AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Clinical Microbiology®

MINIREVIEW

# The Use of Machine Learning for Image Analysis Artificial Intelligence in Clinical Microbiology

Bethany L. Burns,[a] Daniel D. Rhoads,[a,b,c] Anisha Misra[a]

[a]Department of Laboratory Medicine, Cleveland Clinic, Cleveland, Ohio, USA
[b]Department of Pathology, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, Ohio, USA
[c]Infection Biology Program, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA

**ABSTRACT** The growing transition to digital microbiology in clinical laboratories creates the opportunity to interpret images using software. Software analysis tools can be designed to use human-curated knowledge and expert rules, but more novel artificial intelligence (AI) approaches such as machine learning (ML) are being integrated into clinical microbiology practice. These image analysis AI (IAAI) tools are beginning to penetrate routine clinical microbiology practice, and their scope and impact on routine clinical microbiology practice will continue to grow. This review separates the IAAI applications into 2 broad classification categories: (i) rare event detection/classification or (ii) score-based/categorical classification. Rare event detection can be used for screening purposes or for final identification of a microbe including microscopic detection of mycobacteria in a primary specimen, detection of bacterial colonies growing on nutrient agar, or detection of parasites in a stool preparation or blood smear. Score-based image analysis can be applied to a scoring system that classifies images *in toto* as its output interpretation and examples include application of the Nugent score for diagnosing bacterial vaginosis and interpretation of urine cultures. The benefits, challenges, development, and implementation strategies of IAAI tools are explored. In conclusion, IAAI is beginning to impact the routine practice of clinical microbiology, and its use can enhance the efficiency and quality of clinical microbiology practice. Although the future of IAAI is promising, currently IAAI only augments human effort and is not a replacement for human expertise.

**KEYWORDS** artificial intelligence, clinical microbiology, machine learning

Images often serve as primary results in clinical microbiology, which are then interpreted. Examples include stained microscopic slide preparations, microbial colony morphology on nutrient agar, and growth in antimicrobial susceptibility tests. Although these images are traditionally viewed using analog approaches with the human eye and interpreted with the human mind, the growing transition to digital microbiology in clinical laboratories creates the opportunity to interpret images using software. These software analysis tools can be designed to use human-curated knowledge and expert rules (1), but more novel Artificial Intelligence (AI) approaches such as machine learning (ML) are being integrated into clinical microbiology practice. These image analysis AI (IAAI) tools are beginning to penetrate routine clinical microbiology practice, and their scope and impact on routine clinical microbiology practice will continue to grow.

A cursory PubMed search for the term "artificial intelligence" produces over 180,000 results from the last decade. The amount of software development and the number of medical studies using AI can feel overwhelming, and only a small number of these studies will be reviewed in this manuscript. Some authors have written to introduce AI to the novice (2), and others have described AI applications specific to laboratory medicine or described clinical microbiology informatics in general (3, 4). This minireview is not meant to be a general introduction into AI methodology nor an exhaustive report of the opportunities to use AI

**Editor** Romney M. Humphries, Vanderbilt University Medical Center



Address correspondence to Daniel D. Rhoads, daniel.rhoads@case.edu.

The authors declare a conflict of interest. DDR has received research support from Altona, BD, bioMérieux, Bio-Rad, Cepheid, Cleveland Diagnostics, Luminex, HelixBind, Hologic, Qiagen, Q-Linea, Selux, Specific Diagnostics, Thermo Fisher, and Vela; and DDR has or has had advisory relationships with Luminex, Next Gen Diagnostics, Renascent Diagnostics, Roche, and Seegene.

**Published** 3 July 2023

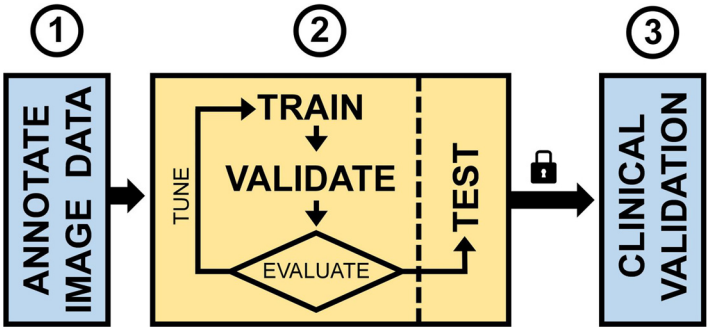


**FIG 1** Typical flow of development and clinical validation of a supervised machine learning tool for use as an *in silico* diagnostic device in the analysis of image data. Blue boxes (1 and 3) represent steps requiring expert clinical microbiology input and evaluation. Yellow box (2) represents algorithm development performed by data scientists. Step 1 requires data curation and labeling, and this labeled data serves as input for the training of the data algorithm. Step 2 uses the annotated data to train, validate, and test the machine learning algorithm. Iterative *in silico* training and validation steps are performed, and the algorithm is tuned to perform optimally. Then, an independent test set is analyzed to confirm the algorithm is performing as expected *in silico*. Once the algorithm development is complete, the algorithm is essentially locked into its final state. In Step 3, the performance of the final algorithm is evaluated in the clinical laboratory as an *in silico* diagnostic device. If the algorithm performs well in the real-world clinical laboratory validation attempt, then the algorithm can be implemented as an *in silico* diagnostic device and used for patient care purposes.

in clinical microbiology. Instead, this brief review discusses a selection of publications that describe IAAI applications designed for use in routine clinical microbiology practice. This review is intended to provide the reader with an overview of the current state and select details relevant to clinical microbiology practice, such as approaches to algorithm development and considerations for clinical validation (Fig. 1). Certain AI terminology are emphasized where the authors feel these details could be helpful to the reader in their future interactions with IAAI in clinical microbiology practice (Box 1).

