Journal of Infection 87 (2023) 287–294

Contents lists available at ScienceDirect

Journal of Infection

journal homepage: www.elsevier.com/locate/jinf

Review

Artificial intelligence, machine learning and deep learning: Potential resources for the infection clinician

Anastasia A. Theodosiou *, Robert C. Read

Clinical and Experimental Sciences and NIHR Southampton Biomedical Research Centre, University Hospital Southampton, Tremona Road, SO166YD Southampton, United Kingdom

ARTICLE INFO

Article history:
Accepted 12 July 2023
Available online 17 July 2023

Keywords:
Artificial intelligence
Machine learning
Deep learning
Clinical decision support systems

SUMMARY

Background: Artificial intelligence (AI), machine learning and deep learning (including generative AI) are increasingly being investigated in the context of research and management of human infection.
Objectives: We summarise recent and potential future applications of AI and its relevance to clinical infection practice.
Methods: 1617 PubMed results were screened, with priority given to clinical trials, systematic reviews and meta-analyses. This narrative review focusses on studies using prospectively collected real-world data with clinical validation, and on research with translational potential, such as novel drug discovery and microbiome-based interventions.
Results: There is some evidence of clinical utility of AI applied to laboratory diagnostics (e.g. digital culture plate reading, malaria diagnosis, antimicrobial resistance profiling), clinical imaging analysis (e.g. pulmonary tuberculosis diagnosis), clinical decision support tools (e.g. sepsis prediction, antimicrobial prescribing) and public health outbreak management (e.g. COVID-19). Most studies to date lack any real-world validation or clinical utility metrics. Significant heterogeneity in study design and reporting limits comparability. Many practical and ethical issues exist, including algorithm transparency and risk of bias.
Conclusions: Interest in and development of AI-based tools for infection research and management are undoubtedly gaining pace, although the real-world clinical utility to date appears much more modest.

Introduction

Artificial intelligence (AI) is the ubiquitous earworm of the current era, increasingly featuring in news, social media, and even medical literature. Indeed, there has been exponential growth in publications relating to AI, machine learning (ML) and deep learning (DL) (Fig. 1). In recent months, this Journal has published several reports of utility of (AI) and machine learning (ML) to improve prediction and diagnosis of infectious diseases. For example, Luo et al.[1] reported the use of ML to generate a predictive model that distinguished active from latent tuberculosis with a sensitivity and specificity of 88% and 91% respectively. However, the global literature to date varies significantly in scope, quality and target audience, ranging from jargon-rich specialist bioinformatics journals to much broader general-interest pieces. The conclusions of authors also vary dramatically, from near-evangelical promises that AI will improve every aspect of medicine, to doomsday prophecies of AI replacing human clinicians outright. In the face of this tsunami, it may be challenging for front-line infection clinicians (including infectious diseases physicians and medical microbiologists) to ascertain the relevance of AI to their own clinical practice. In this Editorial Commentary, we explore key AI publications and milestones over the past five years in the diagnosis and management of human infections.

Artificial intelligence (AI) refers to the use of algorithms and models enabling machines to perform tasks that normally require human intelligence. The earliest applications of AI in healthcare relied on expert rules, in which the knowledge and expertise of human specialists was used to formulate a series of "if-then" rules (e.g. "if the patient is febrile, then request blood cultures"). The first such system relating to human infection was MYCIN, introduced by Stanford University in the 1970s to suggest antibiotics for severe infections based on patient-specific input data, although this system was never widely used in clinical practice.[2] While expert rules are widely used in clinical medicine today, including most clinical decision support systems and calculators, they are often not included

* Corresponding author.
E-mail address: at1u17@soton.ac.uk (A.A. Theodosiou).

https://doi.org/10.1016/j.jinf.2023.07.006

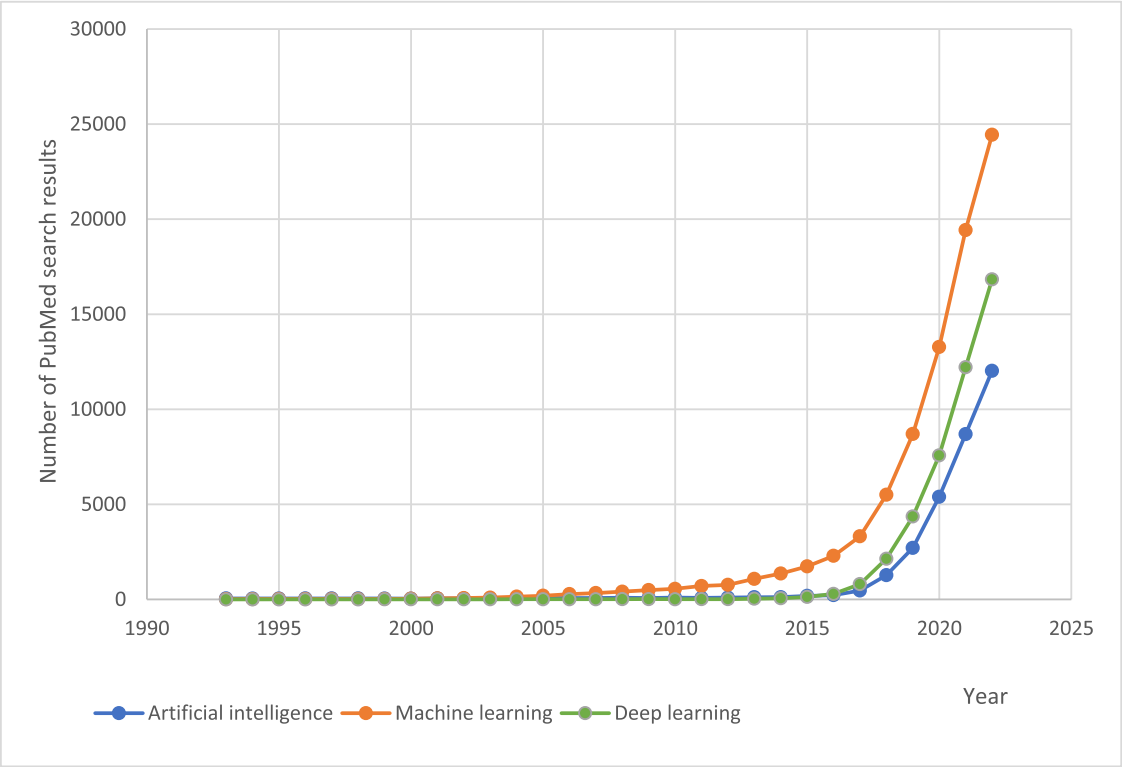

Fig. 1. Recent trends in PubMed search results. Search field: Title/Abstract; Search terms: artificial intelligence, machine learning, deep learning.

in contemporary discussions of AI, which focus primarily on machine learning and, more recently, deep learning.

