoomed-in view of the bacterial labels of some sub-regions containing positive bacilli, positive cocci, negative cocci and negative bacilli, which can be identified by annotators with some experience. To ensure annotation accuracy, images were annotated by a total of three annotators, including a physician with extensive clinical microbiology diagnostic experience. Any discrepancies between the bacterial annotations of the two annotators were referred to the more experienced physician for determination.
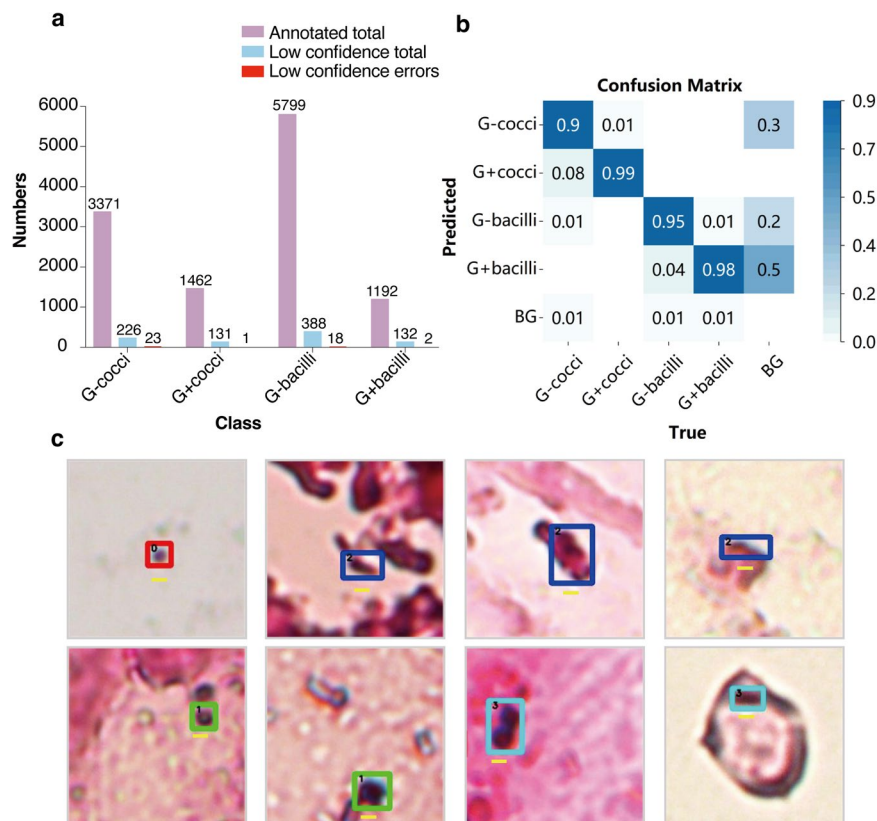
Fig. 2 Statistics and analysis of Data annotation validation. (**a**) Total number of annotated, hard-to-annotate and re-examined inconsistent bacteria. G-cocci: Gram-negative cocci, G + cocci: Gram-positive cocci, G-bacilli: Gram-negative bacilli, G + bacilli: Gram-positive cocci. (**b**) Confusion matrix derived from rechecking the four types of hard-to-annotate bacteria. (**c**) Example of hard-to-annotation situation: Gram-negative cocci (red box); Gram-negative bacilli(blue box); Gram-positive cocci (green box). Gram-positive bacilli (cyan box).The scale bars (yellow line) represent 1 μm.

Data Records

The datasets[22] can be categorized into three types: raw images directly collected with the micro-optical system, partially cropped images, and their annotated datasets. All of them are hosted on https://doi.org/10.5281/zenodo.10526360. These datasets[22] are stored in two zip files. One zip file "RawImageDataSet" includes all raw images with a size of 18.7 Gigabytes. The other zip file is organized into 3 folders, named "640DataSet", "DetectionDataSet" and "SegmentationDataSet". The "640DataSet" consists of cropped images randomly extracted from their corresponding raw images. The "DetectionDataSet" and "SegmentationDataSet" consist of detection labels and semantic segmentation labels, respectively. The detection labels are standardized and stored in text files, while the segmentation label is saved in a Json file that records the shape of each bacteria as well as its positive or negative status. Training, testing and validation datasets are also included in the files "DetectionDataSet" and "SegmentationDataSet". These datasets for training deep learning networks only include the file name. Based on the file name and its corresponding file format, we can obtain the corresponding image and its annotation.