## OPPORTUNITIES FOR ARTIFICIAL INTELLIGENCE IN CLINICAL MICROBIOLOGY

Modern technologies facilitate the creation and storage of large amounts of digital data, but traditional data analysis tools often seem insufficient to thoroughly analyze these expansive data sets. AI algorithms can be trained to recognize patterns and draw inferences from data, including image data. ML is one approach that automates and streamlines analysis of large data sets. Rather than a single entity, ML denotes computer systems that can learn and adapt without explicit instructions. In clinical microbiology, ML can make predictions about antibiotic susceptibility or discover new determinants of antimicrobial resistance (5–7). ML can also enhance the quality and efficiency of image interpretation in clinical microbiology, which is the focus of this minireview.

## USING ARTIFICIAL INTELLIGENCE TO INTERPRET IMAGE DATA

In clinical microbiology practice, image analysis can be performed using static "snapshot" images as input data (e.g., nutrient agar petri dish) or using a set of images encompassing an entire microscope slide (i.e., whole-slide image [WSI]). IAAI uses the image data as an input and then produces an output classification based on a set of rules that can be as simple as expert-determined rules-based logic (8). However, supervised ML uses a different approach wherein unprocessed data (e.g., digital image) are interpreted using a complex algorithm that has been developed by using a set of accurately annotated images as its reference standard. Various types of algorithms are included within the broad category of supervised ML, but one commonly used approach is the convolutional neural network (CNN), which works particularly well for IAAI. CNN mimics the neural networks of the human optical cortex by interconnecting convoluted algorithms (8).

The development of a supervised ML algorithm includes 3 stages: training, validation, and testing (Fig. 1). Once the ML algorithm development is complete, then a clinical

## Definitions and descriptions of key terms

**Artificial intelligence:** Computer algorithms or software designed to perform simple (e.g., expert rules) or complex (e.g., image analysis) interpretations of data, which traditionally have been interpreted by humans.
**Machine Learning:** A type of artificial intelligence algorithm that is commonly used in contemporary practice and which is developed without human input of explicit logic rules but instead is developed by enabling the computer to draw inferences directly from data.
**Deep Learning:** A type of machine learning that uses multiple layers of nodes in its neural network.
**Neural Network:** Computing system inspired by the optic cortex of animal brains that use a collection of nodes loosely modeled on neurons and their numerous interconnections. Multiple inputs can influence the output of a node, and the output of a node can act as an input for multiple other nodes.
**Convolutional Neural Network:** A type of neural network that uses convolution mathematical operations in the design of the algorithm. These convolutional neural networks are commonly used in contemporary image analysis artificial intelligence software.
**Supervised learning:** A type of machine learning that relies on human insight. For example, a microbiologist might label photomicrographs as containing or not containing a certain microbe of interest, and those labeled images could then be used as a data source for developing a machine learning algorithm to identify the microbe in a photomicrograph.
**Unsupervised learning:** A type of machine learning that does not rely on human insight for classification. For example, grouping images based on their similarity and differences can be performed using an unsupervised approach.
**Precision (data science):** Precision is calculated by dividing a machine learning algorithm's true positive results by all positive results reported. In laboratory medicine, this equation is the same as that which is used to calculate positive predictive value.
**Precision (clinical microbiology):** Precision describes the repeatability of achieving the same or similar test results.
**Recall (data science):** Recall is calculated by dividing a machine learning algorithm's detected true positives by all true positives. In laboratory medicine, this equation is the same as that which is used to calculate sensitivity or positive percent agreement.
**Training data set (machine learning):** Input data that is used for the initial development of a machine learning algorithm.
**Validation data set (machine learning):** The machine learning model created using the training data set is evaluated using different data, the validation data set. The findings from the validation are considered, and human intervention is used to create minor changes (aka "tune") the algorithm with the intention of optimizing the algorithm's performance.
**Test data set (machine learning):** The machine learning model that has been trained and validated is assessed using a final independent data set. This final independent data set is the test data set.
**Clinical Validation (clinical microbiology):** Evaluation or establishment of the performance characteristics of a test using clinical samples, wherein the performance characteristics are ultimately deemed acceptable or unacceptable for use as a diagnostic device.
**Feature.** In image analysis artificial intelligence, features are classified by the algorithm. Features can be low-level characteristics like edges or colors, or they can be high-level characteristics comprised of low-level features like a colony or an acid-fast bacillus. A low-level feature is the most primitive component of an image that can be useful in classification, and a high-level feature includes the component of an image that can be classified as a microbial entity.
**Rare event detection.** Identification of only a single high-level feature within an image is needed in order to classify the image as abnormal.
**Score-based or Categorical classification.** Interpretation of the image is not directly linked to the presence or absence of a single high-level microbial feature within the image, but either the image as a whole (categorical) or multiple microbial features within the image (score-based) are considered to determine whether the image is interpreted as normal or abnormal.

validation can be performed. A developed ML algorithm can be considered an *in silico* diagnostic (ISD) device or Software as a Medical Device (SaMD) (9), and it should be clinically and technically validated similarly to an *in vitro* diagnostic (IVD) device before it is deployed for clinical use to inform patient care decisions.

When using an ISD IAAI device, the input image(s) serve as the laboratory results, and the algorithm's output is the interpretation of the results. IAAI ISD applications can vary depending on the diagnostic need, but for simplicity and clarity, this review separates the IAAI applications into 2 broad classification categories: (i) rare event detection/classification or (ii) score-based/categorical classification.

## RARE EVENT DETECTION

IAAI can be used to detect rare events. Rare event detection can be used for screening purposes or for final identification of a microbe. Results can be qualitative or quantitative. Examples of rare event detection IAAI include microscopic detection of mycobacteria in a primary specimen, detection of young bacterial colonies growing on nutrient agar, or detection of parasites in a stool preparation and/or blood smear.

