A Narrative Review

Role of Artificial Intelligence in Revolutionizing the Diagnosis of Fatty Liver Disease

Muhammad Ahmad[1], Shanawar Hayat[2]

1 University of Lahore, Lahore, Pakistan

2 Avicenna Medical College, Lahore, Pakistan

*** Correspondence: shanawarhayat1@gmail.com**

ABSTRACT

Background: Steatotic liver disease (SLD), including metabolic dysfunction-associated steatotic liver disease (MASLD) and non-alcoholic steatohepatitis (NASH), affects nearly 30% of the global population and poses a major public health challenge. Diagnosis still depends heavily on liver biopsy, which is invasive, prone to sampling error, and unsuitable for widespread screening. These limitations have driven growing interest in artificial intelligence (AI) as a non-invasive diagnostic solution. **Objective**: To evaluate the current state and performance of AI technologies used for diagnosing, staging, and predicting progression in steatotic liver disease across imaging, electronic health record (EHR) analysis, and digital pathology. **Methods**: A comprehensive review was conducted, examining recent AI applications in CT, MRI, and ultrasound imaging, EHR-based machine learning models, and digital pathology platforms. Reported diagnostic accuracy, predictive performance metrics, technical strengths, and clinical limitations were analyzed. **Results**: Deep learning models applied to CT imaging demonstrated high accuracy in staging fibrosis, with AUC values of 0.97 for advanced fibrosis (≥F3) and 0.95 for cirrhosis (F4). AI-assisted ultrasound achieved an AUC of 0.98 for NAFLD detection. EHR-based tools, such as NASHmap, showed moderate predictive ability (AUC 0.76). Progression-prediction models reached AUROC values of 0.87 for forecasting fibrosis within four years. In digital pathology, AI systems like qFibrosis® provided superior reproducibility (89–93%) and identified treatment effects missed by conventional histology. **Conclusion**: AI offers accurate, scalable, and objective alternatives to invasive diagnostics in liver disease, particularly for ruling out advanced fibrosis and predicting progression. However, challenges persist, including algorithmic bias, limited generalizability, opacity of deep learning models, regulatory constraints, and slow clinical translation. Future advancement requires multimodal data integration, robust external validation, improved transparency, and clear governance frameworks.

Keywords: Steatotic liver disease, MASLD, NASH, artificial intelligence, deep learning, liver fibrosis, non-invasive diagnostics, digital pathology, EHR models, risk prediction.

Received: 05 August 2025
Revised: 18 October 2025
Accepted: 24 October 2025
Published: 31 October 2025

Citation: Click to Cite

INTRODUCTION

Fatty liver disease, |known as NAFLD, MASLD, or simply steatotic liver disease, has quietly become the most common chronic liver condition in the world. The numbers are staggering: nearly one in three people globally now have this condition, representing a prevalence of about 30% (1). Even more concerning, this figure has jumped by over 50% in just three decades, between 1990 and 2019.

What makes this disease particularly tricky is that it exists on a spectrum. Some people have simple fat accumulation in the liver (what doctors call hepatic steatosis), while others develop a more aggressive inflammatory form known as NASH (non-alcoholic steatohepatitis). This distinction matters enormously because NASH can lead to serious complications: scarring of the liver (fibrosis), cirrhosis, liver cancer, and ultimately liver failure (2).

The progression statistics paint a sobering picture. When someone develops NASH, there's a 30–40% chance they'll go on to develop some degree of liver scarring, and about 15–20% will eventually develop cirrhosis (3). These aren't just numbers, they represent millions of people worldwide who need effective screening and early intervention.

WHY CURRENT DIAGNOSTIC METHODS FALL SHORT

Here's the paradox: despite affecting nearly a third of the global population, the only way to definitively diagnose NASH and accurately stage fibrosis remains the liver biopsy, an invasive procedure that involves inserting a needle into the liver to extract tissue (4). It's simply not feasible to biopsy hundreds of millions of people. Beyond the practical impossibility, biopsies carry risks, cause discomfort, and are expensive.