Machine learning (ML) is the subset of AI involving algorithms that, unlike expert rules, can define their own rules from input data through iterative training and improvement, without explicit human programming.[3] ML was first applied in medicine in the 1980s and 1990s in the form of computer-assisted diagnosis systems in medical imaging.[4] Broadly, machine learning encompasses supervised, unsupervised and reinforcement learning (Fig. 2). Supervised learning algorithms are trained using labelled data (e.g. histological specimens that have already been labelled as normal or diseased by a human expert). When applied to unlabelled categorical or continuous test data, these trained algorithms predict outcomes by classification or regression, respectively. Conversely, unsupervised ML algorithms are trained on unlabelled data by data-driven (rather than human-guided) processes. Such models are used for data clustering, feature extraction and dimensionality reduction (e.g. identifying patient subgroups based on unlabelled clinical data). Finally, reinforcement learning is an environment-driven approach, where iterative learning cycles result in a reward or penalty by comparison with a pre-defined target (e.g. continuous blood glucose monitoring and insulin administration). Clinically-relevant AI models often employ more than one of these approaches, and each approach can be further subclassified into many different algorithm types, the details of which are beyond the scope of this review.[3]

Deep learning (DL) refers to an increasingly-popular branch of ML employing artificial neural networks (ANN) with multiple processing layers, which may employ supervised, unsupervised and reinforcement ML approaches.[5] DL excels particularly in complex tasks involving high-volume and high-dimensional data. However, it is computationally more demanding than traditional ML approaches, and, as many of its processing layers remain hidden from the human user (giving rise to the so-called "black box" of AI), it presents greater challenges for model interpretability and accountability. Large language models (LLM), including chatbots based on generative pre-trained transformer (GPT) architecture like ChatGPT and GPT-4, are DL models trained on large volumes of data to generate human-like text.[6] Such **generative AI** is also capable of producing novel non-text outputs (e.g. images, audio and video), marking it out as a potentially more disruptive technology than traditional non-generative ML. Concerningly, such technology is also being utilised in cases of scientific fraud.[7]

Predictions of the transformational role of AI in healthcare are not new, with the earliest claims dating to the 1950s.[8] However, as anticipated by Moore's law, computing speed, memory, compactness, cost, and algorithmic capabilities have improved exponentially since then, as has the availability of digital clinical data. This rapid rate of development makes accurate forecasting about the future role of AI challenging. Here we focus on recent and current applications of AI in the diagnosis and management of infection, highlighting barriers to uptake in clinical practice and ongoing concerns and limitations. It is not intended as an exhaustive or specialist review, but rather a practical guide for front-line infection clinicians making sense of AI in their own specialty.

Methods

An initial MEDLINE/PubMed search was conducted using the search terms "(((artificial intelligence[Title/Abstract]) OR (machine learning [Title/Abstract]) OR (deep learning[Title/Abstract]) OR (neural network [Title/Abstract]) OR (chatGPT[Title/Abstract])) AND ((microbiology [Title/Abstract]) OR (infectious diseases[Title/Abstract]) OR (antimicrobial[Title/Abstract]) OR (antibiotic[Title/Abstract])))", with publication date limited to the past five years (up to 31st May 2023). The titles of all 1617 results were screened, but full-text articles were only reviewed for systematic reviews, meta-analyses and clinical trials or real-world algorithm validation, rather than papers dealing solely with algorithm development. Additional potentially-relevant papers were