**Mycobacteria detection in primary specimens.** *Mycobacterium* spp. are small (2 to 5 $\mu$ m), acid-fast bacilli (AFB) that require acid-fast staining for routine microscopic visual detection. AFB detection for anatomic pathology applications has been developed. Pantanowitz et al. created IAAI to screen for AFB in WSI (10). The algorithm consisted of a patch-based approach where each WSI was divided into non-overlapping image fragments (i.e., "patches") that were qualitatively interpreted as either positive or negative for AFB. This design was repeated twice; first with a more sensitive approach, followed by a more specific approach to reduce false positive results. When using the area under the receiver operating characteristic curve (AUC) as the marker of algorithm performance, the more sensitive algorithm when used alone yielded 95% AUC, the more specific algorithm when used alone yielded 92% AUC, and the 2 algorithms when used in series yielded a 96% AUC, which outperformed the use of a single algorithm.

Positive patches were digitally circled in the WSI, so a human expert user could view the feature of interest. The study concluded that IAAI improved mycobacterial detection regardless of the number of bacilli present in the tissue, and the algorithm facilitated shorter review duration and decreased perceived effort by the human reader compared to manual review using conventional light microscopy. This algorithm-assisted approach resulted in similar specificity to pathologist microscope review and review using WSI, but improved sensitivity, which yielded fewer false negative interpretations. A significantly higher percentage of algorithm-assisted reviews were classified by pathologists as easy (93.5%) compared with manually screening slides with a microscope (43.8%) or manually reviewing WSI (38.4%).

IAAI could be used to screen primary samples in the clinical microbiology lab for AFB using either light or fluorescence microscopy. Two studies described the use of MetaSystems (Altlussheim, Germany) to analyze fluorescent scans as the data inputs for AFB detection (11, 12). The human time needed to interpret each slide was reduced by 90% when using AI support, but the accuracy of the AI was not the same as analog slide review. At the slide level, the AI had high sensitivity (97%) but low specificity (13%) (11). Using an expert human to review digital images identified by the algorithm as "suspected AFB" helped improve the specificity over using the algorithm alone. The specificity improved from 13% to 89%. One conclusion by the Stanford group was that low resolution in combination with suboptimal focus likely impaired the algorithm's performance. Optimal image focus and adequate resolution of the potential microbe are perpetual challenges when using digital microscopy for clinical microbiology applications (13). This challenge persists regardless of whether the interpreter is a human or computer.

**Bacteria colony detection.** IAAI of photos of nutrient agar could be used to shorten time to organism detection, decrease the time required to review culture plates, and/or increase the accuracy of the interpretation of the growth. Faron et al. used IAAI to iden-tify vancomycin-resistant enterococcus (VRE) colonies from images of more than 100,000 agar plates using routine bacteriology automation equipment (14). The trained algorithm identified the blue colonies growing on the chromogenic VRE media. The reference standard used in the study was a manual human interpretation of the same culture plates. The IAAI had imperfect negative percent agreement (90%) when compared to the manual refer-ence method; however, 4.8% (499/10,348) of the "false positive" IAAI determination were identified to be false negative results using the reference method (manual human

interpretation) when reviewed more closely. The IAAI algorithm had perfect negative predictive value with no false-negatives identified compared to the reference method, which enabled 84% (87,979/104,730) of the cultures to be identified as negative for VRE using IAAI.

**Ova and parasite detection.** Identification of parasites in stool specimens is labor-intensive and requires high technical competence. In the United States, the pretest probability of a true ova and parasite (O&P) infection is low, requiring hours of expert microbiologists' time per positive finding. Mathison et al. developed a deep learning CNN to detect parasites on trichrome-stained fecal samples (15). Slides were digitized at an $\times$82.4 magnification with 0.1214 micron per pixel resolution. Fields were scanned at 3 different depths, and the software selected the most in-focus plane from the Z-stack. The scanned fields were stitched together to form the complete scanned image. A total of 127 positive slides trained the software to identify a panel of common parasites: *Giardia duodenalis* cyst, *Giardia duodenalis* trophozoite, *Blastocystis* spp., *Dientamoeba fragilis*, *Entamoeba* non-*hartmanni* trophozoite, *Entamoeba hartmanni* trophozoite, *Chilomastix mesnili* trophozoite, *Endolimax nana*, and *Iodamoeba buetschlii* trophozoite. The trained model also identified red blood cells, white blood cells, and yeast to help to prevent the algorithm from potentially misidentifying those cell types as parasites.

Human experts supplemented the ML algorithm at clinical implementation as the IAAI ISD device screened WSI data for parasites, but confirmation of the parasites was performed by the human expert before reporting. Excellent positive and negative agreement (99% PPA and 98% NPA) occurred between the AI-augmented workflow and traditional manual microscopy at the specimen-level. The AI model, however, had a 32-fold lower limit of detection compared to the traditional manual workflow. Precision-recall plots evaluated the model's continued performance after the machine learning validation stage which yielded a total recall of 83% to 92%, depending on accepted confidence scores (16).

In daily clinical practice, Mathison et al.'s AI-augmented workflow assisted medical laboratory scientists and parasitologists, and the performance data (positive and negative agreement) incorporated human experts in the workflow (15). For example, if the model flagged slides containing parasites, the glass slides were manually reviewed by a medical laboratory scientist for confirmation and sent to a trained parasitologist for final identification. Likewise, occasional suspicious false positive slides were manually reviewed by medical laboratory scientists, and slides lacking parasites required a shorter duration of human review before confirming the absence of parasites. This AI-augmented approach to identifying and characterizing rare events has the potential to positively impact efficiency in high volume laboratory settings.

**Blood parasites.** The diagnosis of babesiosis requires not only a qualitative identification of the organism, but also a quantitative determination of the percent of parasitemia. Some authors developed IAAI algorithms to automate the detection and quantification processes. Durant et al. trained an IAAI model with binary image classification that calculated the percent parasitemia from a clinical validation set (17). The model demonstrated high precision in the development training and testing; however, the clinical validation phase identified false positive errors, which were attributed to rouleaux, which was not adequately encountered in the development phase. This scenario highlights the need for rigorous clinical validation and quality monitoring regardless of how an algorithm performs in the test environment.

