Digestive and Liver Disease 58 (2026) 747–758

Contents lists available at ScienceDirect

Digestive and Liver Disease

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

Position Paper

Artificial Intelligence in hepatology: A position paper by the Italian Association for the Study of the Liver

Clara Balsano[a,*], Anna Alisi[b], Patrizia Burra[c], Vincenza Calvaruso[d], Calogero Cammà[e], Andrea Campagner[f,g], Piergiorgio Donatelli[h], Giacomo Germani[i], Alessio Gerussi[j,k], Mauro Giuffrè[e], Ana Lleo[l,m], Valeria Panebianco[n], Marcello Persico[o], Nicola Pugliese[l,m], Silvia Rossi[p], Artificial Intelligence Committee[1]

[a] Division of Geriatrics, School of Emergency-Urgency Medicine, Department of Life, Health and Environmental Sciences-MESVA, University of L'Aquila, Piazzale Salvatore Tommasi 1, Coppito, L'Aquila 67100, Italy
[b] Research Unit of Genetics of Complex Phenotypes, "Bambino Gesù" Children's Hospital, IRCCS, Rome, Italy
[c] Gastroenterology, Department of Surgery, Oncology, and Gastroenterology, University of Padua, Padua, Italy
[d] Department of Health Promotion, Mother and Child Care, Internal Medicine and Medical Specialties (PROMISE), Section of Gastroenterology and Hepatology, University of Palermo, 90100 Palermo, Italy
[e] Gastroenterology and Hepatology Unit, Department of Health Promotion, Mother & Child Care, Internal Medicine & Medical Specialties, University of Palermo, Italy
[f] Department of Computer Science, Systems and Communication, University of Milano-Bicocca, Milan, Italy
[g] IRCCS Ospedale Galeazzi Sant'Ambrogio, Milan, Italy
[h] Department of Philosophy, Sapienza University of Rome, 00185 Rome, Italy
[i] Multivisceral Transplant Unit, Department of Surgery, Oncology and Gastroenterology Azienda Ospedale-Università Padova, 35125 Padua, Italy
[j] Division of Gastroenterology, Center for Autoimmune Liver Diseases, European Reference Network on Hepatological Diseases (ERN RARE-LIVER), Fondazione IRCCS San Gerardo dei Tintori, Monza, Italy
[k] Department of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy
[l] Department of Biomedical Sciences, Humanitas University, Pieve Emanuele (MI), Italy
[m] Division of Internal Medicine and Hepatology, Department of Gastroenterology, IRCCS Humanitas Research Hospital, Rozzano (MI), Italy
[n] Department of Radiological Sciences, Oncology and Pathology, Sapienza University of Rome, Rome, Italy
[o] Department of Medicine, Surgery and Dentistry, "Scuola Medica Salernitana", University of Salerno, 84081 Baronissi, SA, Italy
[p] Department of Electrical Engineering and Information Technology, University Federico II-Naples, Naples, Italy

ARTICLE INFO

Article history:
Received 16 December 2025
Accepted 4 March 2026
Available online 23 March 2026

Keywords:
Artificial Intelligence
Ethics
Hepatology
Liver

ABSTRACT

Artificial Intelligence (AI) is transforming medicine, providing unprecedented opportunities to enhance diagnosis, prognosis, and treatment across all areas of hepatology. This Position Paper by the Italian Association for the Study of the Liver (AISF) offers a comprehensive overview of the current and emerging roles of AI in liver diseases, highlighting both its promise and its limitations. The document critically examines methodological, ethical, and educational challenges associated with AI integration in clinical and research settings, and reviews key applications in MASLD/MASH, alcohol-related liver disease, autoimmune and cholestatic liver diseases, viral hepatitis, drug-induced liver injury, hepatocellular carcinoma, cholangiocarcinoma, and liver transplantation.