But there's another problem that doesn't receive enough attention: biopsies themselves aren't as reliable as we'd like to think. When different pathologists examine the same biopsy sample, they often disagree on the scores, a phenomenon called inter-observer variability (5). The liver is a large organ, and a biopsy needle only captures a tiny fragment, which may not represent the overall disease state (sampling error). For clinical trials testing new drugs, these limitations become critical roadblocks. How can we measure subtle improvements in liver health if our measurement tool is inherently inconsistent?

ENTER ARTIFICIAL INTELLIGENCE: A GAME-CHANGING APPROACH

This is where artificial intelligence steps in with genuine promise. AI, particularly through machine learning and deep learning techniques, can analyze massive amounts of data, medical images, electronic health records, lab results, and even molecular markers, to identify patterns invisible to the human eye (6). Early studies suggest that AI systems can match or even exceed human performance in detecting fatty liver disease, diagnosing NASH, and staging fibrosis (7).

The appeal is obvious: it's noninvasive, scalable, objective, and potentially more accurate than traditional methods. But as we'll explore, the journey from promising research to routine clinical use is filled with both technical and societal challenges.

AI'S ROLE IN MEDICAL IMAGING: SEEING WHAT HUMANS MISS

The workhorse of AI in medical imaging is a technology called Convolutional Neural Networks (CNNs), essentially, computer programs that learn to recognize patterns in images the way a child learns to distinguish cats from dogs (8). These networks are trained on thousands of scans to find the signs of fat buildup and scarring in people with liver disease.

Currently, MRI-PDFF (Magnetic Resonance Imaging–Proton Density Fat Fraction) is considered the best non-invasive way to measure liver fat in research settings (9). AI models are now being developed to work alongside or even enhance this technique. One study indicated that a deep learning model could predict liver fat content with reasonable accuracy, achieving a correlation coefficient (R^2) of 0.63 with actual MRI-PDFF

measurements (10), not perfect, but a solid start. Where AI really shines is in detecting advanced liver scarring using CT scans. One deep learning system had an AUC (Area Under the Curve, a measure of diagnostic accuracy) of 0.97 for finding advanced fibrosis and 0.95 for cirrhosis (11). To put that in perspective, an AUC of 0.97 means the system correctly distinguishes between people with and without advanced fibrosis 97% of the time, better than many traditional blood tests.

AI performs better at identifying severe disease (advanced fibrosis stages F3 and F4) than at distinguishing subtle differences in fat levels (12). This makes intuitive sense: severe scarring creates distinctive architectural changes in the liver, bridging fibrosis, nodules, and distorted blood vessels, that are easier for AI to recognize than the quantitative gradations of fat content.

MAKING ULTRASOUND SMARTER AND MORE ACCESSIBLE

Ultrasound remains the most widely available imaging tool worldwide, but it has limitations. Traditional ultrasound struggles with mild fatty liver and depends heavily on the operator's skill. AI is changing both of these equations. Studies show that AI-assisted ultrasound can achieve an impressive AUC of 0.98 for detecting fatty liver disease (13). More granularly, deep learning algorithms have demonstrated AUCs of 0.85, 0.91, and 0.93 for detecting mild, moderate, and severe steatosis, respectively (14).

Perhaps most significantly, AI integration into elastography devices (which measure liver stiffness as a proxy for fibrosis) has already received regulatory approval. The FDA cleared the Velacur™ liver diagnostic tool, which uses AI to automatically identify liver tissue and assess the quality of shear wave measurements (15). This means that less experienced operators can now perform high-quality liver assessments, democratizing access to advanced diagnostics and making large-scale screening feasible.

WHERE AI EXCELS (AND WHERE IT STRUGGLES)

A closer look at performance data reveals an important pattern. AI systems are exceptionally good at ruling out advanced disease, their strength lies in identifying who doesn't need a biopsy rather than providing fine-grained staging of early disease (16). For example, while AI achieves an AUC of 0.85 for basic steatosis detection (fat vs. no fat), the performance drops to 0.67 when trying to distinguish moderate from severe fat accumulation (17). This suggests the current clinical sweet spot for AI imaging is as a triage tool: quickly identifying high-risk patients who need further evaluation while confidently reassuring others that invasive procedures aren't necessary.

MINING ELECTRONIC HEALTH RECORDS: AI AS A CLINICAL DETECTIVE

Beyond imaging, machine learning can extract diagnostic insights from the data already sitting in electronic health records (EHRs), demographics, vital signs, lab results, medication lists, and clinical notes.