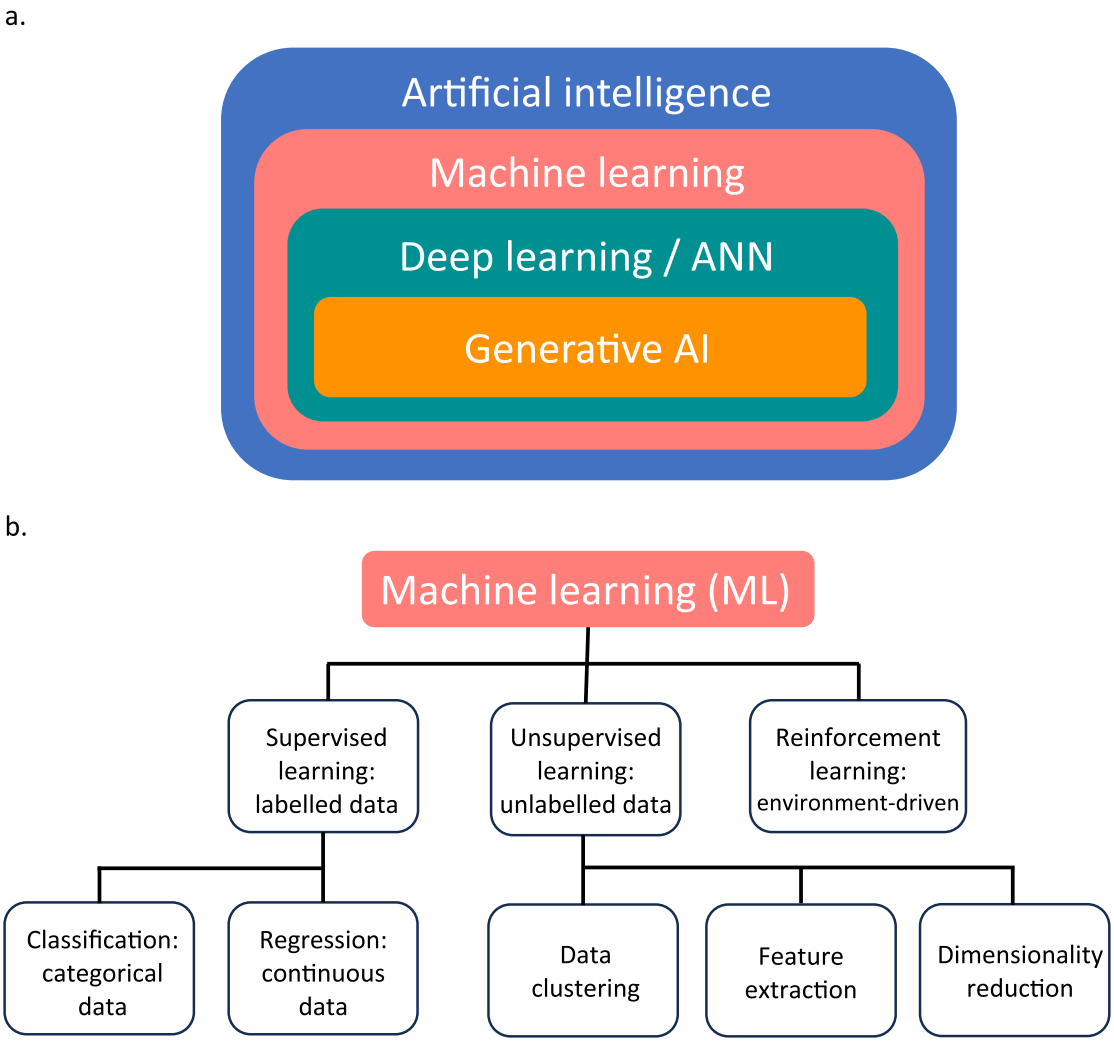

Fig. 2. Relationship between artificial intelligence (AI), machine learning (ML), deep learning/artificial neural networks (ANN), and generative AI (a); subdivisions and applications of ML (b).

identified using the reference lists of included full-text articles. As review articles published within the past five years made references to papers published earlier than five years ago, some older full-text articles were assessed where they were felt to add significant value to this review, such as randomised clinical trials employing AI algorithms (very rare overall in the literature to date).

Many papers were identified describing the development of AI algorithms for use in the diagnosis and management of human infections, as well as several reviews summarising these. However, most articles dealt purely with algorithm development rather than clinical deployment, with validation in an external dataset either lacking or applying only to retrospective data. Furthermore, algorithm performance was almost always reported using statistical metrics such as sensitivity, specificity, or area under the receiver-operating characteristic curve (AUROC) (Table 1a), with very few reporting any clinical utility metrics, such as impact on patient care or efficiency. The focus here is on studies that included prospectively collected real-world data, and any reporting clinical outcomes of AI algorithm use. Furthermore, we focus on ML, DL and generative AI, rather than manually-programmed expert rules. Where research applications of AI are explored, we focus primarily on research with translational potential, such as development of novel antimicrobials, vaccines or microbiome-based therapies.

Laboratory and imaging diagnostics

ML and DL algorithms have been investigated in the **clinical microbiology laboratory**, including microorganism detection, quantification, and antimicrobial resistance profiling.[9] In 2019, the automated plate assessment system APAS Compact (Clever Culture Systems, Switzerland) became the first AI-based clinical microbiology system to receive approval by the US Food and Drug Administration (FDA) as a Class II Medical Device.[10] Comparing digital image analysis of culture plates with traditional plate reading by microbiologists, APAS Compact achieved 90.8% sensitivity and 92.8% specificity,[11] while the more recent APAS Independence improved throughput to 200 plates per hour.[10] Clever Culture System's module for detecting methicillin-resistant *Staphylococcus aureus* (MRSA) received FDA approval in 2021, achieving 100% negative predictive value over a five-month implementation period involving digital plate reading of 5913 nasal swab cultures.[12] However, the positive predictive value was only 60.8%, and cost-effectiveness and clinical uptake have not yet been reported. A further AI-based system (using supervised ML) integrates urine microscopy results and patient clinical data to determine the need for culture, with Burton et al. reporting a classification sensitivity of over 95% and a resulting reduction in relative workload of 41% since implementation.[10]

Table 1
Reporting metrics commonly used in published evaluations of artificial intelligence algorithms. AUROC: area under receiver operating curve; DOR: diagnostic odds ratio; FN: false negatives; FP: false positives; FPR: false positive rate; MCC: Matthews correlation coefficient; PPV: positive predictive value; TN: true negatives; TNR: true negative rate; TP: true positives; TPR: true positive rate; *: multiplied by.

a. Description and calculation of algorithm reporting metrics

Metric	Description [metric range]	Equation
Sensitivity (recall, TPR)	Ratio of correctly labelled positives to all condition positives [0 to 1]	$\frac{TP}{TP+FN}$
Specificity (TNR)	Ratio of correctly labelled negatives to all condition negatives [0 to 1]	$\frac{TN}{TN+FP}$
Precision (PPV)	Ratio of correctly labelled positives to all condition positives [0 to 1]	$\frac{TP}{TP+FP}$
Accuracy	Ratio of correctly labelled datapoints to all datapoints in dataset [0 to 1]	$\frac{TP+TN}{TP+TN+FN+FP}$
F1-score	Harmonic mean of precision and sensitivity [0 to 1]	$\frac{2*sensitivity*precision}{(sensitivity\ +\ precision)}$
MCC	Considers all four binary classification variables (TP, TN, FP, FN) [-1 to 1]	$\frac{(TP*TN)-(FP*FN)}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}}$
DOR	Odds ratio of positive labelling for condition positive versus negative [0 to positive infinity]	$\frac{TP*TN}{FP*FN}$
AUROC	Area under TPR (sensitivity) plotted against FPR (1-specificity) [0 to 1]	Various approaches exist, relying on integration of ROC

b. Hypothetical dataset with calculated reporting metrics

	Condition positive = 92	Condition negative = 8
Labelled positive = 90	TP = 88	FP = 2
Labelled negative = 10	FN = 4	TN = 6

Metric	Calculation	Result
Sensitivity	88 / (88+4)	0.957
Specificity	6 / (6+2)	0.75
Precision	88 / (88+2)	0.978
Accuracy	(88+6) / (88+6+2+4)	0.94
F1-score	2*0.978*0.957/(0.978*0.957)	0.968
MCC	(88*6)-(2*4)/ √(88+2)*(88+4)*(6+2)*(6+4)	0.845
DOR	88*6/2*4	66

AI has been applied to the diagnosis of pulmonary tuberculosis using computer-assisted analysis of **clinical images** (including chest radiographs and computed tomography).[13] In a meta-analysis of 23 clinical studies pertaining to 124,959 patients, of which 12 were conducted prospectively, pooled sensitivity and specificity were 91% (95% CI, 89–93%) and 65% (54–75%), respectively, with most comparing digital imaging analysis to a human reader. However, there was significant heterogeneity in study design, sample size, validation and reporting metrics, and no cost-effectiveness analyses were reported. Computer-assisted image analysis of blood films has also been applied to the diagnosis of malaria,[14] with some models achieving accuracy and sensitivity over 99%. However, as for tuberculosis, the evidence to date varies in dataset size (a few hundred to nearly 30,000 images per study), study design, and reporting metrics, with most studies lacking independent clinical validation, and generally disappointing results with non-falciparum species. ML-assisted blood spectroscopy has also been evaluated for malaria diagnosis, with a systematic review summarising results from 58 studies, although real-world validation and prospective testing to date remain limited.[15]