Particular emphasis is placed on data quality, algorithmic fairness, explainability, and reproducibility, as well as on the necessity of robust external validation and adherence to international reporting standards (TRIPOD-AI, CONSORT-AI, DECIDE-AI, CHAMAI and others). The AISF advocates a responsible and evidence-based adoption of AI through multidisciplinary collaboration, professional education, and patient engagement. This Position Paper outlines a national roadmap to align AI innovation with ethics,

* Corresponding author at: Department of Life, Health and Environmental Sciences MESVA, University of L'Aquila, Piazzale Salvatore Tommasi 1, Coppito, L'Aquila, 67100, Italy.
E-mail address: clara.balsano@univaq.it (C. Balsano).
[1] Artificial Intelligence Committee - Italian Association for the Study of the Liver (AISF)

https://doi.org/10.1016/j.dld.2026.03.005

transparency, and equity, ensuring that AI becomes a genuine ally of hepatology and patient-centered care.

1. General introduction

Artificial Intelligence (AI) has emerged as one of the most transformative forces in modern hepatology, providing new tools for disease prediction, diagnosis, prognosis, and personalized management [1,2].

In hepatology, AI plays a dual role. First, it functions as a discovery engine, identifying biomarkers, predicting fibrosis stages, and integrating radiological and histological features for earlier diagnosis. Second, it can operate as a clinical decision-support system, enhancing accuracy, reducing variability, and assisting in treatment planning [3–6].

The World Health Organization (WHO) emphasizes that responsible AI must remain human-centered, ensuring that ethical, legal, and societal dimensions are integrated into technological development [1]. Within this context, AISF (Associazione Italiana per lo Studio del Fegato) encourages the adoption of AI that respects patient rights and medical ethics. The EU AI Act (2025) classifies medical AI as "high-risk," mandating strict requirements, including transparency, traceability, human oversight, and ongoing post-market surveillance.

Nevertheless, while AI offers remarkable promise, its adoption in hepatology reveals challenges. Bias, transparency, and trustworthiness are central, as biased models can amplify pre-existing inequalities or propagate spurious associations [5,7,8]. In addition, the introduction of Generative AI (GenAI) further expands possibilities, enabling text and image synthesis, large-scale knowledge extraction, and personalized patient interaction. Yet, its clinical reliability depends on future rigorous validation and domain-specific regulation [6].

Given the scope, complexity, and rapid evolution of AI in medicine, and in hepatology in particular, a position paper in this field cannot be conceived as a narrowly prescriptive or highly operational document. Moreover, the spectrum of AI applications is highly heterogeneous. In this context, the AISF Position Paper has been designed with a dual and complementary purpose. In particular, AISF briefly articulates a set of institutional statements that define the principles, priorities, and conditions under which AI should be developed, validated, and integrated into hepatological practice.

Moreover, we provides a comprehensive and critical overview of the current state of clinical and translational research on AI in hepatology, highlighting both its achievements and its limitations.

2. Algorithmic bias, fairness, and trust

The success of AI solutions in healthcare is often measured by accuracy on a hold-out set drawn from the same population used for model development (technically called internal validation). However, accuracy alone is often insufficient to demonstrate the practical clinical usefulness of AI systems [9]. This is not solely due to the limitations of accuracy as a performance measure but more generally to the limits of so-called discrimination metrics (including area-under-the-curve (AUC), sensitivity, positive predictive value), which focus only on agreement with a given ground truth while disregarding other relevant dimensions [10].

Historically, performance evaluation of medical AI has been dominated by accuracy and AUC metrics. However, these indicators are insufficient for clinical decision-making, where safety, reliability, and fairness are equally critical. Proper assessment should consider five interrelated dimensions: calibration, robustness and transportability, interpretability and explainability, fairness, and trust calibration. In this section, we will briefly examine such dimensions, particularly relevant to the healthcare domain.

Calibration: clinicians and healthcare personnel often need risk estimates in order to support medical decision-making [11]. To this aim, an AI system should not only provide a categorical support (e.g., state that a certain condition is present or not) but also offer an assessment of the probability of that same support (e.g., state that certain condition is present with a certain probability). An AI system is calibrated if it provides reliable risk estimates: formally speaking, calibration entails that if we consider all cases in which an AI system estimates a x % risk for a given condition or outcome, exactly x % of cases present the desired condition or outcome [12]. As discrimination metrics only focus on matching with a given ground truth, they are not able to measure calibration: indeed, even a perfectly accurate AI can be highly uncalibrated [13].