Note that the relationship between the cropped image and the original image is indicated by the file name. For example, for a cropped image with the file name 000001_2_3 (source image name_3rd row_4th column), the first six characters denote the file name of the source image, while the subsequent characters represent different locations within that source image from which it was extracted.

Technical Validation

Data annotation validation. We have annotated a total of 3371 negative cocci, 1462 positive cocci, 5799 negative bacilli, and 1192 positive bacilli as shown in Fig. 2a. We trained all annotated datasets with a detection network that selects hard-to-annotate bacteria. For a bacterium, the IOU between its predicted and annotated boxes is less than 0.88. The bacterium is thought to be hard-annotated. The number of hard annotated bacteria is indicated by the pale blue bar in Fig. 2a. One of the annotators rechecked these screened bacteria and found 44 bacteria that were inconsistent with the annotated set, which are shown by red bars in Fig. 2a.

We also presented the confusion matrix for manual identification in the hard-to-annotate bacteria. In the confusion matrix, the x-axis horizontal coordinates of the matrix represent the true categories and the y-axis vertical coordinates represent the predicted categories. The values in the boxes represent the rate at which the true data is predicted to be in that category. The distribution of identification errors is summarized as follows: (1)

a few errors in identifying negative and positive bacteria; (2) some bacteria with smaller morphological scales are confused with impurities in the background (Fig. 2b). Furthermore, we give some of the hard-to-annotate bacteria in Fig. 2c,where the top row shows examples of hard-to-annotate cocci and bacilli. These examples stem from the small scale of the bacteria, interference from other signals, or irregular morphology. The lower row shows examples of difficult annotations for negative and positive bacteria, mainly due to deviations of these objections from the typical features of Gram-positive or Gram-negative bacteria. From the above results, there are only a small number of inconsistent discriminants in our annotated set, which is therefore acceptable given the complexity of clinical data and the diversity of sample data.

Deep learning networks for bacteria identification.   We used two deep learning networks to validate the effectiveness of data annotation and provide the benchmark algorithms for bacterial identification in our data. The annotated datasets are split into training, validation, and test datasets according to a ratio of 7:2:1 and used for training and testing the deep learning network. In the bacterial classification of our data, the manual labels are mainly based on differences in bacterial morphology and color, and accordingly, deep learning networks should capture the differences to enable automatic identification of bacterial types with high accuracy and excellent agreement with the manual annotations. Based on this, we employ the widely used detection network and semantic segmentation network to test our labelled data. The detection network is YOLOv5 (https://github.com/ultralytics/yolov5), which detects positive/negative cocci and positive/negative bacilli in images. The segmentation network is the UNet network (https://github.com/wolny/pytorch-3dunet), which is a semantic segmentation network that separates positive and negative bacteria in an image. The loss of the detection network YOLOv5 is a weighted sum of the rectangular box loss, classification loss and confidence loss with weighting factors of 0.05, 0.5, and 1.0. The loss function for semantic segmentation is the sum of the cross entropy and Dice loss.

Evaluation metrics.   We evaluate the performance of object detection algorithms using Precision, Recall, F1score and mAP. In the calculation of these metrics, it is necessary to determine whether the target is correctly detected or not. The algorithm detection box and ground truth box with $IoU > 0.45$ is judged to be correctly identified. After obtaining the correctly detected targets, we compute true positive, false positive and false negative to obtain performance metrics for evaluating detection algorithms, defined as below[23,24].

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{3}$$

$$AP = \int_0^1 p(r)dr \tag{4}$$

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{5}$$

Here, TP denotes the number of correctly detected objects, FP is the number of additional objects that do not match the ground truth, and FN is the number of objects in the ground truth that are not detected. K is the number of detected categories and p(r) refers to precision-recall curve. We still adopt the metrics in formulas (1–5) for semantic segmentation evaluation. In semantic segmentation, different values are assigned to the regions of Gram-positive and Gram-negative bacteria. If algorithmically segmented regions are dropped in annotated bounding boxes, these regions are directly binned, with each binned region representing a bacterial region. Otherwise, the region in the background region is treated as a false positive. We partition these regions into a series of connected regions, each representing a false positive bacterium. With the above operations, we can compute Precision, Recall, and F1score for bacterial semantic segmentation.

