Journal of Biological and Sustainability Sciences
An International Peer-Reviewed Open Access Journal
Vol. 2 • No. 1 • 2026 | DOI: 10.5281/zenodo.23923742 | jbss.biosuss.com

REVIEW ARTICLE

Role of Artificial Intelligence in Clinical Microbiology: Evidence, Implementation Challenges, and Future Directions

Anisha Khan[1], Wooda Farooq[1], Bahzad Ahmad Farhan[2*], Hafiz Muhammad Sultan[3*], Shahzeen Mushier[1]

[1] Institute of Biological Sciences, Faculty of Biological, Health Sciences & Technology, Khwaja Fareed University of Engineering & Information Technology, Rahim Yar Khan, Pakistan

[2] Biochemistry and Biomolecular Engineering Research Division, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung, West Java, Indonesia

[3] Institute of Crop Sciences, Graduate School of Chinese Academy of Agricultural Sciences (GSCAAS) Beijing, China.

[*]***Corresponding authors: Bahzad Ahmad Farhan & Hafiz Muhammad Sultan;*** *farhanbehzad@gmail.com* **&** *Sultanjameel909@gmail.com*

PUBLICATION DETAILS
Received: 2026-10-15
Revised: 2026-10-28
Accepted: 2026-11-01
Published: 2026-11-03
Article ID: JBSS-2026-0013

ABSTRACT

Artificial intelligence (AI) is increasingly applied to clinical microbiology workflows, but strong technical performance does not necessarily establish clinical usefulness. This narrative review, based on literature searched through 30 September 2026, critically examines AI-assisted culture-plate analysis, microscopy, MALDI-TOF mass spectrometry, genomic and rapid phenotypic antimicrobial susceptibility prediction, molecular-result interpretation, and laboratory-linked surveillance. Evidence was evaluated according to the independence of the test data, validation design, measurement unit, clinical comparator, and implementation outcomes. Culture-image screening has been assessed in a three-laboratory study involving 104,730 specimens, whereas some promising spectral models rely on very few independent test isolates. Machine-learning-assisted nanomotion AST has demonstrated 89.5–98.9% accuracy in an independent test series, but performance varies across antibiotic–organism combinations. A published PCR-interpretation study reported agreement across 22,200 interpretations and reduced manual analysis time. Separately, total laboratory automation shortened urine-culture reporting time in a before–after study, although this improvement cannot be attributed to AI alone. Evidence for genomic resistance prediction, prescribing support, and outbreak prediction remains heterogeneous, predominantly retrospective, or insufficiently externally validated. Important barriers include dataset shift, in adequate reference standards, sample-level data leakage, limited prospective clinical-impact studies, workflow costs, equity, and regulatory oversight. The central priority is therefore not higher accuracy alone but transparent reporting, independent multicentre validation, and evaluation of patient and laboratory outcomes. AI should be introduced as a monitored addition to microbiologist-led practice, with implementation decisions matched to the strength of evidence for each specific task.

KEYWORDS

Clinical microbiology; Artificial intelligence; Antimicrobial resistance; Diagnostic accuracy; MALDI-TOF; Laboratory automation; Model validation

© 2026 The Author (s). Published in the Journal of Biological and Sustainability Sciences (JBSS), a journal of BIOSUSS Publishing (Pvt.) Limited. Authors retain copyright and grant BIOSUSS Publishing (Pvt.) Limited the non-exclusive right of first publication. This open-access article is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). **Journal information and contact: jbss.biosuss.com**

Cite this article: Khan A, Farooq W, Farhan BA, Sultan HM, Mushier S. The Role of Artificial Intelligence in Clinical Microbiology: Evidence, Implementation Challenges, and Future Directions. J Biol Sustain Sci. 2026;2(1):37-45. DOI: https://doi.org/10.5281/zenodo.23923742

1. Introduction

Conventional microbiology depends on culture, microscopy, molecular testing, mass spectrometry, and phenotypic antimicrobial susceptibility testing (AST). Turnaround time varies substantially: identification from a cultured colony by MALDI-TOF may be completed rapidly, whereas culture and routine phenotypic AST commonly require one or more days. Growing test volumes, workforce pressures, and antimicrobial resistance (AMR) increase the demand for reliable and efficient workflows. Globally, an estimated 1.27 million deaths were attributable to bacterial AMR in 2019, with 4.95 million associated deaths [1].

AI offers methods for classifying culture images and stained slides, interrogating mass spectra and genomes, estimating resistance phenotypes, reviewing molecular results, and detecting unusual laboratory patterns. However, accuracy in a development dataset, reproducibility at another laboratory, and demonstrable improvement in patient care are three different endpoints. The number of images or spectra is also not equivalent to the number of independent patients or isolates. These distinctions determine whether an AI tool is ready for routine clinical use.

1.1 Scope and contribution

This review asks: which AI applications have evidence sufficient to support implementation in the clinical microbiology laboratory, and where do validation and translation gaps remain? The core scope includes bacteriology, mycology,

parasitology, and molecular microbiology tests conducted in or directly reported by a laboratory. Antimicrobial stewardship and infection surveillance are discussed when laboratory results materially inform their outputs. Broader bedside sepsis alerts, PPE-camera systems, vaccine design, and antibiotic discovery are considered adjacent activities rather than primary diagnostic evidence. The latter has an important AI literature [2] but is outside the principal analysis here.