An intriguing application of AI is expedited diagnosis of infections caused by **antimicrobial resistant (AMR) organisms** by alternatives to conventional culture-based techniques, such as whole genome sequencing[16,17] and matrix-assisted laser desorption/ionisation–time of flight mass spectrometry (MALDI-TOF).[18,19] Using a dataset of over 300,000 mass spectra and over 750,000 AMR phenotypes, ML combined with MALDI-TOF achieved good detection of resistant pathogens, including *S. aureus, E. coli* and *K. pneumoniae* (AUROC 0.80, 0.74 and 0.74, respectively).[18] In a retrospective case review, this model would have improved treatment in eight out of 63 patients, but may have led to inappropriate treatment in one case. ML modelling of whole genome sequencing data has yielded promising initial results for highly clonal organisms with well-defined single nucleotide variants, such as *M. tuberculosis.* However, ML model performance varies significantly with assembly quality, feature selection, resistance metrics, and drug or organism of interest, and clinical utility of this approach remains unclear.

Clinical decision support

Beyond the laboratory, AI algorithms have been developed and applied to offer clinical decision support, including prediction and stratification of sepsis, antimicrobial prescribing and stewardship advice, and prediction of colonisation with AMR organisms.[9,20]

A trio of papers was identified relating to real-world implementation of ML algorithms (using supervised gradient-boosted decision trees) by the USA-based company Dascena to facilitate **early sepsis detection.**[21–23] A prospective open-label multi-centre study of 17,758 adult patients at nine hospitals (2017–2018) reported clinical improvement for at least one month following ML algorithm use, compared with a period prior to implementation (in-hospital mortality 2.34%, length of stay (LOS) 3.27 days, 30-day readmission 28.12%, compared with 3.86%, 4.83 days and 36.4% respectively, $p < 0.001$ for all three outcomes).[23] However, no control group was included, and so it possible that the improvements were due to factors other than algorithm deployment, such as clinician engagement during the intervention period. In a smaller study using the same algorithm, 67 intensive care unit (ICU) patients were randomised to monitoring using either the ML algorithm (n = 67) or standard of care (n = 75) over a three-month period.[22] Improvements were noted in hospital LOS (10.3 days versus 13.0 days, $p = 0.042$), ICU LOS (6.31 days versus 8.40 days, $p = 0.03$) and in-hospital mortality (8.96% versus 21.3%, $p = 0.018$), with patients having blood cultures taken and receiving antibiotics approximately 2.8 h earlier than controls when sepsis was suspected. Dascena has also trialled an ML algorithm for predicting need for invasive ventilation in COVID-19 infection.[24] This prospective multi-centre trial of 197 patients reported superior sensitivity, specificity and diagnostic odds ratio compared to the Modified Early Warning Score (MEWS). However, the authors assumed a linear relationship between MEWS

score and probability of needing mechanical ventilation without offering any evidence that the MEWS score can or should be used in this way, limiting the usefulness of this comparison.

Regarding clinical support for **antimicrobial prescribing**, we identified several studies describing expert rules systems in clinical practice or ML algorithm development with no clinical validation, but only one study reporting real-world use of an ML algorithm.[25] During a five-week period, 515 piperacillin-tazobactam prescriptions were prospectively evaluated by a trained supervised ML algorithm, and also by the hospital's existing expert rules system and pharmacists. The pharmacists issued 43 recommendations to correct inappropriate prescribing, of which 38 were predicted by the expert rules system compared with only 17 by the ML algorithm (although these included all 5 that the expert rules system failed to identify). However, the authors did not comment on whether these systems detected any inappropriate prescribing missed by the pharmacists. An additional paper described the protocol for a cluster-randomised open-label cross-over controlled trial comparing standard of care with a neural network-based algorithm providing real-time feedback to antimicrobial prescribers, although the results of this trial have not yet been published.[26]

In clinical management, the facility to predict infection with AMR organisms may expedite prescribing of appropriate antimicrobial therapy. A systematic review of twenty-two algorithms predicting colonisation with AMR organisms reported that previous admission, recent antibiotic exposure, age and sex were the most useful risk factors for predicting carriage, although almost all studies relied on retrospective data or lacked clinical validation.[27] Only one study, involving an artificial neural network for MRSA prediction, had a prospective case-control design with clinical validation. The model predicted MRSA colonisation with 85.6% accuracy (91.3% sensitivity, 80.0% specificity), although no patient care outcomes were reported, and validation in an independent dataset was not performed.[28]

Infectious disease surveillance and public health

A recent systematic review of 237 articles on AI for infectious disease surveillance and **biopreparedness** summarised applications for tracking temporal incidence, disease risk factors and spatial movement of people, although the authors highlight ongoing issues with uncertainty quantification, handling missing data, and inconsistent reporting metrics.[29] However, some ML and DL-based systems have already seen real-world application in surveillance and public health, with many leveraging the ever-growing availability of person-specific data arising from social media, global positioning systems, and wearable healthcare trackers.[30]

Resources such as HealthMap employ natural language processing of online news media and expert-curated resources (such as ProMED) to provided automated global outbreak alerts, including Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV2).[31,32] Similarly, ensemble ML models based on regional and seasonal epidemiological data are used by the US Centres for Disease Control and Prevention (CDC) to forecast annual influenza patterns.[33] Combining whole genome sequencing of clinical bacterial isolates, electronic health record mining and ML has proven effective at retrospectively identifying undetected hospital outbreaks, with Sundermann et al. offering cost-effectiveness analyses compared with traditional infection prevention approaches, although prospective clinical trial data are lacking.[34]

More recently, the COVID-19 pandemic has seen ML applied to rapid and scalable genomic classification and lineage mapping, assisting in both outbreak epidemiology and probing **potential zoonotic origins** of the virus.[35] The reinforcement ML system Eva was implemented at all Greek borders to reduce SARS-CoV2 transmission[36]; ML-targeted testing based on demographic data and test results of previous travellers led to detection of 1.85 times more asymptomatic infected travellers than random surveillance, and up to four times more during peak travel periods. Impressively, the ZOE COVID Study used a smartphone app to collect potential COVID-19 symptom data from over 2.6 million participants, used to predict probable infection, and providing valuable real-world public health information on disease trajectory and vaccine efficacy.[37]