Similar to babesiosis, malaria requires organism identification and quantitative determination of percent parasitemia. The World Health Organization (WHO) originally created a 55-slide set for their External Competence Assessment of Malaria Microscopists (ECAMM) program, which later was used as a clinical validation to evaluate the performance of a fully automated malaria diagnostic IAAI platform, EasyScan GO. The EasyScan GO was developed using more than 500 slides from 11 countries. In the WHO reference set, EasyScan GO accurately detected the presence or absence of malaria in 33/35 (94% sensitivity) of the slides with parasites, and it detected no false positives cases (20/20; 100% specificity). In the cases in which malaria was detected, the IAAI identified parasite quantitation within 25% of the reference count in half of cases (18). The EasyScan GO algorithm met criteria for the highest accreditation level (1) according to the ECAMM evaluation slide set rubric. The

algorithm was less successful at species identification accuracy (82.9%), corresponding with a lower accreditation level (2). The authors concluded that the EasyScan GO platform could be a valuable tool for assessing subsequent serial blood smears for parasitemia when assessing treatment efficacy.

Although IAAI systems are not yet routinely used in malaria diagnosis and monitoring, algorithms have been developed and have demonstrated performance on par with expert parasitologists. It is likely that IAAI will play a future role in clinical laboratory detection and monitoring of malaria.

## SCORE-BASED OR CATEGORICAL CLASSIFICATION

Besides rare event detection, image analysis can be applied to a scoring system that classifies images *in toto* as its output interpretation. The interpretation is not necessarily directly linked to the presence or absence of a single feature in the image.

**Nugent score for classification of bacterial vaginosis.** Bacterial vaginosis (BV) is a disruption of the normal microbiome resulting in an inappropriate overgrowth of anaerobic bacteria. The reference standard for laboratory diagnosis of BV is the Nugent score by Gram stain (19). The Nugent score uses the semi-quantification of Gram-positive rods, Gram-negative coccobacilli, and curved Gram-negative rods on Gram stain, which ideally correlates with *Lactobacillus* spp., *Gardnerella* spp., and *Mobiluncus* spp., respectively. The Nugent score corresponds to categories with clinical interpretations; score 0 to 3 (*Lactobacillus* spp. dominant; consistent with normal microbiome), score 4 to 6 (mixed morphotypes indicative of altered vaginal flora), and score 7 to 10 (absence of *Lactobacillus* spp. and predominance of other morphotypes diagnostic of BV).

Wang et al. developed a CNN model (NugentNet) that analyzed and scored vaginal Gram stains (20). The model was trained using more than 23,000 images and validated using another 5,000 images. The software NugentNet was able to interpret 100 images in 2.4 s. After tuning the model by retraining the CNN to accurately interpret images originating from different laboratories (e.g., cameras with variable white balances and resolutions), the model showed diagnostic improvement in sensitivity (24% increase), specificity (9.5% increase), and accuracy (10.2% increase) when compared to the original CNN model trained on data from a single laboratory using a single camera. The accuracy of NugentNet (75.1%) rivaled an average human reader composite interpretation reference standard including medical laboratory scientists and obstetricians. NugentNet was less sensitive than the average human reader (89.0% versus 94.9%) in detecting BV, but was more specific (85.0% versus 74.6%). In general, the model performed slightly below in sensitivity, specificity, and accuracy (89.0%, 85.0%, and 75.1%) when compared to an obstetrician (94.4%, 93.9%, and 80.9%); but it was more specific and accurate when compared to the average medical laboratory scientist (96.5%, 62.2%, and 68.5%).

This performance demonstrates a general truth that is observed in well-performing IAAI, in which the software learns to interpret as the average of the expert users who supervised the categorization of the reference images. For this reason, it is essential to use expert-curated images for IAAI algorithm training to enable the highest quality algorithm performance.

**Bacterial culture interpretation.** Bacteriology automation includes digital image capture of nutrient agar plates as part of standard of care clinical microbiology practice (21). Software to detect rare events from culture plates exists, which may enable earlier or more sensitive detection of growth on chromogenic media used for screening and classification of mixtures using urine chromogenic media (14, 22–27). Digital plate reading software has the potential to automate the interpretation of primary cultures using non-chromogenic media as well. Copan WaspLab (Brescia, Italy), BD Kiestra (Drachten, Netherlands), and Clever Culture Systems APAS (Zurich, Switzerland) offer IAAI solutions to aid in the classification of growth for bacterial cultures including urine cultures. APAS uses an approach that classifies urine culture plates based on colony counts, and this system was the first to receive FDA clearance for using IAAI for culture plate interpretation (28, 29). Copan markets PhenoMatrix (30, 31) and BD markets a Urine Culture App (32, 33), which support a medical laboratory scientist in the interpretation of bacterial growth (or no

growth) on agar media from primary specimens. These BD and Copan solutions detect colonies, enumerate them, and classify them. Then, expert rules create decision trees based on the number of colonies, variety of colony types, and features of the colony (e.g., color). This classification by IAAI with an overlay of expert rules can help to create workflow efficiency in the laboratory (30).

Other groups found similarly high sensitives when using image software to quantify and differentiate bacterial colonies on chromogenic agar. Faron et al. created an image analysis model that demonstrated high sensitivity (99.8%) for bacterial colony growth detection with less specificity (68.5%) when compared to manual reading for growth detection (quantitation agreement at the 10,000 CFU/mL interpretive breakpoint point was 88.9%) (25). With the assistance of IAAI, the average result time decreased for both positive and negative results (25). One study reported success with automated plate reading software using MacConkey agar, demonstrating high sensitivity (99.8%) for growth detection and acceptable specificity for screening culture growth (72%) (31).