The contribution of the review is an application-specific assessment of evidence maturity, rather than a ranking of unrelated accuracy percentages. Existing reviews address genomic AMR [3-5], culture-image AI [6], or the breadth of clinical microbiology AI applications [7]. The present synthesis links diagnostic performance with the unit of independent evaluation, study design, operational value, and post-deployment responsibilities (Table 1). This is a qualitative interpretive framework, not a validated scoring tool or a systematic evidence grade.

Table 1. Relationship of this review to representative earlier reviews and syntheses.

Source	Primary emphasis	What this review adds
Li et al. [3]; Ardila et al. [4]	Genomic and ML approaches to AMR prediction	Places genomic results alongside phenotypic AST and specifies external-validation limitations.
Anahtar et al. [5]	Translational applications of machine learning to AMR	Separates reported discrimination from patient-level clinical impact and implementation readiness.
Burns et al. [6]	AI image analysis in clinical microbiology	Contrasts clinical image-screening evidence with spectral, molecular, and laboratory-workflow evidence.
Dien Bard et al. [7]	Conference synthesis of laboratory AI applications and adoption	Uses an explicit qualitative maturity lens and distinguishes AI-specific effects from general automation.

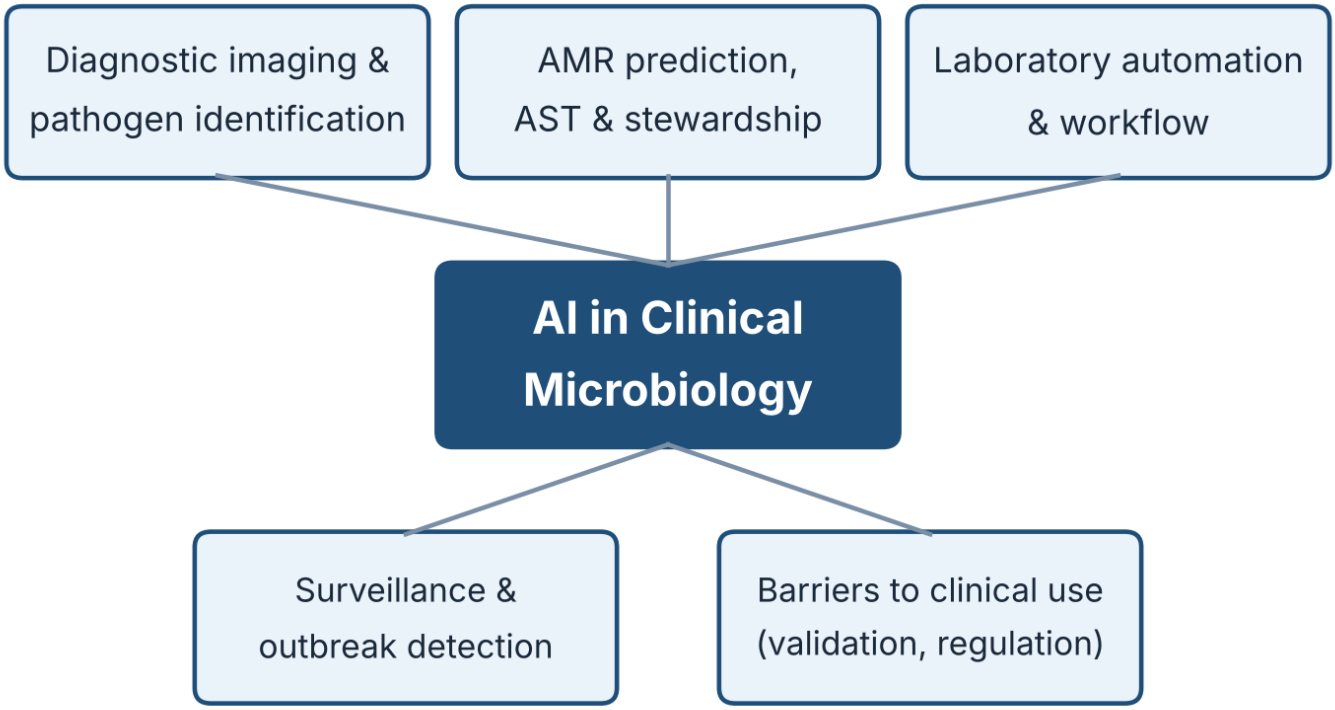


Figure 1. Laboratory-focused application areas addressed in this review. Adjacent clinical uses are discussed separately.

2. Literature Search and Critical Appraisal

This article is a narrative review rather than a systematic review or meta-analysis. The reported literature search used PubMed/MEDLINE, Scopus, and Google Scholar, initially in early 2026 and updated on 30 September 2026. Search terms combined “artificial intelligence”, “machine learning”, or “deep learning” with “clinical microbiology”, “antimicrobial resistance”, “antimicrobial susceptibility testing”, “MALDI-TOF”, “laboratory automation”, “outbreak detection”, or “infection surveillance”. Syntax was adapted for each database. The first 200 relevance-ranked Google Scholar results per query were screened according to the reported search approach, and reference lists were searched for addition al original studies. Only English-language publications were eligible; selected older foundational works were retained. Only one reviewer performed screening, which is a methodological limitation and does not describe the number of manuscript authors.