Detection results.   We manually classify bacteria into negative/positive cocci and negative/positive bacilli based on their morphological features, and accordingly, the deep learning network is also able to capture these features to achieve a fourfold classification of bacteria. Based on this, we train a detection network model to verify the effectiveness of the labelled data. We split the labelled data into a training set, a validation set, and a test set with a 7:2:1 ratio.

When the network is trained to converge, the index of the validation set reaches 0.87. This metric is a comprehensive measure of multi-category detection. This is considered correct when the overlap region between the detection box and the labelled box reaches 0.45 and they identify the same class. According to the criterion, we compute recall, precision, and mAP 0.5 scores for the training, validation, and test sets. For both validation and test sets, the mAP 0.5 score exceeds 0.73 as shown in Fig. 3b. In addition, we also present confusion matrix for the validation and test sets as shown in Fig. 3c,d.
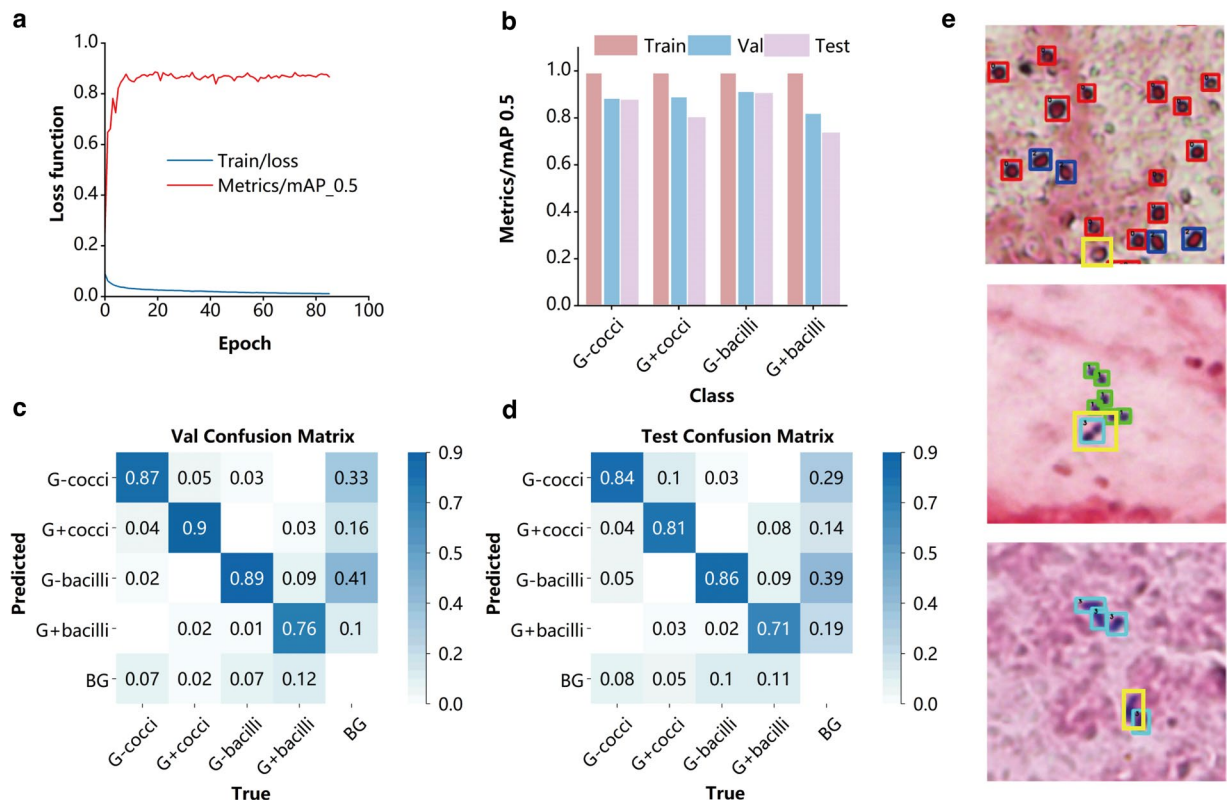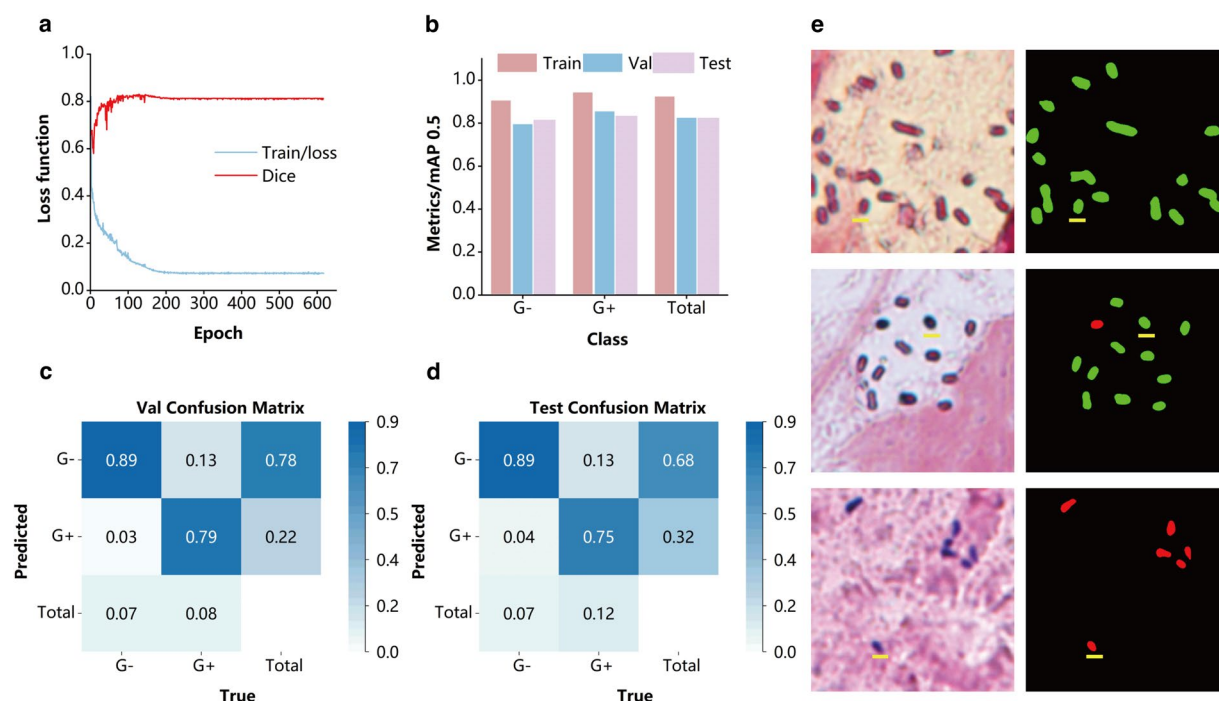
Fig. 3 The classification performance of the detection network in analysis of annotation data. (**a**) The loss function and mean Average Precision (mAP) of YOLOv5 model during the training. (**b**) The AP of four types of bacteria classifications in YOLOv5 model. (**c**) Confusion matrix for the validation sets. (**d**) Confusion matrix for test sets. (**e**) Detection examples include some typical identification errors (yellow boxes). Top, the undetected bacterium with a certain degree of deformation. Middle, two cocci that connect and arrange in a straight line are easily misidentified as bacilli. Bottom, the bacterium whose morphology is relatively fuzzy, limited by the imaging resolution.