Another study demonstrated the ability to forgo the supervised feature annotation where an IAAI algorithm is trained to identify and classify individual colonies on a plate. Instead of training the algorithm to accurately detect, enumerate, and classify colonies on a plate (which subsequently requires expert rules to yield an interpretation), the images were instead analyzed and classified *in toto* without specifically teaching the IAAI algorithm how to count colonies or discriminate mixed growth (34).

**Antimicrobial susceptibility interpretation.** IAAI can be used to aid in antimicrobial susceptibility testing. Investigators attempted to manually measure disk diffusion zones of inhibition using the BD Kiestra digital images (35), but at this time, the measurement is not yet automated by AI with the BD Kiestra system. Copan Radian (36–39), SIRscan (40), and BioMic V3 (Giles Scientific, New York, NY) (41, 42) offer IAAI support for measurement of disk diffusion zones of inhibition. BioMic V3 (https://www.biomic.com/broth-microdilution.html) and Sensititre Vizion (ThermoFisher; Waltham, Massachusetts) use IAAI for interpretation of broth microdilution susceptibility testing (43).

## BENEFITS OF IAAI

IAAI implementation in microbiology laboratories can decrease result turnaround times, decrease hands on time, and in some cases improve sensitivity of detection. In one study, enhanced workflow using bacteriology automation in combination with IAAI yielded an estimated decrease in hands on time of 80% for negative cultures (44). In another study using IAAI, a 4 h decrease in time-to-results was achieved for urine cultures (25). IAAI assisted algorithms can also decrease perceived work effort and substantially shorten operator time required to detect rare events (10).

Often IAAI algorithms demonstrate high sensitivity and negative predictive value, so negative results (e.g., no growth on culture or no amebae identified in a stool exam) are highly reliable and could potentially be reported without human intervention or manual review. Streamlining the workflow for managing positive results from IAAI applications could enhance quality and efficiency as medical laboratory scientists and clinical microbiologists can potentially focus on pathogen identification and avoid some of the fatigue that is associated with arduous review of negative samples.

## CHALLENGES OF IAAI

The quality of the specimens and images should be carefully considered when attempting to use IAAI as an ISD device. Like humans, IAAI algorithms may reject or not process insufficient quantity specimens or specimens with poor preparation (44). It is important that unusual samples or samples not encountered in adequate quantity during IAAI development are not inaccurately given an errant interpretation by an AI platform (17). Changing the resolution, brightness, and physical area of sampling can impact IAAI interpretation (20). Although some studies reported blurriness as a criterion for rejection, other platforms interpreted data even if images were out of focus (17, 18, 44, 45). Background artifact may also

impact IAAI's ability to reliably assess specimens and the presence of other disease processes may become confounding variables (17).

Suitable image resolution poses a challenge when implementing IAAI in the clinical microbiology lab. For example, Wang et al. recreated their original CNN model to a new CNN model, NugentNet, adding more convolutional layers to adapt to the resolution of Gram stains obtained with different cameras in different laboratories (20). *Gardnerella vaginalis* image data essentially disappeared once the image size was compressed to 224 × 224 pixels. The new model increased the resolution to 1,024 × 768 pixels to maintain *G. vaginalis* detection.

Although IAAI algorithms can improve sensitivity, accuracy is not always significantly improved (10). Increased sensitivity can lead to false positive findings that ultimately skew IAAI findings (16). IAAI model developers can influence the sensitivity and specificity of models by training multiple models, implementing algorithms and picking the best fit, and adjusting the pretest probability to influence statistical performance (17). IAAI models are more accurate when trained with large data sets (17). Finding ample testing samples can take significant time and effort and may not be available during the development period (10).

IAAI models have the potential to learn and grow, which may lead them to deviate from their initial validation testing. However, learning IAAI approaches are likely not to be used in clinical practice without discrete clinical validation phases used to verify the performance of the ISD after each iterative software change. Thorough clinical validation is imperative (16), and even reliable IAAI are likely to use a human review to double-check and verify important diagnostic findings. Currently, IAAI fits best into routine clinical practice when used to augment medical laboratory scientist and clinical microbiologist efforts and not to replace them. These IAAI tools are not yet used routinely as ISD devices that operate without human review of positive samples.

In addition to technical challenges of IAAI described above, limitations in laboratory personnel's expertise in AI and the lack of widely implemented equipment capable of routine automated image-capture are two barriers slowing the development and implementation of IAAI tools in clinical microbiology practice.

## FUTURE CONSIDERATIONS

**Imagining the future impact of AI on clinical microbiology labs.** Unrealized opportunities remain for IAAI to impact the clinical microbiology lab. We expect IAAI to play a major role in routine workflows in clinical microbiology over the coming decade (15, 46). The high sensitivity and NPV of IAAI tools can be used to screen out negative samples, which can make high volume workflows more efficient and decrease time spent on labor-intensive tasks. IAAI tools could also be used to provide support in austere settings where on-site expertise is limited but digital clinical microbiology workflows are present (47). Additionally, IAAI software can be designed and used as part of a laboratory's quality management system to double-check manual image interpretations (44, 46).

Although IAAI will create efficiency and improve quality, we do not anticipate IAAI to solve the staffing shortages that have been felt in clinical microbiology labs for decades (48, 49). Currently, IAAI systems require human expertise to develop, validate, and maintain them. We expect this interdependent relationship between IAAI and human expertise to remain the standard of practice for the foreseeable future when using IAAI as an ISD device.

**Preparing for IAAI implementation opportunities.** Clinical microbiology laboratory personnel should work to become familiar with the IAAI applications and recognize that IAAI will likely become part of standard of care in clinical microbiology practice during this decade (49). One forthcoming application of IAAI in the clinical microbiology lab is quality assurance and passive quality monitoring. Specifically, IAAI can facilitate quality assurance for a variety of laboratory tests including PCR, lateral flow assays, and bacterial plate cultures (34, 46, 50). We expect IAAI to expand the possibilities of quality assurance beyond what is currently implemented or imagined for bacteriology and microscopy, and clinical microbiologists should look for opportunities to use IAAI to improve quality and efficiency.