Original model-development, validation, and laboratory implementation studies were prioritized for numerical findings. Narrative and systematic reviews were used for contextual synthesis and citation discovery. Conference reports and preprints were considered with their publication status explicitly identified. Papers focusing primarily on general medical imaging, drug discovery, vaccine design, or non-laboratory AI were excluded from the core evidence synthesis. The article does not report verified aggregate record counts or an independently duplicated screening process; therefore, its selected examples must not be interpreted as an exhaustive sample of the field.

For each highlighted study, the appraisal considered the clinical task, reference method, study design, data type, number of independent sites/patients/isolates when available, internal versus external validation, performance measure, confidence interval if reported, and operational or clinical relevance. A split of many spectra from the same isolate can inflate apparent sample size if correlated observations occur in both training and testing. The qualitative stages in Table 2 are used only to structure discussion; they do not replace formal risk-of-bias assessment or a systematic certainty-of-evidence approach.

Table 2. Working evidence-maturity framework used for this narrative review (not a validated scoring instrument).

Stage	Minimum type of evidence	Principal question
Exploratory	Model development, resampling, or an internal train–test split	Can the model learn the task under the original data conditions?
Independent validation	New isolates/patients, a different instrument, or another clinical laboratory	Does performance transfer beyond correlated training data?
Prospective evaluation	Locked model assessed on newly arising cases under a real workflow	Can the system perform safely and reliably in practice?
Clinical implementation	Evaluated effects on turnaround, workload, treatment, cost, or safety	Does adoption improve meaningful outcomes compared with standard care?
Post-deployment assurance	Ongoing drift, fairness, error, and process monitoring	Does the benefit persist after implementation?

3. How AI Is Used in Clinical Microbiology

Supervised machine learning maps laboratory inputs to known outcomes, such as pathogen identity or resistant/susceptible phenotype. Convolutional neural networks are often applied to microscopy and colony images; random forests and boosting models are frequent choices for tabular genomic, spectral, and clinical features. Unsupervised approaches may identify clusters and anomalies. Quality of labels, leakage between training and testing, and the independence of evaluation data are as important as model architecture.

In practice, automated interpretation should be compared with the relevant clinical reference standard and the existing laboratory workflow. A useful tool must also indicate uncertainty, route ambiguous findings to qualified staff, integrate with the laboratory information system, and accommodate shifts in local prevalence, instruments, and resistance patterns. No single diagnostic-accuracy metric captures all these requirements.

4. AI for Diagnostic Microbiology

4.1 Digital culture-plate analysis

Automated analysis of chromogenic agar plates has one of the clearer laboratory-specific evaluation examples. Faron et al. [8] assessed vancomycin-resistant enterococcal (VRE) screening across three clinical laboratories and 104,730 specimens. Relative to the initial manual reading comparator, the automated approach achieved 100% sensitivity (95% CI 99–100) and 89.5% specificity (95% CI 89–90). Review of discordant plates identified some initially overlooked positive cultures. This indicates potential for negative-plate triage and quality assurance, rather than automatic replacement of microbiological interpretation. Findings remain specific to the test system, media and organisms evaluated.

4.2 Microscopy and clinical parasitology

AI classification has been explored for malaria blood smears and acid-fast-bacillus microscopy. In an early conference study, Dong et al. [9] reported 98.1% accuracy in identifying malaria-infected versus uninfected erythrocyte images. Malaria microscopy belongs within clinical laboratory parasitology, but an image-classification result is not equivalent to patient-level diagnostic sensitivity or performance across different microscopes and staining conditions. Investigations of tuberculosis smear analysis likewise require external laboratory testing, a clinically relevant denominator, and comparisons against established microscopy procedures [18].

Table 3. Selected primary evidence, with study design, measurement unit and limitations kept distinct.