AI in infection research

Computer-assisted **drug discovery** is nothing new. For example, zanamivir and oseltamivir were designed using structure-based strategies, in which virtual screening (so-called "docking") was combined with traditional X-ray crystallography.[38] Conventional drug development is usually characterised by very high rates of costly pre-clinical testing attrition, which may be partly overcome through drug discovery using ML (particularly DL) models.[39] Such approaches have recently been applied to prediction of drug targets, molecular structure, pharmacokinetics and toxicity, including de novo in silico design and virtual screening of large drug databases for putative novel antimicrobials.[40,41] More recently, the novel antimicrobial halicin was identified by a deep neural network screening over 6000 compounds in the Drug Repurposing Hub database, along with a further eight antibacterial compounds by screening over 107 million molecules in the ZINC15 database.[42] Much pre-clinical research is currently in progress investigating the antimicrobial properties of AI-discovered non-ribosomal peptides, bacteriocins, and marine products, although clinical trials using AI-designed drugs are still lacking. One of the most significant milestones in the field of computer-assisted drug design is the ability to accurately predict three-dimensional protein structure, including the neural network-based model AlphaFold.[43] Such advances may also expedite the development of novel vaccines, and ML-based resources such as VaxiJen and Vaxign-ML have already been deployed in reverse vaccinology over the past decade to assist in antigen prediction.[44]

The **human microbiome** presents a further research area with significant translational potential and scope to benefit from AI advances. The microbiome is the overall community of microbes inhabiting a host, and is increasingly recognised as playing a crucial role in human health and disease.[45] However, inherent features of microbiome datasets can make such research computationally demanding, including their high-dimensional, compositional, heterogeneous, and sparse nature.[46] ML methods are well-suited to navigating these challenges, and have been applied to microbiome feature selection and regression, prediction of host phenotypes from microbiome data, and identifying environmental or clinical risk factors from microbiome signatures. Such applications may facilitate microbiome-based therapies and personalised medical or lifestyle interventions. For example, an ML algorithm has been shown to reliably predict faecal transplant donors associated with greater success of microbiome engraftment in the recipient.[47] Moreover, the ZOE Predict study employed AI-based analysis of continuous blood glucose monitoring, faecal metagenomics and detailed dietary, lifestyle and biometric measures to study the relationship between diet, gut microbiome and metabolism in over 1000 participants. ZOE now offer personalised dietary advice commercially to its customers, based on the user's own faecal microbiome, glucose monitoring and biometric data.[48] As the focus of microbiome research moves progressively from purely descriptive analyses to more systems-based and clinically-focussed questions, AI is predicted to play an increasingly important role in expediting microbiome-based therapeutics.[49] However, gold standards for conducting microbiome research are lacking, and the significant heterogeneity between studies and between human participants mean that promise of truly personalised medicine is still some years from being realised.

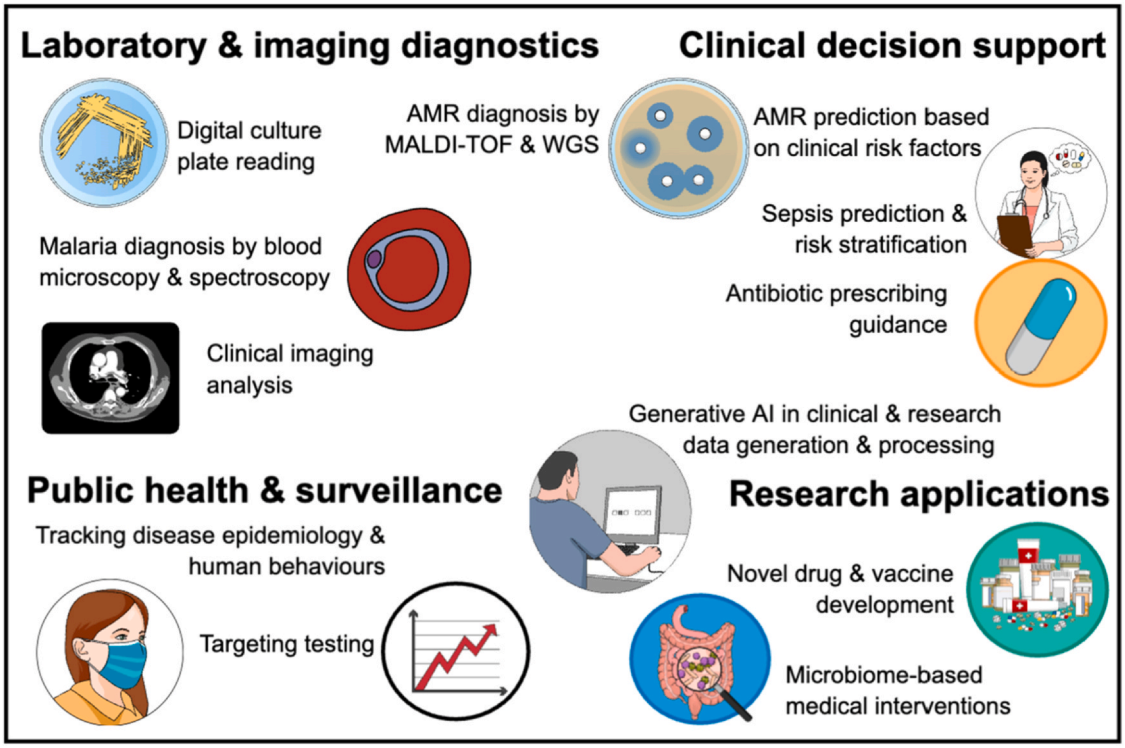

Fig. 3. Examples of artificial intelligence (AI) applications in research and management of human infections. AMR: antimicrobial resistance; MALDI-TOF: matrix-assisted laser desorption/ionisation–time of flight mass spectrometry; WGS: whole genome sequencing.