Standardized approaches that can be used for clinical validation of IAAI ISD devices are beginning to be imagined. When considering developing and/or implementing an IAAI tool as an ISD, clinical microbiologists should be informed by best practices that have been identified by others who have already attempted clinical validation (Fig. 1) and current U.S. Food & Drug Association guidance (51). The reader is referred to the studies of Mathison, Wang, and Alouani for good examples of how to approach IAAI validation (15, 20, 34).

## CONCLUSION

IAAI is beginning to impact the routine practice of clinical microbiology. The use of IAAI can improve the efficiency and quality of clinical microbiology practice. IAAI can augment human effort but is not a replacement for human expertise. Future studies should continue to describe the successes and failures of the development and clinical validation of IAAI ISD devices, so we can collectively learn how to use IAAI in clinical microbiology to maximize the benefit it can provide.

## ACKNOWLEDGMENT

D.D.R. has received research support from Altona, BD, bioMérieux, Bio-Rad, Cepheid, Cleveland Diagnostics, Luminex, HelixBind, Hologic, Qiagen, Q-Linea, Selux, Specific Diagnostics, Thermo Fisher, and Vela; and D.D.R. has or has had advisory relationships with Luminex, Roche, and Seegene; and equity in Next Gen Diagnostics and Renascent Diagnostics.

## REFERENCES

1. Winstanley T, Courvalin P. 2011. Expert systems in clinical microbiology. Clin Microbiol Rev 24:515–556. https://doi.org/10.1128/CMR.00061-10.
2. Harrison JH, Gilbertson JR, Hanna MG, Olson NH, Seheult JN, Sorace JM, Stram MN. 2021. Introduction to artificial intelligence and machine learning for pathology. Arch Pathol Lab Med 145:1228–1254. https://doi.org/10.5858/arpa.2020-0541-CP.
3. Herman DS, Rhoads DD, Schulz WL, Durant TJS. 2021. Artificial intelligence and mapping a new direction in laboratory medicine: a review. Clin Chem 67:1466–1482. https://doi.org/10.1093/clinchem/hvab165.
4. Rhoads DD, Sintchenko V, Rauch CA, Pantanowitz L. 2014. Clinical microbiology informatics. Clin Microbiol Rev 27:1025–1047. https://doi.org/10.1128/CMR.00049-14.
5. Pearcy N, Hu Y, Baker M, Maciel-Guerra A, Xue N, Wang W, Kaler J, Peng Z, Li F, Dottorini T. 2021. Genome-scale metabolic models and machine learning reveal genetic determinants of antibiotic resistance in *Escherichia coli* and unravel the underlying metabolic adaptation mechanisms. mSystems 6:e0091320. https://doi.org/10.1128/mSystems.00913-20.
6. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davis JJ. 2019. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal *Salmonella*. J Clin Microbiol 57:e01260-18. https://doi.org/10.1128/JCM.01260-18.
7. Humphries RM, Bragin E, Parkhill J, Morales G, Schmitz JE, Rhodes PA. 2023. Machine-learning model for prediction of cefepime susceptibility in *Escherichia coli* from whole-genome sequencing data. J Clin Microbiol 61. https://doi.org/10.1128/jcm.01431-22.
8. Smith KP, Wang H, Durant TJ, Mathison BA, Sharp SE, Kirby JE, Long SW, Rhoads DD. 2020. Applications of artificial intelligence in clinical microbiology diagnostic testing. Clinical Microbiology Newsletter 42:61–70. https://doi.org/10.1016/j.clinmicnews.2020.03.006.
9. FDA U.S. Food and Drug Administration. 2020. Software as a medical device (SaMD). https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd.
10. Pantanowitz L, Wu U, Seigh L, LoPresti E, Yeh F-C, Salgia P, Michelow P, Hazelhurst S, Chen W-Y, Hartman D, Yeh C-Y. 2021. Artificial intelligence-based screening for *Mycobacteria* in whole-slide images of tissue samples. Am J Clin Pathol 156:117–128. https://doi.org/10.1093/ajcp/aqaa215.
11. Tomasello G, Foroughi F, Padron D, Moreno A, Banaei N. 2022. Evaluation of MetaSystems automated fluorescent microscopy system for the machine-assisted retection of acid-fast Bacilli in clinical samples. J Clin Microbiol 60:e0113122. https://doi.org/10.1128/jcm.01131-22.
12. Horvath L, Hänselmann S, Mannsperger H, Degenhardt S, Last K, Zimmermann S, Burckhardt I. 2020. Machine-assisted interpretation of auramine stains substantially increases through-put and sensitivity of microscopic tuberculosis diagnosis. Tuberculosis (Edinb) 125:101993. https://doi.org/10.1016/j.tube.2020.101993.
13. Rhoads DD, Mathison BA, Bishop HS, da Silva AJ, Pantanowitz L. 2016. Review of telemicrobiology. Arch Pathol Lab Med 140:362–370. https://doi.org/10.5858/arpa.2015-0116-RA.
14. Faron ML, Buchan BW, Coon C, Liebregts T, van Bree A, Jansz AR, Soucy G, Korver J, Ledeboer NA. 2016. Automatic digital analysis of chromogenic media for vancomycin-resistant-*Enterococcus* screens using Copan WASPLab. J Clin Microbiol 54:2464–2469. https://doi.org/10.1128/JCM.01040-16.
15. Mathison BA, Kohan JL, Walker JF, Smith RB, Ardon O, Couturier MR. 2020. Detection of intestinal protozoa in trichrome-stained stool specimens by use of a deep convolutional neural network. J Clin Microbiol 58:e02053-19. https://doi.org/10.1128/JCM.02053-19.
16. Saito T, Rehmsmeier M. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10:e0118432. https://doi.org/10.1371/journal.pone.0118432.
17. Durant TJS, Dudgeon SN, McPadden J, Simpson A, Price N, Schulz WL, Torres R, Olson EM. 2021. Applications of digital microscopy and densely connected convolutional neural networks for automated quantification of babesia-infected rrythrocytes. Clin Chem 68:218–229. https://doi.org/10.1093/clinchem/hvab237.
18. Horning MP, Delahunt CB, Bachman CM, Luchavez J, Luna C, Hu L, Jaiswal MS, Thompson CM, Kulhare S, Janko S, Wilson BK, Ostbye T, Mehanian M, Gebrehiwot R, Yun G, Bell D, Proux S, Carter JY, Oyibo W, Gamboa D, Dhorda M, Vongpromek R, Chiodini PL, Ogutu B, Long EG, Tun K, Burkot TR, Lilley K, Mehanian C. 2021. Performance of a fully-automated system on a WHO malaria microscopy evaluation slide set. Malar J 20:110. https://doi.org/10.1186/s12936-021-03631-3.
19. Nugent RP, Krohn MA, Hillier SL. 1991. Reliability of diagnosing bacterial vaginosis is improved by a standardized method of gram stain interpretation. J Clin Microbiol 29:297–301. https://doi.org/10.1128/jcm.29.2.297-301.1991.
20. Wang Z, Zhang L, Zhao M, Wang Y, Bai H, Wang Y, Rui C, Fan C, Li J, Li N, Liu X, Wang Z, Si Y, Feng A, Li M, Zhang Q, Yang Z, Wang M, Wu W, Cao Y, Qi L, Zeng X, Geng L, An R, Li P, Liu Z, Qiao Q, Zhu W, Mo W, Liao Q, Xu W. 2021. Deep neural networks offer morphologic classification and diagnosis of bacterial vaginosis. J Clin Microbiol 59:e02236-20. https://doi.org/10.1128/JCM.02236-20.
21. Bailey AL, Ledeboer N, Burnham C-AD. 2019. Clinical microbiology is growing up: the total laboratory automation revolution. Clin Chem 65:634–643. https://doi.org/10.1373/clinchem.2017.274522.