Study/laboratory task	Design and independent unit	Validation	Reported finding	Clinical interpretation / limitation
Faron et al. [8]; VRE plates	104,730 specimens; 3 laboratories	Multisite clinical evaluation	100% sensitivity (95% CI 99–100); 89.5% specificity (95% CI 89–90) vs manual screening	Supports plate triage for specified chromogenic media; not universal pathogen detection.
Dong et al. [9]; malaria smear	Infected/non infected erythrocyte images; conference study	Research image dataset; no established independent multi-site test	98.1% classification accuracy	Clinical parasitology is in scope, but image accuracy is not specimen-level clinical validation.
Mohammad et al. [10]; MALDI-TOF	*M. abscessus*: 187 test spectra from 8 independent isolates	Held-out test spectra; few independent test isolates	84–95% balanced accuracy for species/subspecies; 91–99% for resistance prediction across benchmark tasks	Performance estimates need task-specific reading; 187 spectra are not 187 independent isolates.
Ren et al. [11]; *E. coli* WGS	Sequenced isolates; organism- and drug-specific datasets	Internal model evaluation	83.2–83.5% random forest; 85.5% CNN accuracy in reported comparisons	Predictions vary by resistance phenotype; clinical transfer not established.
Hicks et al. [12]; gonococcal WGS	Genome-linked antibiotic phenotype datasets	Internal model evaluation	>95% accuracy for ciprofloxacin; 78–88% for azithromycin	Strong drug dependence; results are not interchangeable across agents.
Sturm et al. [13]; rapid phenotypic AST	364 isolates in development; 223 strains in independent test	Independent strain testing, including clinical setting	89.5–98.9% test accuracy across antibiotic conditions	Promising for rapid AST; errors near MIC breakpoints and polymicrobial samples require study.
Kanjilal et al. [14]; UTI stewardship	Retrospective test cohort of 3,629 patients	Retrospective comparison with clinician selections	67% fewer second-line recommendations; 18% fewer inappropriate recommendations	Counterfactual recommendations; clinical treatment benefits not prospectively shown.
Khalid et al. [15]; urine workflow	16,210 manual-period and 60,474 automated-period cultures; one centre	Non-randomised before–after implementation	Negative TAT 49.68→34.38 h; positive 50.42→34.62 h	Effect of total automation, not isolated effect of AI algorithms.
MacLean & Gunson [16]; PCR.Ai	22,200 molecular-result interpretations	Prospective clinical comparison with manual analysis	100% agreement; 45 min saved per respiratory run; 32 min per norovirus run	Single-setting operational evaluation; manual comparator, not external multicentre validation.
Goncalves et al. [17]; biosurveillance	4.5 million EHR records; 123 simulated outbreak events	Retrospective, simulated prospective analysis; preprint	41/123 outbreaks (33.3%) detected 1–24 days earlier	Not a deployed hospital microbiology labor atory system or a prospective outbreak trial.

Note. VRE, vancomycin-resistant enterococci; WGS, whole-genome sequencing; MIC, minimum inhibitory concentration; TAT, turnaround time. Accuracy, balanced accuracy, sensitivity/specificity, and operational time are distinct outcomes. A multicentre evaluation does not automatically imply random assignment, blinded reader evaluation, or prospective patient benefit. Figures reproduced in this table refer to the cited task and data; they cannot be directly ranked across organisms or studies.

4.3 MALDI-TOF identification and AMR prediction

MALDI-TOF mass spectrometry already supports rapid species identification in clinical laboratories. Machine learning may extend spectral interpretation to difficult taxonomic distinctions and resistance-related signals. Mohammad et al. [10] benchmarked multiple deep-learning architectures using diverse clinical and vector-surveillance spectra. The study reported balanced accuracies of 84–95% for species/subspecies identification and 91–99% for resistance prediction across the benchmark tasks. These outcomes must not be reassigned to the unrelated mosquito identification task. Critically, the *M. abscessus* test dataset comprised 187 spectra but only eight independent test isolates (with 814 training spectra from 33 isolates). Thus, repeated spectra may support technical evaluation without establishing broad clinical transportability. Independent isolate-level validation across instruments and laboratories remains necessary.

Across image and spectral applications, the most defensible near-term role is targeted decision support: directing staff to likely positives, resistant isolates, or unusual results while retaining established confirmatory pathways. Diagnostic accuracy and operational benefit should be demonstrated separately rather than assumed to coincide.

5. AI for Antimicrobial Resistance and Susceptibility Testing

5.1 Genomic resistance prediction

Genomic prediction links sequence-derived features with resistance phenotypes. Ren et al. [11] reported random-forest accuracies of 83.2–83.5% and a CNN accuracy of 85.5% for selected *E. coli* antibiotic-resistance classification tasks. Hicks

et al. [12] found notably different performance by drug in *N. gonorrhoeae*: above 95% for ciprofloxacin and approximately 78–88% for azithromycin in reported experiments. These results demonstrate that model quality cannot be generalized from one species–antibiotic pair to another. Narrative reviews have described accuracies as high as 95% for particular feature and organism combinations [3,19], but those figures are secondary reports, not new independent validations.

Systematic reviews emphasize heterogeneity of samples, reference AST procedures, predictor features, and evaluation designs [4,20]. Ardila et al. [20] reported a mean AUROC of approximately 0.80 for gradient-boosted approaches in the reviewed health care datasets, but pooled summaries of retrospective studies do not establish clinical usefulness. Genomic prediction can be complicated by novel resistance mechanisms, non-genetic tolerance, altered epidemiology, and differences in phenotypic breakpoints [5,21]. The model must ultimately be evaluated against drug-specific susceptibility outcomes, with attention to false-susceptible predictions that could affect patient treatment.

5.2 Machine-learning-assisted rapid phenotypic AST

Rapid phenotypic AST is particularly attractive because it directly interrogates bacterial response rather than inferring phenotype from genetic markers. Sturm et al. [13] analysed bacterial nanomotion recordings from spiked positive blood cultures containing *E. coli* and *K. pneumoniae*. Development involved 364 isolates, and an independent evaluation of 223 strains produced test accuracies ranging from 89.5% to 98.9%, depending on the tested antibiotic. The reported assay could return results in hours rather than the conventional growth-based interval. Errors were more common near clinical MIC breakpoints, and polymicrobial specimens were not assessed. Accordingly, the promising independent evaluation supports further clinical study, not unqualified replacement of reference AST.