The results show that the sources of error detections mainly include: (1) some bacteria were misidentified as background, partly due to shrinkage during the staining process; (2) some errors in identifying negative and positive bacteria. The sources of error have in common with the inconsistencies that occur in our annotations. We present some detection examples of deep learning networks. The algorithm was able to identify most of the bacteria and give the correct class, with recognition errors mainly due to mutual interference between the bacteria and some blurred edges of the bacteria. Bacterial morphology can also be affected by phagocytosis and administration of antibiotics. The presented trained deep learning networks do not give targeted solutions for these situations.

Based on the above results, we can conclude that deep learning networks can capture and detect bacterial class features in annotated data with high accuracy. Simultaneously, there is a correlation between the sources of algorithmic errors and the scenarios where annotation inconsistencies occur, confirming the effectiveness of our annotations. Moreover, it can inspire new approaches to enhance the accuracy of algorithmic detection.

Semantic segmentation results. Through our prior manual and testing validation of bacterial labelling effectiveness, we have identified some errors in the identification of both negative and positive bacteria. As a result, we have employed a semantic segmentation network to further investigate these error characteristics. The labelled set has been divided into a training set, a validation set, and a test set in a 7:2:1 ratio. Upon convergence of the network, the metrics on the validation set are 0.81, demonstrating that the semantic segmentation network is more proficient in recognizing negative and positive bacteria. The mAP values for negative and positive bacteria in the training set, validation set, and test set are presented in Fig. 4b. While these values surpass 0.8, they are still lower than the mAP values in the training set. This suggests that there are challenging cases where the algorithm struggles to recognize both negative and positive bacteria.

Through an analysis of the confusion matrix, we observed that a small percentage of bacteria are not segmented and are incorrectly identified as background, particularly some positive bacteria. These positive bacteria are more likely to be mistaken for background than negative cells due to the reduction in scale during the staining process. This reduction in scale further complicates the identification of positive cells, as their morphological characteristics become entangled with impurities in the background image, making recognition more challenging. Additionally, there is a relatively high proportion of positive bacteria incorrectly identified as negative, mainly because some positive bacteria retain red regions with complex size and distribution. These

Fig. 4 Semantic segmentation performance from annotation data. (**a**) The loss function and mean Average Precision (mAP) of U-Net model during the training. (**b**) The AP of 4 bacteria classifications in U-Net model. (**c**) Confusion matrix for the validation sets. (**d**) Confusion matrix for test sets. (**e**) Typical segmentation results of negative and positive bacteria, red for Gram-positive bacteria, green for Gram-negative bacteria.

red regions require careful manual identification, posing difficulties for automatic algorithms. Fig. 4e provides typical semantic segmentation results for negative and positive bacteria. The analysis of these results reveals that the semantic segmentation network can effectively capture the features of both negative and positive bacteria, successfully identifying the majority of instances. However, there is room for improvement in the recognition accuracy of positive bacteria, necessitating enhancements in both the detection network and the semantic segmentation network.

Usage Notes

To the best of our knowledge, this is the first publicly available clinical dataset for bacterial classification. This dataset[22] can be used to train and evaluate different AI algorithms. In addition, the dataset and the proposed benchmark algorithms encourage developers to develop algorithms and contribute to more involvement in analyzing clinical bacteria datasets. All these datasets uploaded in Zenodo allows for use without any restriction.

Limitations. Although the presented dataset includes a wide variety of images that took two years to collect, there are some limitations. First, all images were collected at the same hospital in Beijing, using a relatively fixed procedure. However, due to the standardized data acquisition process and diverse sample sources, the submitted data remains representative. Second, there are still some extremely dense bacteria that have not been labelled because it is difficult for experienced taggers to accurately judge the types of these bacteria in the image, especially their negative and positive cases.

Code availability

All codes are publicly available in the following GitHub repository https://github.com/Quanlab-Bioimage/301RoseDataSet (software licence: Apache version 2.0). The repository contains already training detection network and semantic segmentation network both of which can be used directly. Running these two networks can obtain the results presented in Figs. 3–4. The repository also contains the python codes for retraining the networks and resetting training parameters. Most of these codes are from YOLOv5 and UNet.

References

1. GBD 2019 LRI Collaborators. Age–sex differences in the global burden of lower respiratory infections and risk factors, 1990–2019: results from the Global Burden of Disease Study 2019. *Lancet Infect. Dis.* **22**, 1626–1647 (2022).
2. Ac, K. *et al.* Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **63**, (2016).
3. Gasparini, S. It is time for this 'ROSE' to flower. *Respir. Int. Rev. Thorac. Dis.* **72**, 129–131 (2005).

4. Musher, D. M., Montoya, R. & Wanahita, A. Diagnostic value of microscopic examination of Gram-stained sputum and sputum cultures in patients with bacteremic pneumococcal pneumonia. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **39**, 165–169 (2004).

5. Anevlavis, S. *et al*. A prospective study of the diagnostic utility of sputum Gram stain in pneumonia. *J. Infect.* **59**, 83–89 (2009).