A landmark example is NASHmap, the first machine learning model specifically designed to predict NASH using biopsy-confirmed cases for training (18). Built using an advanced algorithm called XGBoost (Extreme Gradient Boosting), NASHmap analyzes 14 clinical features to predict the likelihood of NASH. In real-world validation using data from Optum's massive EHR database, it achieved an AUC of 0.76, good, though not perfect. Importantly, a simplified version using just 5 features maintained strong performance (AUC of 0.74), making it practical for resource-limited settings (18). This kind of tool

doesn't replace biopsies for definitive diagnosis, but it helps doctors prioritize, who should be referred to a specialist? Who needs more aggressive monitoring? In a world where healthcare resources are perpetually stretched thin, this kind of intelligent triage is invaluable.

Perhaps even more valuable than diagnosing current disease is predicting who will progress. One study used XGBoost to analyze patients with simple fatty liver and predict who would develop NASH or significant fibrosis within four years (19). The results were impressive: an AUROC of 0.79 for progression to NASH and an even higher 0.87 for progression to fibrosis (19).

Think about what the result means clinically. If we can identify high-risk patients years before they develop irreversible damage, we can intervene with lifestyle modifications, experimental therapies, or closer monitoring. This is the promise of precision medicine: treating the right patient, at the right time, with the right intervention.

The power of this approach was demonstrated dramatically in a Veterans Affairs study that analyzed 4.2 million patient records (20). Using machine learning, researchers identified over 500,000 veterans (12% of the at-risk population) who likely had undiagnosed NASH. The top predictive factors? Age, obesity, and abnormal liver function tests, common, easily obtainable variables that, when combined intelligently by AI, become powerful diagnostic tools (20). Even when biopsies are performed, interpreting them remains problematic. The NASH-CRN scoring system, which pathologists use to grade steatosis, inflammation, ballooning, and fibrosis, is inherently subjective (21). Different pathologists looking at the same slide often assign different scores, a major problem when you're trying to measure whether an experimental drug is working in a clinical trial.

The issue runs deeper: the categorical nature of these scores (F0, F1, F2, F3, F4) means that small improvements might be missed entirely, or a patient teetering between stages could be scored differently by different observers (22). Pharmaceutical companies, investing hundreds of millions in drug development, find this inconsistency to be a significant challenge.

AI-POWERED QUANTITATIVE HISTOLOGY

Digital pathology powered by AI is addressing these limitations by providing objective, quantitative measurements rather than subjective categorical scores. A standout example is qFibrosis® by HistoIndex (23). This system uses second harmonic generation microscopy, a specialized technique that makes collagen fibers visible without chemical staining, to measure fibrosis on a continuous numerical scale. The reproducibility is exceptional: 89% agreement when the same sample is measured twice by the same system and 86% agreement between different systems (24). When pathologists use qFibrosis as a reference, their agreement jumps to 93%, nearly perfect concordance (25).

Critically, qFibrosis has detected treatment effects in clinical trials that conventional scoring missed (26). This isn't just an academic curiosity, it directly impacts drug development. Other tools like AIM-NASH and FibroNest™ offer similar quantitative approaches, with FibroNest even providing sub-staging within traditionally monolithic categories like F1 and F4 (27,28).

ACCELERATING DRUG DEVELOPMENT

The FDA has signaled that improvement in liver fibrosis can serve as a surrogate endpoint for accelerated approval of NASH drugs (29). AI-powered digital pathology tools make

measuring these improvements both more sensitive and more reliable, potentially shortening the timeline for getting effective therapies to patients.

By standardizing how we measure disease, these tools also ensure consistency in determining who's eligible for clinical trials and whether treatments are working, fundamental requirements for regulatory approval (29).

THE LONG ROAD FROM LAB TO CLINIC

Traditional medical device regulation was designed for static technologies, a stethoscope is a stethoscope, unchanged from the day it's manufactured. But AI algorithms are different: they can learn and evolve, constantly improving as they encounter new data (30). This situation creates a regulatory conundrum. The FDA and European regulators are developing new frameworks for what they call Software as a Medical Device (SaMD), which emphasize continuous monitoring, documentation of changes, and, critically, explainability (31). Regulators want to understand why an AI makes a particular diagnosis, not just that it's statistically accurate. This need is both reasonable and challenging, given the "black box" nature of many deep learning systems.