5.3 Stewardship as a laboratory-linked clinical application

Microbiology results frequently guide antimicrobial prescribing; therefore, algorithmic stewardship is a relevant downstream use rather than an intrinsic microbiology test. Kanjilal et al. [14] retrospectively assessed an algorithm for uncomplicated urinary-tract infection in 3,629 patients. Compared with the antibiotics selected by clinicians in the dataset, the algorithm would have made 67% fewer second-line selections and 18% fewer inappropriate selections. These figures describe a retrospective comparison of treatment choices, not an observed effect after live deployment. Patient allergies, contraindications, local resistance, and access to results still require clinician or pharmacist oversight. Prospective trials are needed to demonstrate whether recommendations improve therapy and safety.

6. Laboratory Automation and Molecular Result Interpretation

Total laboratory automation (TLA) may integrate robotics, automated incubation, specimen transport, imaging, and software. It should not be treated as synonymous with AI. In a single-centre comparison, Khalid et al. [15] analysed 16,210 cultures processed manually in 2023 and 60,474 after implementation of WASPLab in 2025. Median reporting time decreased from 49.68 to 34.38 hours for negative urine cultures and from 50.42 to 34.62 hours for positive cultures; reporting-time variability also declined. The before–after design cannot disentangle machine learning from robotics, workflow redesign, staffing, or secular changes. These are credible benefits of the integrated TLA system, not a measured independent causal effect of AI.

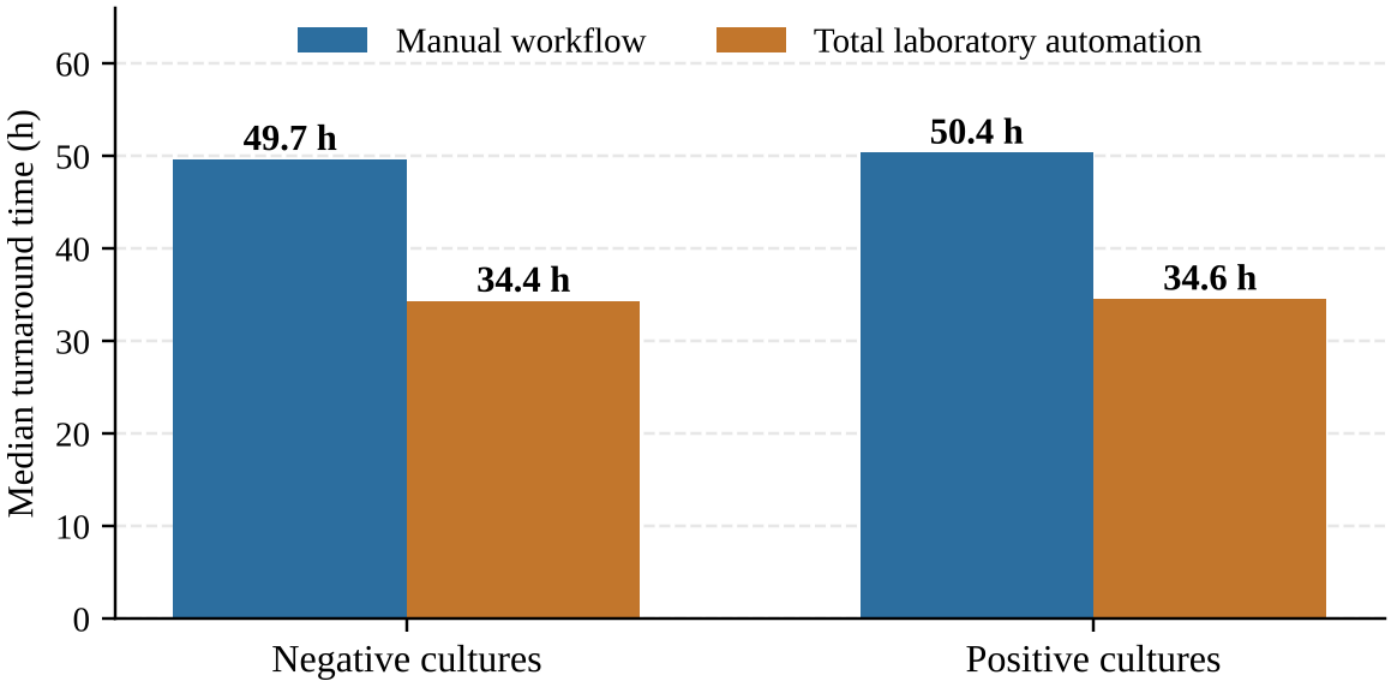


Figure 2. Median urine-culture turnaround times before and after total laboratory automation in a single-centre, non-randomised study (Khalid et al. [15]). Negative cultures: 49.68 versus 34.38 h; positive cultures: 50.42 versus 34.62 h. The comparison does not isolate the effect of AI.