Generative AI in infection research and clinical practice

Generative AI has dominated news, social media and popular culture in recent months, since the chatbot ChatGPT became the fastest-growing consumer application in history, with over 100 million monthly users within two months of its release. There have been a flurry journal articles on the role of generative AI in medicine, including editorials, examples of academic manuscripts or discharge summaries written entirely by ChatGPT, and reports of ChatGPT passing medical exams and suggesting management of hypothetical clinical cases. A recent systematic review of 60 such papers highlighted the potential for ChatGPT and other generative AI resources to assist in medical research (e.g. scientific writing, literature review, and data analysis) and clinical practice (e.g. medical record keeping and administration, generating diagnoses and management plans, improving public health messaging and patient health literacy).[50] On the negative side, commonly cited concerns included risk of bias, plagiarism, data privacy, transparency, inaccurate information, misinformation, copyright and legal issues. The World Association of Medical Editors have issued recommendations for the use of generative AI in academic writing, stating that chatbots cannot be listed as authors, editors need appropriate tools to detect AI-generated content, and authors should be transparent about chatbot involvement and remain ultimately responsible for their work.[51]

Discussion

It is clear that interest in AI has extended to almost all aspects of human infection management and research (Fig. 3), although concrete evidence of clinical utility remains much more modest. While accurate forecasting remains challenging, several concerns, limitations and opportunities have been widely highlighted, and awareness of these could help front-line clinicians keep abreast of AI developments in their own practice.

At present, AI-based research lacks consensus standards for study design, data sources and handling, reporting metrics, and model validation, leading to significant heterogeneity in study quality and limiting reproducibility and meta-analysis (Table 1b). Furthermore, although future clinical utility is often alluded to, very few studies address how to integrate AI into existing clinical and laboratory workflow. It is unclear to what extent models developed using a particular dataset can be extrapolated to other populations; of note, the majority of clinical support models developed to date are based on secondary care (especially ICU) rather than primary care data, and on data from high-income rather than low- and middle-income countries (LMIC).[9] A notable exception to the latter is smartphone-based applications designed for use in resource-poor settings, such as an antibiogram image analysis point-of-care tool spearheaded and used in the field by the charity Medecins Sans Frontieres.[52]

For ML algorithms to achieve mainstream clinical utility, issues of accountability must first be addressed. As ML (especially DL) models include several hidden processing layers, it may be impossible for clinicians (and even the programming engineer) to determine how any given model has arrived at a particular output.[53] This so-called "black box" is especially pertinent when considering **AI hallucinations**, the phenomenon in which a generative AI tool confidently asserts a factual inaccuracy.[50] Commentators have also warned that over-reliance on AI in clinical practice may lead to physician de-skilling, and that AI models may be less able to navigate uncertainty, nuance and individual patient contexts than a human clinician.[53] Such issues highlight the importance of the user's own expertise when using AI models, as a non-specialist practitioner or member of the public may be more susceptible to these pitfalls than a specialist aiming to supplement (rather than replace) their clinical expertise.

Several **ethical, social and legal issues** relating to AI use in medicine have been identified and systematically reviewed, including data governance, privacy, and equitable access.[54] These are particularly relevant when considering disease surveillance models based on data from the general public, rather than consenting

patients. There is also concern that incorporation of AI-based tools into clinical practice could magnify existing inequalities and biases. Several studies and systematic reviews have identified implicit bias amongst healthcare professionals,[55] based on race, ethnicity, gender, age, and weight, and, concerningly, no widely used interventions have been shown to cause a meaningful and lasting reduction in such biases.[56] Social and health inequalities also disproportionately impact particular groups, and training datasets may not be representative of all target populations, leading to both data-driven and algorithm bias.[57] Thus, ML algorithms trained on labelled or unlabelled data may arrive at biased conclusions, and even amplify underlying inequalities. For example, an AI algorithm using health costs as a proxy for healthcare needs falsely interpreted lower per-capita health spending on black patients as an indication that this group was healthier than corresponding white patients, resulting in inappropriate racial prioritisation of white patients in the algorithm's outputs.

There have been calls to address these issues, to facilitate constructive uptake of AI into healthcare. The US FDA, UK Medicines and Healthcare products Regulatory Agency (MHRA) and Health Canada have jointly identified ten guiding principles to inform the development of Good Machine Learning Practice, promoting safe, effective and high-quality medical devices that use AI and ML. These emphasise the need for multi-disciplinary expertise, robust cybersecurity and risk management, representative population data, independent training and validation data sets, and focus on the role of human factors and interpretability ("human in the loop"). Other commentators have advocated for open science practices, including code and data sharing, participant-centred algorithm development, and more clinical trial data.[57]

Conclusions

Artificial intelligence has undoubtedly generated a range of intriguing tools with potential application to research and management of human infections. However, at the time of writing, the real-world utility of almost all such AI applications remains unclear, and several practical and ethical issues must be addressed before widespread uptake of AI in infection medicine can be expected. Provided further clinical trial data, regulation and reporting standards are forthcoming, AI is well-placed to improve accuracy and efficiency in laboratory diagnostics, clinical support, public health and infection research. It does not, however, appear poised to replace front-line infection clinicians or academics any time soon.

Declaration of Competing Interest

Both authors conceived of and planned this Editorial Commentary. AT screened the literature and prepared the first draft of the manuscript, both authors discussed and edited the manuscript. Artificial intelligence was not used in the preparation of this manuscript. The authors have no competing interests to declare.

Acknowledgements

Figures are original and were produced using Mindthegraph.com and Microsoft Office. AT is funded by a Medical Research Council Clinical Research Training Fellowship (MR/V002015/1).