Fortunately, there is positive news to report. Regulatory agencies are actively embracing advanced quantitative tools. The FDA's focus on objective biomarkers and surrogate endpoints in NASH trials aligns well with what AI digital pathology offers (29).

THE VALLEY OF DEATH: WHY EFFECTIVE AI TAKES YEARS TO REACH PATIENTS

It takes a long time to move AI models from research to clinical practice, even when they work well in research settings. The U-Net architecture, a foundational deep learning approach, was published in 2015 but didn't see significant adoption in research on fatty liver disease until 2019–2020 (32,33). That's a four-to-five-year lag.

Why so long? The bottleneck isn't usually the technology itself, it's the system. Healthcare workflows need to be redesigned. Doctors need to be trained. Data needs to be standardized. Hospitals need to invest in infrastructure. Insurance companies need to decide on reimbursement. None of this happens quickly (34).

Data standardization deserves special mention. Deep learning is data-hungry, requiring enormous, high-quality datasets (35). A single hospital rarely has enough diverse cases to train a robust model. But sharing data between institutions raises privacy concerns, requires legal agreements, and faces technical hurdles because of incompatible data formats. Until we solve these foundational issues, many promising AI tools will remain research curiosities rather than clinical realities (35).

ETHICAL MINEFIELDS: BIAS, TRANSPARENCY, AND TRUST

Here's an uncomfortable truth: AI systems inherit the biases present in their training data. If an algorithm is trained predominantly on data from Western, urban, well-resourced hospitals, it may perform poorly on patients from different ethnic backgrounds, rural settings, or resource-constrained environments (36). Liver disease manifestation varies by ethnicity and socioeconomic status. An AI that underperforms in African or Asian populations doesn't just fail scientifically, it exacerbates existing health inequities (37). Imagine deploying a screening tool that misses disease in already underserved populations. The harm would compound over time.

The solution requires mandatory bias audits, diverse training datasets, and validation across different populations before deployment (38). This isn't just ethically right, it's practically necessary for tools meant to serve global populations.

Deep learning models, especially complex CNNs, often function as "black boxes", producing accurate predictions without explaining their reasoning (39). A doctor might be told, "This patient has a 78% probability of NASH," but not why the AI reached that conclusion. This opacity creates multiple problems. Clinically, doctors are hesitant to trust recommendations they can't verify or explain to patients. Legally, who's liable when an unexplainable algorithm makes a mistake? Ethically, patients have a right to understand the basis of their medical decisions (40).

Explainable AI (XAI) techniques like saliency mapping and Grad-CAM are helping by highlighting which parts of an image or which data features most influenced the AI's decision (41). Imagine an AI showing you: "I diagnosed fibrosis because of these septal patterns here and this nodular architecture here." That makes the technology transparent and teachable.

Importantly, patients are often unaware that AI is being used in their care. Ethical practice demands disclosure, consent forms and patient information should state when algorithms contribute to diagnosis or treatment decisions (42).

Healthcare data is a uniquely sensitive national resource. Misuse can violate privacy, erode public trust, and create institutional vulnerabilities (43). Policy frameworks must balance patient data rights with the societal benefits of AI-driven medical advances, a delicate equilibrium.

Equally important is establishing clear accountability. When an AI makes an error, who's responsible? Is it the algorithm developer? Is it the hospital that implemented the algorithm? Who was the doctor who relied on the algorithm? Clear chains of responsibility, meticulous documentation of system performance, and transparent version control are essential governance mechanisms (34). Without them, we risk creating systems where nobody feels accountable for failures.

THE FUTURE: TOWARD TRULY PERSONALIZED LIVER CARE

The next frontier is multimodal data fusion, seamlessly integrating clinical data, advanced imaging, and molecular information (44). Imagine an AI that considers not just your CT scan and lab results, but also genetic markers, immune cell profiles, and metabolic signatures. This holistic approach is the foundation of precision medicine.

AI is already being applied to quantify immune cell infiltration in liver biopsies, automatically identifying and counting macrophages and T cells to understand the inflammatory environment (45). Elsewhere, machine learning analyzes liquid biopsies (blood samples containing circulating DNA and cellular fragments) to detect early signs of liver cancer (46). These technologies offer comprehensive risk assessment that was previously impossible.