AI-assisted interpretation of real-time PCR is a separate, more direct software use case. In a prospective clinical comparison, MacLean and Gunson [16] assessed PCR.Ai across 22,200 interpretations of respiratory-virus and norovirus tests and reported 100% concurrence with their manual analysis. Time savings were 45 minutes per respiratory test run and 32 minutes per norovirus run, not a uniform 40 minutes across all assays. These results support analytical workflow simplification within the tested service, but independent multicentre testing and error analysis are needed before assuming equivalent performance elsewhere. More generally, specimen triage and automated verification remain opportunities requiring rigorous evaluation [6,22]. A practical implementation should document who reviews algorithmic exceptions, which specimens must enter urgent STAT pathways, and how incorrect automated outputs are detected and corrected. Avoiding manual checks is not itself a validated reduction in analytical error.

7. Infection Surveillance and Adjacent Clinical Uses

Laboratory-linked surveillance can use microbiology results, time, location, and patient data to identify clusters that may warrant investigation. Goncalves et al. [17] evaluated an approach combining machine-learning-based diagnostic prediction with spatiotemporal outbreak analysis in approximately 4.5 million electronic health records. In a retrospective simulated-prospective assessment, 41 of 123 outbreaks (33.3%) were detected one to 24 days earlier than conventional diagnosis-based reporting. This was reported as a preprint and was not a live prospective clinical deployment. Earlier signals are only useful if false alerts, verification procedures, and the operational consequences of action are also evaluated.

General emergency-department sepsis prediction and camera-based personal protective equipment monitoring are adjacent to, rather than core examples of, AI used inside a microbiology laboratory. For example, a sepsis-alert implementation has been associated with improved outcomes in a non-randomised health care evaluation [23], but its mixed vital-sign and EHR inputs make it in appropriate to use as direct evidence that a clinical microbiology AI assay improves mortality. Keeping these domains separate prevents overstatement of laboratory-specific effectiveness.

8. Translational Barriers and Critical Appraisal

8.1 Validation, leakage, and meaningful comparators

Internal resampling, cross-validation, held-out test spectra, genuinely independent isolates, external laboratory datasets, and prospective patient-level evaluation provide different levels of evidence. For models with repeated images or spectra from the same specimen, patient- or isolate-level grouping is essential to reduce leakage between training and testing. Transportability should be challenged across instruments, operators, hospital populations, pathogen prevalence, and laboratory procedures. Calibration, sensitivity for clinically hazardous errors, subgroup performance, and workload effects may matter more than overall accuracy.

Applicable reporting resources include TRIPOD+AI for prediction-model studies, STARD-AI for diagnostic-accuracy studies, and DECIDE-AI for early clinical evaluation [24]. These checklists help identify missing data, test-set definitions, comparison standards, and transparency concerns, but using a checklist alone does not provide external validation.

8.2 Drift, bias, and implementation cost

Covariate shift occurs when input distributions change, for example after a new staining protocol or MALDI-TOF instrument. Prevalence shift changes the frequency of pathogens or resistance phenotypes, while concept drift changes the relation between inputs and target outcomes. New resistance mechanisms, revised breakpoints, different sampling practices, or altered case mix can degrade performance silently. Laboratory governance should therefore include baseline error rates, periodic external checks, trigger thresholds for investigation, a model-update policy, and clear responsibilities for staff review.

Datasets from well-resourced institutions may underrepresent uncommon organisms, patients from low-resource settings, or laboratories using other equipment. This can lead to uneven diagnostic performance and inequitable benefit [25]. Even technically successful systems may be difficult to implement because of IT integration, cybersecurity, licensing fees, instrument compatibility, quality management, and addition al staff training. A small conference-attendee survey reported 17% image-analysis use among 32 respondents [7], but the non-representative convenience sample cannot be generalized to clinical laboratories globally.

8.3 Regulation, explainability, and distinct biological endpoints

Regulatory requirements differ by jurisdiction and by whether software qualifies as a medical device. The FUTURE-AI consensus guideline emphasizes fairness, explainability, accountability, and lifecycle monitoring [26]. Explanations may help investigate error patterns, but an explanation is not proof of causal validity. Resistance based on an AST breakpoint must also be distinguished from antibiotic tolerance and persistence, where organisms survive exposure without necessarily showing an elevated MIC; these phenotypes should not be merged into one prediction target [21].

9. Future Priorities for Clinical Translation

A practical pathway begins with representative, well-labelled laboratory data and locked evaluation datasets; proceeds to independent external validation; and then tests safety, human factors, workload, cost and clinical outcomes prospectively. Post-deployment surveillance is required if the model becomes routine. Explainable AI, federated learning and multi-

omics integration may support these goals, but they are parallel technical approaches rather than compulsory sequential steps. The priority should reflect the laboratory task and available evidence, not the novelty of the algorithm.

Explainable models. SHAP [27] and LIME [28] can help visualize which features contributed to individual predictions. Their stability and usefulness for microbiologists should be tested rather than assumed.

Federated collaboration. Federated learning can train models across institutions without routinely pooling raw patient records. Privacy remains dependent on governance, secure aggregation and local safeguards; federation alone does not guarantee privacy.

Multi-omics and multimodal integration. Combining sequence, spectra, growth phenotypes and clinical context may reveal patterns missed by a single assay [21,29]. These systems require careful handling of in complete data and independent validation of each claimed application.

Prospective evaluation and monitoring. Reporting should distinguish performance by organism, drug, patient setting, reference standard, and diagnostic-error type. Meaningful impact includes time to actionable treatment, unnecessary antibiotic use, staff workload, cost, and adverse events—not solely AUROC or image-level accuracy.