22. Baker J, Timm K, Faron M, Ledeboer N, Culbreath K. 2020. Digital image analysis for the detection of group B *Streptococcus* from ChromID Strepto B medium using PhenoMatrix algorithms. J Clin Microbiol 59:e01902-19. https://doi.org/10.1128/JCM.01902-19.
23. Van TT, Mata K, Dien Bard J. 2019. Automated detection of *Streptococcus pyogenes* pharyngitis by use of colorex strep A CHROMagar and WASPLab artificial intelligence chromogenic detection module software. J Clin Microbiol 57:e00811-19. https://doi.org/10.1128/JCM.00811-19.
24. Cherkaoui A, Renzi G, Charretier Y, Blanc DS, Vuilleumier N, Schrenzel J. 2019. Automated incubation and digital image analysis of chromogenic media using Copan WASPLab enables rapid detection of vancomycin-resistant *Enterococcus*. Front Cell Infect Microbiol 9:379. https://doi.org/10.3389/fcimb.2019.00379.
25. Faron ML, Buchan BW, Samra H, Ledeboer NA. 2019. Evaluation of WASPLab software to automatically read chromID CPS elite agar for reporting of urine cultures. J Clin Microbiol 58:e00540-19. https://doi.org/10.1128/JCM.00540-19.
26. Faron ML, Buchan BW, Vismara C, Lacchini C, Bielli A, Gesu G, Liebregts T, van Bree A, Jansz A, Soucy G, Korver J, Ledeboer NA. 2016. Automated scoring of chromogenic media for detection of methicillin-resistant *Staphylococcus aureus* by use of WASPLab image analysis software. J Clin Microbiol 54:620–624. https://doi.org/10.1128/JCM.02778-15.
27. Gammel N, Ross TL, Lewis S, Olson M, Henciak S, Harris R, Hanlon A, Carroll KC. 2021. Comparison of an automated plate assessment system (APAS independence) and artificial intelligence (AI) to manual plate reading of methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* CHROMagar surveillance cultures. J Clin Microbiol 59:e0097121. https://doi.org/10.1128/JCM.00971-21.
28. Glasson J, Hill R, Summerford M, Olden D, Papadopoulos F, Young S, Giglio S. 2017. Multicenter evaluation of an image analysis device (APAS): comparison between digital image and traditional plate reading using urine cultures. Ann Lab Med 37:499–504. https://doi.org/10.3343/alm.2017.37.6.499.
29. Gitterman. 2019. K183648. APAS independence with urine analysis module. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K183648.
30. Dauwalder O, Michel A, Eymard C, Santos K, Chanel L, Luzzati A, Roy-Azcora P, Sauzon JF, Guillaumont M, Girardo P, Fuhrmann C, Lina G, Laurent F, Vandenesch F, Sobas C. 2021. Use of artificial intelligence for tailored routine urine analyses. Clin Microbiol Infect 27:1168.e1–1168.e6. https://doi.org/10.1016/j.cmi.2020.09.056.
31. Faron ML, Buchan BW, Relich RF, Clark J, Ledeboer NA. 2020. Evaluation of the WASPLab segregation software to automatically analyze urine cultures using routine blood and MacConkey agars. J Clin Microbiol 58:e01683-19. https://doi.org/10.1128/JCM.01683-19.
32. Uwamino Y, Nagata M, Aoki W, Kato A, Daigo M, Ishihara O, Igari H, Inose R, Hasegawa N, Murata M. 2022. Efficient automated semi-quantitative urine culture analysis via BD urine culture app. Diagn Microbiol Infect Dis 102:115567. https://doi.org/10.1016/j.diagmicrobio.2021.115567.
33. Patel P, Droske LE, Patel J, Barza R, Lindgren R, McElvania E. 2019. 2152. Detection of uropathogens using BD Kiestra™ total laboratory automation with urine culture application. Open Forum Infect Dis 6:S730. https://doi.org/10.1093/ofid/ofz360.1832.
34. Alouani DJ, Ransom EM, Jani M, Burnham C-A, Rhoads DD, Sadri N. 2022. Deep convolutional neural networks implementation for the analysis of urine culture. Clin Chem 68:574–583. https://doi.org/10.1093/clinchem/hvab270.
35. Thomson GK, Jamros K, Snyder JW, Thomson KS. 2021. Digital imaging for reading of direct rapid antibiotic susceptibility tests from positive blood cultures. Eur J Clin Microbiol Infect Dis 40:2105–2112. https://doi.org/10.1007/s10096-021-04249-8.
36. Cherkaoui A, Schorderet D, Azam N, Crudeli L, Fernandez J, Renzi G, Fischer A, Schrenzel J. 2022. Fully automated EUCAST rapid antimicrobial susceptibility testing (RAST) from positive blood cultures: diagnostic accuracy and implementation. J Clin Microbiol 60:e0089822. https://doi.org/10.1128/jcm.00898-22.