The highest value of AI may lie not in diagnosing current disease but in forecasting future trajectories. An AUROC of 0.87 for predicting progression to fibrosis four years in advance is genuinely remarkable (19). It means we can identify high-risk individuals while their livers are still relatively healthy and intervene aggressively.

In drug development, AI can accelerate every stage, from identifying therapeutic targets to predicting which compounds will work to optimizing clinical trial design (47). AI makes it

possible to choose the best treatment for each person by looking at genes and molecular pathways that are specific to the disease (47).

To realize its potential, AI research must transition from retrospective, single-center studies to prospective, multicenter validations across varied populations (48). We need institutional platforms that standardize data collection and sharing, making it easier to train and validate AI systems across different healthcare settings (48).

The goal isn't just publishable research, it's deployable tools that improve patient outcomes in the real world. That requires collaboration between AI developers, clinicians, regulators, and patients themselves to ensure the technology meets actual needs rather than solving hypothetical problems.

CONCLUSION: A PROMISING BUT CHALLENGING PATH FORWARD

Artificial intelligence is genuinely transforming how we diagnose and manage fatty liver disease. Deep learning applied to CT imaging achieves near-perfect accuracy (AUC ≥0.95) in detecting advanced fibrosis, making it an excellent tool for ruling out severe disease (11). Machine learning algorithms analyzing EHR data, like NASHmap (AUC 0.76), enable population-level screening (18), while predictive models (AUROC 0.87 for fibrosis progression) allow early intervention (19). AI-powered digital pathology tools like qFibrosis are solving the reproducibility crisis in clinical trials, potentially accelerating drug development (23–26).

Yet challenges remain formidable. Overcoming algorithmic bias requires diverse training data and rigorous validation across global populations (36–38). Addressing the "black box" problem demands greater transparency and explanation (39–41). Navigating complex regulatory landscapes for continuously learning systems requires new frameworks that balance innovation with safety (29,31). And translating research successes into routine clinical practice requires addressing systemic barriers around data standardization, workflow integration, and institutional acceptance (30,34,35).

The trajectory is clear: we're moving toward a future of precision hepatology where intelligent systems guide personalized, patient-centered care. But realizing this vision requires not just better algorithms but also thoughtful policy, ethical governance, and a commitment to equity in how these powerful tools are developed and deployed.

DECLARATIONS

Ethical Approval
This study was approved by the Institutional Review Board of Avicenna Medical College
Informed Consent
NA
Conflict of Interest
The authors declare no conflict of interest.
Funding
This research received no external funding.
Authors' Contributions
Concept: MA; Design: SH; Data Extraction and Management: MA; Drafting: SH.
Data Availability
The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.
Acknowledgments
Not applicable.
Study Registration
Not applicable.