References

1. Luo Y, Xue Y, Song H, Tang G, Liu W, Bai H, et al. *Machine learning based on routine laboratory indicators promoting the discrimination between active tuberculosis and latent tuberculosis infection. J Infect* 2022;**84**(5):648–57. https://doi.org/10.1016/j.jinf.2021.12.046
2. Shortliffe E. *Computer-based medical consultations*. MYCIN. Elsevier; 1976. ⟨https://shop.elsevier.com/books/computer-based-medical-consultations-mycin/shortliffe/978-0-444-00179-5⟩ [Accessed 10 July 2023].
3. Jiang Y, Luo J, Huang D, Liu Y, Li D dan. *Machine learning advances in microbiology: a review of methods and applications. Front Microbiol* 2022;**13**:925454. https://doi.org/10.3389/fmicb.2022.925454
4. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. *Artificial intelligence in radiology. Nat Rev Cancer* 2018;**18**(8):500–10. https://doi.org/10.1038/s41568-018-0016-5
5. Sidey-Gibbons JAM, Sidey-Gibbons CJ. *Machine learning in medicine: a practical introduction. BMC Med Res Method* 2019;**19**(1):64. https://doi.org/10.1186/s12874-019-0681-4
6. Birhane A, Kasirzadeh A, Leslie D, Wachter S. *Science in the age of large language models. Nat Rev Phys* 2023;**5**(5):277–80. https://doi.org/10.1038/s42254-023-00581-4
7. Liverpool L. *AI intensifies fight against 'paper mills' that churn out fake research. Nature* 2023;**618**(7964):222–3. https://doi.org/10.1038/d41586-023-01780-w
8. Haug CJ, Drazen JM. *Artificial intelligence and machine learning in clinical medicine.* Drazen JM, Kohane IS, Leong TY (Eds.) N Engl J Med; 388(13); 2023: p. 1201–8. DOI: ⟨10.1056/NEJMra2302038⟩.
9. Peiffer-Smadja N, Dellière S, Rodriguez C, Birgand G, Lescure FX, Fourati S, et al. *Machine learning in the clinical microbiology laboratory: has the time come for routine practice? Clin Microbiol Infect Publ Eur Soc Clin Microbiol Infect Dis* 2020;**26**(10):1300–9. https://doi.org/10.1016/j.cmi.2020.02.006
10. Brenton L, Waters MJ, Stanford T, Giglio S. *Clinical evaluation of the APAS® Independence: automated imaging and interpretation of urine cultures using artificial intelligence with composite reference standard discrepant resolution. J Microbiol Methods* 2020;**177**:106047. https://doi.org/10.1016/j.mimet.2020.106047
11. Glasson J, Hill R, Summerford M, Olden D, Papadopoulos F, Young S, et al. *Multicenter evaluation of an image analysis device (APAS): comparison between digital image and traditional plate reading using urine cultures. Ann Lab Med* 2017;**37**(6):499–504. https://doi.org/10.3343/alm.2017.37.6.499
12. Gammel N, Ross TL, Lewis S, Olson M, Henciak S, Harris R, et al. *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* 2021;**59**(11):e0097121. https://doi.org/10.1128/JCM.00971-21
13. Zhan Y, Wang Y, Zhang W, Ying B, Wang C. *Diagnostic accuracy of the artificial intelligence methods in medical imaging for pulmonary tuberculosis: a systematic review and meta-analysis. J Clin Med* 2022;**12**(1):303. https://doi.org/10.3390/jcm12010303
14. Aqeel S, Haider Z, Khan W. *Towards digital diagnosis of malaria: how far have we reached. J Microbiol Methods* 2023;**204**:106630. https://doi.org/10.1016/j.mimet.2022.106630
15. Goh B, Ching K, Soares Magalhães RJ, et al. *The application of spectroscopy techniques for diagnosis of malaria parasites and arboviruses and surveillance of mosquito vectors: a systematic review and critical appraisal of evidence. PLoS Negl Trop Dis* 2021;**15**(4):e0009218. https://doi.org/10.1371/journal.pntd.0009218
16. 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. https://doi.org/10.1371/journal.pcbi.1007349
17. Kim JI, Maguire F, Tsang KK, Gouliouris T, Peacock SJ, McAllister TA, et al. *Machine learning for antimicrobial resistance prediction: current practice, limitations, and clinical perspective. Clin Microbiol Rev* 2022;**35**(3):e0017921. https://doi.org/10.1128/cmr.00179-21
18. Weis C, Cuénod A, Rieck B, Dubuis O, Graf S, Lang C, et al. *Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. Nat Med* 2022;**28**(1):164–74. https://doi.org/10.1038/s41591-021-01619-9
19. Weis CV, Jutzeler CR, Borgwardt K. *Machine learning for microbial identification and antimicrobial susceptibility testing on MALDI-TOF mass spectra: a systematic review. Clin Microbiol Infect* 2020;**26**(10):1310–7. https://doi.org/10.1016/j.cmi.2020.03.014
20. Luz CF, Vollmer M, Decruyenaere J, Nijsten MW, Glasner C, Sinha B. *Machine learning in infection management using routine electronic health records: tools, techniques, and reporting of future technologies. Clin Microbiol Infect Publ Eur Soc Clin Microbiol Infect Dis* 2020;**26**(10):1291–9. https://doi.org/10.1016/j.cmi.2020.02.003
21. McCoy A, Das R. *Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. BMJ Open Qual* 2017;**6**(2):e000158. https://doi.org/10.1136/bmjoq-2017-000158
22. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. *Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Respir Res* 2017;**4**(1):e000234. https://doi.org/10.1136/bmjresp-2017-000234
23. Burdick H, Pino E, Gabel-Comeau D, McCoy A, Gu C, Roberts J, et al. *Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. BMJ Health Care Inf* 2020;**27**(1):e100109. https://doi.org/10.1136/bmjhci-2019-100109
24. Burdick H, Lam C, Mataraso S, Siefkas A, Braden G, Dellinger RP, et al. *Prediction of respiratory decompensation in Covid-19 patients using machine learning: the READY trial. Comput Biol Med* 2020;**124**:103949. https://doi.org/10.1016/j.compbiomed.2020.103949