10. Remaining Research Gaps and Limitations

The most consequential research gaps are: independent multi-institution test sets with patient- or isolate-level separation; prospective diagnostic-impact studies; relevant low-resource laboratory evaluations; shared data and annotation standards; comparative cost-effectiveness; and procedures for monitoring bias, error, and drift after deployment. Each application needs its own clinical comparator and decision threshold. For example, a false-negative resistance prediction may be more hazardous than a false-positive plate-screening alert, even when their reported overall accuracies appear similar.

10.1 Limitations of this review

The synthesis is narrative, selective and descriptive. A formal protocol was not registered, and one reviewer handled screening without independent duplicate assessment. Only English-language material was eligible, and database coverage was limited to PubMed/MEDLINE, Scopus and Google Scholar, supplemented by backward citation searches. No comprehensive search-flow counts or trial-registry search were available for reporting, and no formal QUADAS-2, PROBAST, or equivalent risk-of-bias assessment was performed. Some cited evidence came from conference research or preprints, and no quantitative meta-analysis was attempted. These limitations restrict reproducibility and preclude confident comparative ranking of application maturity. Therefore, conclusions should be interpreted as a transparent critical narrative rather than a systematic estimate of pooled effectiveness.

11. Conclusion

AI has credible near-term roles in selected clinical microbiology tasks, particularly targeted culture-plate screening, assisted interpretation of molecular test results, and rapid phenotypic signals for susceptibility assessment. Nevertheless, even striking percentages can be misleading when test cases are highly correlated, denominators are unclear, or models have not been evaluated on independent patients, isolates, and laboratories. Total laboratory automation can shorten reporting times, but those gains should not automatically be attributed to AI. Stewardship recommendations and outbreak forecasts have addition al prospective and clinical-impact requirements. Future adoption should therefore depend on task-specific external validation, safety and cost analysis, human oversight, and continuous monitoring. The essential question is not simply whether AI can classify laboratory data accurately, but whether it produces reliable, equitable and clinically meaningful improvements under routine working conditions.

Declarations

Author contributions: Anisha Khan and Bahzad Ahmad Farhan contributed to the conceptualization and planning of the manuscript. Anisha Khan, Wooda Farooq, Bahzad Ahmad Farhan, and Shahzeen Mushier contributed to the original manuscript drafting. Anisha Khan and Wooda Farooq prepared the figures. Bahzad Ahmad Farhan, Hafiz Muhammad Sultan and Wooda Farooq critically reviewed and revised the manuscript and contributed to its final formatting and preparation for submission. All authors read and approved the final version of the manuscript.

Funding: This work received no external funding, as reported by the authors.

Ethics approval and informed consent: Not applicable; this narrative review uses previously published literature and includes no original participants, patient-level data collection, or animal experiments.

Data availability: No new research datasets were generated. The reviewed publications are identified in the References; an exhaustive database-screening log is not included.

Acknowledgements: The authors acknowledge academic guidance and critical feedback received during manuscript preparation from corresponding authors.

Conflicts of interest: The authors declare no conflicts of interest.

Use of artificial intelligence tools: Generative AI tools (Claude and ChatGPT) assisted with language editing, literature and citation checking, figure planning, and manuscript formatting. The authors take full responsibility for the final manuscript and must verify every cited source, scientific claim, and submission requirement.