6. Yoshimura, J. *et al*. Effect of Gram Stain–Guided Initial Antibiotic Therapy on Clinical Response in Patients With Ventilator-Associated Pneumonia: The GRACE-VAP Randomized Clinical Trial. *JAMA Netw. Open* **5**, e226136 (2022).

7. Li, T. *et al*. Microbiology rapid on-site evaluation: a better method for Mucoid Pseudomonas Aeruginosa diagnosis in bronchiectasic patients. *Eur. Rev. Med. Pharmacol. Sci.* **26**, 1738–1742 (2022).

8. Tao, Y. *et al*. Application of microbiological rapid on-site evaluation in respiratory intensive care units: a retrospective study. *Ann. Transl. Med.* **10**, 7 (2022).

9. Zhang, X., Ye, L. & Hong, L. Rapid on-site evaluation of pulmonary cryptococcosis: A preliminary assessment. *Asian J. Surg.* **43**, 1101–1102 (2020).

10. Katayama, Y. *et al*. Usefulness of Gram staining of tracheal aspirates in initial therapy for ventilator-associated pneumonia in extremely preterm neonates. *J. Perinatol. Off. J. Calif. Perinat. Assoc.* **30**, 270–274 (2010).

11. Gottesman, T. *et al*. The accuracy of Gram stain of respiratory specimens in excluding Staphylococcus aureus in ventilator-associated pneumonia. *J. Crit. Care* **29**, 739–742 (2014).

12. Yoshimura, J. *et al*. Impact of Gram stain results on initial treatment selection in patients with ventilator-associated pneumonia: a retrospective analysis of two treatment algorithms. *Crit. Care Lond. Engl.* **21**, 156 (2017).

13. Meyer, K. C. *et al*. An Official American Thoracic Society Clinical Practice Guideline: The Clinical Utility of Bronchoalveolar Lavage Cellular Analysis in Interstitial Lung Disease. *Am. J. Respir. Crit. Care Med.* **185**, 1004–1014 (2012).

14. Junan Medical cell platform expert committee. Chinese expert consensus on cell morphological examination of bronchoalveolar lavage fluid. *J Mod Lab Med.* **38**(03), 11–16+23 (2023).

15. Alouani, D. J. *et al*. Deep Convolutional Neural Networks Implementation for the Analysis of Urine Culture. *Clin. Chem.* **68**, 574–583 (2022).

16. Makrai, L. *et al*. Annotated dataset for deep-learning-based bacterial colony detection. *Sci. Data* **10**, 497 (2023).

17. Smith, K. P., Kang, A. D. & Kirby, J. E. Automated Interpretation of Blood Culture Gram Stains by Use of a Deep Convolutional Neural Network. *J. Clin. Microbiol.* **56**, e01521–17 (2018).

18. Signoroni, A. *et al*. Hierarchical AI enables global interpretation of culture plates in the era of digital microbiology. *Nat. Commun.* **14**, 6874 (2023).

19. Treebupachatsakul, T. & Poomrittigul, S. Bacteria Classification using Image Processing and Deep learning. in *2019 34th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)* 1–3 (IEEE, 2019).

20. Sajedi, H., Mohammadipanah, F. & Pashaei, A. Automated identification of Myxobacterial genera using Convolutional Neural Network. *Sci. Rep.* **9**, 18238 (2019).

21. Russell, B. C., Torralba, A., Murphy, K. P. & Freeman, W. T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **77**, 157–173 (2008).

22. Wang, X. Clinical Bacteria Dataset. *Zenodo.* https://doi.org/10.5281/zenodo.10526360 (2024).

23. Goutte, C. & Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. in *Advances in Information Retrieval* (eds. Losada, D. E. & Fernández-Luna, J. M.) **3408** 345–359 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2005).

24. Everingham, M., Van Gool, L., Williams, C. K., Winn, J. & Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **88**, 303–338 (2010).

Acknowledgements

Author contributions

X. Wang contributed to research plan determination, acquisition of all data, data analysis, result discussion, and manuscript writing. Y. Shi, F. Xie, Z. Duan, Y. Hu, H. Fu, X. Shi collected the samples. T. Quan, S. Guo and X. Qu carried out data proposing and analysis. L. Xie and K. Wang supervised the project, designed the experiment, and reviewed the manuscript. All authors contributed to the study design.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.W. or L.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.