25. Beaudoin M, Kabanza F, Nault V, Valiquette L. *Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. Artif Intell Med* 2016;**68**:29–36. https://doi.org/10.1016/j.artmed.2016.02.001
26. Chang Y, Yao Y, Cui Z, Yang G, Li D, Wang L, et al. *Changing antibiotic prescribing practices in outpatient primary care settings in China: study protocol for a health information system-based cluster-randomised crossover controlled trial. PLoS One* 2022;**17**(1):e0259065. https://doi.org/10.1371/journal.pone.0259065
27. Jeon D, Chavda S, Rennert-May E, Leal J. *Clinical prediction tools for identifying antimicrobial-resistant organism (ARO) carriage on hospital admissions: a systematic review. J Hosp Infect* 2023;**134**:11–26. https://doi.org/10.1016/j.jhin.2023.01.003
28. Hsu CC, Lin YE, Chen YS, Liu YC, Muder RR. *Validation study of artificial neural network models for prediction of methicillin-resistant Staphylococcus aureus carriage. Infect Control Hosp Epidemiol* 2008;**29**(7):607–14. https://doi.org/10.1086/588588
29. Keshavamurthy R, Dixon S, Pazdernik KT, Charles LE. *Predicting infectious disease for biopreparedness and response: a systematic review of machine learning and deep learning approaches. One Health* 2022;**15**:100439. https://doi.org/10.1016/j.onehlt.2022.100439
30. Brownstein JS, Rader B, Astley CM, Tian H. *Advances in artificial intelligence for infectious-disease surveillance*. Drazen JM, Kohane IS, Leong TY (Eds.) N Engl J Med; 388(17); 2023: p. 1597–607. DOI: ⟨10.1056/NEJMra2119215⟩.
31. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. *HealthMap: global infectious disease monitoring through automated classification and visualization of internet media reports. J Am Med Inf Assoc* 2008;**15**(2):150–7. https://doi.org/10.1197/jamia.M2544
32. Hossain N, Househ M. *Using healthmap to analyse middle east respiratory syndrome (MERS) data. Stud Health Technol Inf* 2016;**226**:213–6.
33. Reich NG, McGowan CJ, Yamana TK, Tushar A, Ray EL, Osthus D, et al. *Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the U.S. PLOS Comput Biol* 2019;**15**(11):e1007486. https://doi.org/10.1371/journal.pcbi.1007486
34. Sundermann AJ, Chen J, Kumar P, Ayres AM, Cho ST, Ezeonwuka C, et al. *Whole-genome sequencing surveillance and machine learning of the electronic health record for enhanced healthcare outbreak detection. Clin Infect Dis Publ Infect Dis Soc Am* 2022;**75**(3):476–82. https://doi.org/10.1093/cid/ciab946
35. Randhawa GS, Soltysiak MPM, El Roz H, de Souza CPE, Hill KA, Kari L. *Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. PLoS One* 2020;**15**(4):e0232391. https://doi.org/10.1371/journal.pone.0232391
36. Bastani H, Drakopoulos K, Gupta V, Vlachogiannis I, Hadjichristodoulou C, Lagiou P, et al. *Efficient and targeted COVID-19 border testing via reinforcement learning. Nature* 2021;**599**(7883):108–13. https://doi.org/10.1038/s41586-021-04014-z
37. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. *Real-time tracking of self-reported symptoms to predict potential COVID-19. Nat Med* 2020;**26**(7):1037–40. https://doi.org/10.1038/s41591-020-0916-2
38. Mallipeddi PL, Kumar G, White SW, Webb TR. *Recent advances in computer-aided drug design as applied to anti-influenza drug discovery. Curr Top Med Chem* 2014;**14**(16):1875–89. https://doi.org/10.2174/1568026614666140929153812
39. Sarkar C, Das B, Rawat VS, Wahlang JB, Nongpiur A, Tiewsoh I, et al. *Artificial intelligence and machine learning technology driven modern drug discovery and development. Int J Mol Sci* 2023;**24**(3):2026. https://doi.org/10.3390/ijms24032026
40. Koutroumpa NM, Papavasileiou KD, Papadiamantis AG, Melagraki G, Afantitis A. *A systematic review of deep learning methodologies used in the drug discovery process with emphasis on in vivo validation. Int J Mol Sci* 2023;**24**(7):6573. https://doi.org/10.3390/ijms24076573
41. Talat A, Khan AU. *Artificial intelligence as a smart approach to develop antimicrobial drug molecules: a paradigm to combat drug-resistant infections. Drug Discov Today* 2023;**28**(4):103491. https://doi.org/10.1016/j.drudis.2023.103491
42. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. *A deep learning approach to antibiotic discovery*. e13 *Cell* 2020;**180**(4):688–702. https://doi.org/10.1016/j.cell.2020.01.021
43. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. *Highly accurate protein structure prediction with AlphaFold. Nature* 2021;**596**(7873):583–9. https://doi.org/10.1038/s41586-021-03819-2
44. Ong E, Wang H, Wong MU, Seetharaman M, Valdez N, He Y. *Vaxign-ML: supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. Bioinformatics* 2020;**36**(10):3185–91. https://doi.org/10.1093/bioinformatics/btaa119
45. de Vos WM, Tilg H, Van Hul M, Cani PD. *Gut microbiome and health: mechanistic insights. Gut* 2022;**71**(5):1020–32. https://doi.org/10.1136/gutjnl-2021-326789
46. Marcos-Zambrano LJ, Karaduzovic-Hadziabdic K, Loncar Turukalo T, Przymus P, Trajkovik V, Aasmets O, et al. *Applications of machine learning in human microbiome studies: a review on feature selection, biomarker identification, disease prediction and treatment. Front Microbiol* 2021;**12** Accessed July 10, 2023 ⟨https://www.frontiersin.org/articles/10.3389/fmicb.2021.634511⟩.
47. Ianiro G, Punčochář M, Karcher N, Porcari S, Armanini F, Asnicar F, et al. *Variability of strain engraftment and predictability of microbiome composition after fecal microbiota transplantation across different diseases. Nat Med* 2022;**28**(9):1913–23. https://doi.org/10.1038/s41591-022-01964-3
48. Asnicar F, Berry SE, Valdes AM, Nguyen LH, Piccinno G, Drew DA, et al. *Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals. Nat Med* 2021;**27**(2):321–32. https://doi.org/10.1038/s41591-020-01183-8
49. Li P, Luo H, Ji B, Nielsen J. *Machine learning for data integration in human gut microbiome. Microb Cell Fact* 2022;**21**(1):241. https://doi.org/10.1186/s12934-022-01973-4
50. Sallam M. *ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare* 2023;**11**(6):887. https://doi.org/10.3390/healthcare11060887
51. Zielinski C, Winker M, Aggarwal R, Ferris L, Habibzadeh F. Chatbots, generative AI, and scholarly manuscripts. WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. WAME; 2023https://wame.org/page3.php?id=106.
52. Pascucci M, Royer G, Adamek J, Asmar MA, Aristizabal D, Blanche L, et al. *AI-based mobile application to fight antibiotic resistance. Nat Commun* 2021;**12**(1):1173. https://doi.org/10.1038/s41467-021-21187-3
53. Cabitza F, Rasoini R, Gensini GF. *Unintended consequences of machine learning in medicine. JAMA* 2017;**318**(6):517–8. https://doi.org/10.1001/jama.2017.7797
54. Zhao IY, Ma YX, Yu MWC, Liu J, Dong WN, Pang Q, et al. *Ethics, integrity, and retributions of digital detection surveillance systems for infectious diseases: systematic literature review. J Med Internet Res* 2021;**23**(10):e32328. https://doi.org/10.2196/32328
55. FitzGerald C, Hurst S. *Implicit bias in healthcare professionals: a systematic review. BMC Med Ethics* 2017;**18**(1):19. https://doi.org/10.1186/s12910-017-0179-8
56. Sabin JA. *Tackling implicit bias in health care. N Engl J Med* 2022;**387**(2):105–7. https://doi.org/10.1056/NEJMp2201180
57. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. *Addressing bias in big data and AI for health care: a call for open science. Patterns* 2021;**2**(10):100347. https://doi.org/10.1016/j.patter.2021.100347