References

1. Murray CJL, Ikuta KS, Sharara *F*, Swetschinski L, Robles Aguilar G, Gray A, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629–55. doi: 10.1016/S0140-6736(21)02724-0.
2. Stokes JM, Yang K, Swanson K, Jin *W*, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688–702.e13. doi: 10.1016/j.cell.2020.01.021.
3. Li Y, Cui X, Yang X, Liu G, Zhang J. Artificial intelligence in predicting pathogenic microorganisms' antimicrobial resistance: challenges, progress, and prospects. Front Cell Infect Microbiol. 2024;14. doi: 10.3389/fcimb.2024.1482186.
4. Ardila CM, Yadalam PK, González-Arroyave D. Integrating whole genome sequencing and machine learning for predicting antimicrobial resistance in critical pathogens: a systematic review of antimicrobial susceptibility tests. PeerJ. 2024;12: e18213. doi: 10.7717/peerj.18213.
5. Anahtar MN, Yang JH, Kanjilal S. Applications of machine learning to the problem of antimicrobial resistance: an emerging model for translational research. J Clin Microbiol. 2021;59(7): e01260-20. doi: 10.1128/JCM.01260-20.
6. Burns BL, Rhoads DD, Misra A. The use of machine learning for image analysis artificial intelligence in clinical microbiology. J Clin Microbiol. 2023;61(9): e02336-21. doi: 10.1128/jcm.02336-21.
7. Dien Bard J, Prinzi AM, Larkin PMK, Peaper DR, Rhoads DD. Proceedings of the Clinical Microbiology Open 2024: artificial intelligence applications in clinical microbiology. J Clin Microbiol. 2025;63(4): e01804-24. doi: 10.1128/jcm.01804-24.
8. Faron ML, Buchan BW, Coon C, Liebregts *T*, van Bree A, Jansz A, et al. Automatic digital analysis of chromogenic media for vancomycin-resistant-enterococcus screens using Copan WASPLab. J Clin Microbiol. 2016;54(10):2464–9. doi: 10.1128/JCM.01040-16.
9. Dong Y, Jiang Z, Shen H, Pan WD, Williams LA, Reddy VVB, et al. Evaluations of deep convolutional neural networks for automatic identification of malaria infected cells. In: 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); 2017 Feb 16–19; Orlando, FL, USA. IEEE; 2017. p. 101–104. doi:10.1109/BHI.2017.7897215.
10. Mohammad *N*, Godmer A, Piarroux *R*, Nabet C, Tannier X. Benchmarking deep learning architectures for MALDI-TOF mass spectrometry in infectious disease diagnostics and vector-borne disease surveillance. npj Digit Med. 2026;9:717. doi: 10.1038/s41746-026-02816-x.
11. Ren Y, Chakraborty *T*, Doijad S, Falgenhauer L, Falgenhauer J, Goesmann A, et al. Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. Bioinformatics. 2022;38(2):325–34. doi: 10.1093/bioinformatics/btab681.
12. Hicks AL, Wheeler *N*, Sánchez-Busó L, Rakeman JL, Harris SR, Grad YH. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLoS Comput Biol. 2019;15(9): e1007349. doi: 10.1371/journal.pcbi.1007349.
13. Sturm A, Jóźwiak G, Pla Verge M, Munch L, Cathomen G, Vocat A, et al. Accurate and rapid antibiotic susceptibility testing using a machine learning-assisted nanomotion technology platform. Nat Commun. 2024;15(1):2037. doi: 10.1038/s41467-024-46213-y.
14. Kanjilal S, Oberst M, Boominathan S, Zhou *H*, Hooper DC, Sontag D. A decision algorithm to promote out patient antimicrobial stewardship for uncomplicated urinary tract infection. Sci Transl Med. 2020;12(568): eaay5067. doi: 10.1126/scitranslmed.aay5067.
15. Khalid *F*, Alzahrani AJ, Mohammed *H*, et al. Impact of total laboratory automation on urine culture turnaround time: a comparative study between manual workflow and WASPLab™. Diagnostics (Basel). 2026;16(8):1235. doi: 10.3390/diagnostics16081235.
16. MacLean AR, Gunson *R*. Automation and standardisation of clinical molecular testing using PCR.Ai: a comparative performance study. J Clin Virol. 2019;120:51–56. doi: 10.1016/j.jcv.2019.08.005.
17. Goncalves AR, Cadena Pico J, Hu Y, Schlessinger D, Greene J, O'Suilleabhain L, et al. AI-enabled diagnostic prediction within electronic health records to enhance biosurveillance and early outbreak detection [Preprint]. medRxiv. 2025 May 16. doi:10.1101/2025.05.14.25327606.
18. Kandilci M, Yakıcı G, Kayar MB. Artificial intelligence and microbiology. Exp Appl Med Sci. 2024;5(2):119–28. doi: 10.46871/eams.1458704.
19. Ali *T*, Ahmed S, Aslam M. Artificial intelligence for antimicrobial resistance prediction: challenges and opportunities towards practical implementation. Antibiotics (Basel). 2023;12(3):523. doi: 10.3390/antibiotics12030523.
20. Ardila CM, González-Arroyave D, Tobón S. Machine learning for predicting antimicrobial resistance in critical and high-priority pathogens: a systematic review considering antimicrobial susceptibility tests in real-world health care settings. PLoS One. 2025;20(2): e0319460. doi: 10.1371/journal.pone.0319460.
21. Ghosh A, Vang CK, Brenner EP, Ravi J. Unlocking antimicrobial resistance with multi omics and machine learning. Trends Microbiol. 2025;33(10):1048–51. doi: 10.1016/j.tim.2025.04.017.
22. Singhal *N*, Vardhan *H*, Jain *R*, Gupta *P*, Pandey A, Wagri NK, et al. Role of artificial intelligence in automating diagnostic procedures in clinical microbiology laboratories. Curr Res Biotechnol. 2025;10:100351. doi: 10.1016/j.crbiot.2025.100351.
23. Boussina A, Shashikumar SP, Malhotra A, Owens RL, El-Kareh *R*, Longhurst CA, et al. Impact of a deep learning sepsis prediction model on quality of care and survival. NPJ Digit Med. 2024;7:14. doi: 10.1038/s41746-023-00986-6.
24. Sounderajah *V*, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025;31:3283–3289. doi: 10.1038/s41591-025-03953-8.

25. Joseph J. Algorithmic bias in public health AI: a silent threat to equity in low-resource settings. Front Public Health. 2025;13:1643180. doi: 10.3389/fpubh.2025.1643180.

26. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in health care. BMJ. 2025;388: e081554. doi: 10.1136/bmj-2024-081554.

27. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017. p. 4765–4774.

28. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16); 2016; San Francisco, CA, USA. ACM; 2016. p. 1135–1144. doi:10.1145/2939672.2939778.

29. Jandoubi B, Akhloufi MA. Multimodal artificial intelligence in medical diagnostics. Information. 2025;16(7):591. doi: 10.3390/info16070591.