REFERENCES

1. Younossi ZM, et al. (2023). The global epidemiology of NAFLD and NASH in 2023: A systematic review and meta-analysis. Hepatology, 77(4), 1335–1347.
2. Rinella ME, et al. (2023). A multisociety Delphi consensus statement on new fatty liver disease nomenclature. Hepatology, 78(6), 1966–1986.
3. Powell EE, et al. (2021). Non-alcoholic fatty liver disease. The Lancet, 397(10290), 2212–2224.
4. Chalasani N, et al. (2018). The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. Hepatology, 67(1), 328–357.
5. Bedossa P, et al. (2021). Histological assessment of NAFLD: Current standards and future perspectives. Journal of Hepatology, 74(5), 1219–1229.
6. Sato K, et al. (2024). Revolutionizing MASLD: How artificial intelligence is shaping the future of liver care. Diagnostics, 14(3), 256.
7. Kim BH, et al. (2023). Artificial intelligence in NAFLD: Current status and future directions. Clinical and Molecular Hepatology, 29(2), 235–254.
8. LeCun Y, Bengio Y, Hinton G. (2015). Deep learning. Nature, 521(7553), 436–444.
9. Caussy C, et al. (2020). MRI proton density fat fraction for assessment of hepatic steatosis. Hepatology, 71(4), 1371–1383.
10. Park HJ, et al. (2020). Deep learning regression model for fat fraction estimation using clinical MRI sequences. Journal of Magnetic Resonance Imaging, 51(5), 1387–1396.
11. Wu CC, et al. (2021). Deep learning system for staging liver fibrosis on computed tomography. Radiology, 299(2), 338–347.
12. Byra M, et al. (2022). Deep learning models for grading hepatic steatosis: Comparing performance across disease stages. Radiology, 303(2), 445–454.
13. Byra M, et al. (2020). Artificial intelligence-based ultrasound imaging for detection and quantification of hepatic steatosis. IEEE Transactions on Medical Imaging, 39(8), 2422–2431.
14. Han A, et al. (2020). Deep learning for hepatic steatosis quantification in ultrasound imaging. European Radiology, 30(5), 2889–2898.
15. Sonic Incytes. (2023). FDA 510(k) clearance of AI-powered features for Velacur liver ultrasound. Company Press Release.
16. Zhang YN, et al. (2022). Artificial intelligence for liver disease diagnosis: Strengths and limitations. Frontiers in Medicine, 9, 850656.
17. Cao W, et al. (2021). Deep learning performance in staging steatosis severity: A comparative analysis. Medical Image Analysis, 69, 101965.
18. Loomba R, et al. (2021). Noninvasive diagnosis of nonalcoholic steatohepatitis using machine learning: Development of the NASHmap model. JMIR Medical Informatics, 9(5), e26565.

19. Ioannou GN, et al. (2020). Machine learning to predict progression of non-alcoholic fatty liver to steatohepatitis or fibrosis. Alimentary Pharmacology & Therapeutics, 51(10), 1038–1049.

20. Konerman MA, et al. (2021). Artificial intelligence in identifying patients with undiagnosed nonalcoholic steatohepatitis. Journal of Health Economics and Outcomes Research, 8(1), 70–78.

21. Kleiner DE, et al. (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology, 41(6), 1313–1321.

22. Ratziu V, et al. (2011). Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology, 128(7), 1898–1906.

23. Goodman ZD, et al. (2021). Second harmonic generation microscopy with artificial intelligence for quantitative fibrosis assessment in NASH. Hepatology Communications, 5(7), 1209–1221.

24. Sun Y, et al. (2021). Repeatability and reproducibility of artificial intelligence-based digital pathology for liver fibrosis evaluation. Journal of Pathology Informatics, 12, 28.

25. Tsochatzis EA, et al. (2022). qFibrosis-assisted reads improve interpathologist agreement in NASH clinical trials. Hepatology Communications, 6(9), 2341–2351.

26. Loomba R, et al. (2023). qFibrosis detects treatment effects missed by conventional NASH-CRN scoring in clinical trials. Hepatology, 77(3), 712–724.

27. PathAI. (2022). AIM-NASH: AI-powered digital pathology for reproducible NASH assessment. PathAI Technical Report.

28. PharmaNest. (2023). FibroNest: Sub-staging fibrosis with AI-based quantitative pathology. PharmaNest White Paper.

29. FDA. (2023). Use of biomarkers for diagnosing and assessing treatment response in noncirrhotic NASH trials: Guidance for industry. U.S. Food and Drug Administration.

30. Topol EJ. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

31. FDA. (2021). Artificial intelligence and machine learning in software as a medical device. U.S. Food and Drug Administration.

32. Ronneberger O, Fischer P, Brox T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In: MICCAI. 234–241.

33. Zhou T, et al. (2021). The evolution of artificial intelligence technology in non-alcoholic fatty liver disease research. Journal of Clinical Medicine, 10(16), 3528.

34. Char DS, et al. (2020). Implementing machine learning in healthcare: Addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983.

35. He J, et al. (2020). The practical implementation of artificial intelligence technologies in medicine. Nature Medicine, 25(1), 30–36.

36. Obermeyer Z, et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

37. Rajkomar A, et al. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866–872.

38. Gianfrancesco MA, et al. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA Internal Medicine, 178(11), 1544–1547.

39. Rudin C. (2019). Stop explaining black box machine learning models and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.

40. Vayena E, et al. (2018). Machine learning in medicine: Addressing ethical challenges. PLOS Medicine, 15(11), e1002689.

41. Selvaraju RR, et al. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: ICCV. 618–626.