37. Herroelen PH, Heestermans R, Emmerechts K, Vandoorslaer K, Wybo I, Piérard D, Muyldermans A. 2022. Validation of rapid antimicrobial susceptibility testing directly from blood cultures using WASPLab, including Colibri™ and Radian in-line carousel. Eur J Clin Microbiol Infect Dis 41:733–739. https://doi.org/10.1007/s10096-022-04421-8.
38. Cherkaoui A, Renzi G, Vuilleumier N, Schrenzel J. 2021. Performance of fully automated antimicrobial disk diffusion susceptibility testing using Copan WASP Colibri coupled to the radian in-line carousel and expert system. J Clin Microbiol 59:e0077721. https://doi.org/10.1128/JCM.00777-21.
39. Hombach M, Jetter M, Blöchliger N, Kolesnik-Goldmann N, Böttger EC. 2017. Fully automated disc diffusion for rapid antibiotic susceptibility test results: a proof-of-principle study. J Antimicrob Chemother 72:1659–1668. https://doi.org/10.1093/jac/dkx026.
40. Cherkaoui A, Renzi G, Fischer A, Azam N, Schorderet D, Vuilleumier N, Schrenzel J. 2020. Comparison of the Copan WASPLab incorporating the BioRad expert system against the SIRscan 2000 automatic for routine antimicrobial disc diffusion susceptibility testing. Clin Microbiol Infect 26:619–625. https://doi.org/10.1016/j.cmi.2019.11.008.
41. Korgenski EK, Daly JA. 1998. Evaluation of the BIOMIC video reader system for determining interpretive categories of isolates on the basis of disk diffusion susceptibility results. J Clin Microbiol 36:302–304. https://doi.org/10.1128/JCM.36.1.302-304.1998.
42. Berke I, Tierno PM. 1996. Comparison of efficacy and cost-effectiveness of BIOMIC VIDEO and Vitek antimicrobial susceptibility test systems for use in the clinical microbiology laboratory. J Clin Microbiol 34:1980–1984. https://doi.org/10.1128/jcm.34.8.1980-1984.1996.
43. Franssens BT, Fluit AC, Rentenaar RJ. 2019. Reproducibility between two readout methods of a commercial broth microdilution assay for *Pseudomonas aeruginosa* isolates from patients with Cystic Fibrosis. Infect Dis (Lond) 51:50–55. https://doi.org/10.1080/23744235.2018.1500705.
44. Nowag A, Wisplinghoff H, Quante X, Giglio S, Wirth S, Pohl B, Jazmati N. 2021. Evaluation of the use of artificial intelligence for the detection of VRE using two different agar types. https://cleverculturesystems.com/evaluation-of-ai-detection-of-vre/.
45. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. 2020. Introduction to machine learning, neural networks, and deep learning. Transl Vis Sci Technol 9:14.
46. Alouani DJ, Rajapaksha RRP, Jani M, Rhoads DD, Sadri N. 2021. Specificity of SARS-CoV-2 real-time PCR improved by deep learning analysis. J Clin Microbiol 59:e02959-20. https://doi.org/10.1128/JCM.02959-20.
47. Karah N, Antypas K, Al-Toutanji A, Suveyd U, Rafei R, Haraoui L-P, Elamin W, Hamze M, Abbara A, Rhoads DD, Pantanowitz L, Uhlin BE. 2022. Teleclinical microbiology: an innovative approach to providing web-enabled diagnostic laboratory services in Syria. Am J Clin Pathol 157:554–560. https://doi.org/10.1093/ajcp/aqab160.
48. Williams RE, Trotman RE. 1969. Automation in diagnostic bacteriology. J Clin Pathol Suppl Coll Pathol 3:8–13. https://doi.org/10.1136/jcp.s2-3.1.8.
49. Doern CD, Miller MB, Alby K, Bachman MA, Brecher SM, Casiano-Colon A, Couturier MR, Johnson JK, Kirby JE, McElvania E, Newton DW, Nolte FS, Pancholi P, McNult P, Dharmarha V, Dunbar S. 2022. Proceedings of the Clinical Microbiology Open 2018 and 2019 - a discussion about emerging trends, challenges, and the future of clinical microbiology. J Clin Microbiol 60:e00092-22. https://doi.org/10.1128/jcm.00092-22.
50. Bermejo-Peláez D, Medina N, Álamo E, Soto-Debran JC, Bonilla O, Luengo-Oroz M, Rodriguez-Tudela JL, Alastruey-Izquierdo A. 2023. Digital platform for automatic qualitative and quantitative reading of a cryptococcal antigen point-of-care assay leveraging smartphones and artificial intelligence. JoF 9:217. https://doi.org/10.3390/jof9020217.
51. Marin MJ, Van Wijk XMR, Durant TJS. 2022. Machine learning in healthcare: mapping a Path to Title 21. Clin Chem 68:609–610. https://doi.org/10.1093/clinchem/hvab285.