Transportability and Robustness: a critical limitation of internal validation is that it only evaluates the effectiveness of an AI system on the same population it was developed on. This may critically limit the applicability of an AI system, as it may fail when its use is translated to different settings, such as different hospitals, countries, or even the same setting but at a different time [14]. Transportability refers to the capacity of an AI system to be applied in different settings and scenarios while maintaining a satisfactory level of performance. Similarly, robustness refers to the related notion of an AI system being resistant to changes to its reference population in terms of patients' characteristics, data availability, or related dimensions [10]. Evaluating robustness and transportability is crucial to ensure reliability: standard practice is to accompany internal validation with external validation, which consists in testing an AI system's accuracy on data completely separate from the development set [7,12].

Interpretability and Explainability: Healthcare personnel often need justifications and understandable evidence to support their decision-making. Commonly employed modern AI systems, such as deep learning (DL) or large language models (LLMs), are often impenetrable black boxes, which may limit their applicability. Interpretability and explainability describe two distinct—but complementary—approaches to making AI systems understandable to human users [15]. Interpretability refers to methods that aim at clarifying the internal working of an AI system, while, by contrast, explainability refers to methods that aim at justifying the decisions and predictions issued by an AI system. In both cases, one can distinguish between inherent methods for interpretability (also called transparency) and explainability, which include white-box models such as decision trees or linear models as well as causal models, and post-hoc methods, which aim at interpreting or explaining a black-box model and include approaches such as surrogates or local explanation methods such as feature importance methods [16,17]. Evaluating interpretability and explainability is more complex than other dimensions, as they are socio-technical properties

requiring user-centered assessment of how well clinicians or personnel understand the system's support [18].

Fairness: commonly adopted performance metrics aim at evaluating the effectiveness of an AI system at the population level. Yet, such an evaluation can potentially hide discrimination or unfair treatment of different population sub-groups, wherein an AI system may exhibit unequal performance on minority groups [19]. Such differences may be due to intrinsic population differences or to systemic discrimination or bias in the development data. Fairness refers to an AI system ability to avoid undue discrimination [20]. Multiple ways exist to evaluate the fairness of an AI system, which include accuracy-inspired metrics for group fairness, such as demographic parity, predictive parity or equalized odds, calibration metrics, similarity-based metrics for individual fairness, or also causality-based metrics, which can bridge between individual and group-level criteria [21].

Usability, Appropriate Reliance and Trust Calibration: With the exception of explainability and interpretability, all described dimensions related either to data or to an AI system, separated from the socio-technical setting in which the data and the AI themselves are embedded. The socio-technical dimensions of performance concern the relation between an AI system and the organization they are embedded in, and are of critical importance. Usability refers to how the AI system enables users to interact effectively and satisfactorily to interact with the system, enabling user appropriation through effective user interfaces and support, as well as by clarifying its functioning and behavior [22]. Usability usually manifests in terms of appropriate reliance, that is the users should be able to discriminate when to accept and when to reject the support of an AI system [20,22,23].

3. The productivity paradox and the importance of reporting guidelines for AI in healthcare

The transformative potential of AI in healthcare remains largely unrealized, despite significant technological advances and substantial investments, a phenomenon exemplified by the well-documented "productivity paradox" (PP) that has historically characterized the adoption of general-purpose technologies [24]. The PP highlights the slow and uneven realization of productivity gains from revolutionary innovations, particularly in complex sectors such as healthcare. Historically, technologies like computing and digital connectivity have shown long delays between their introduction and measurable productivity impacts arising not only from technical limits but also from necessary complementary organizational, cultural, and workforce adaptations [24,25].

When applied to AI, this paradox emerges through four main drivers including overly optimistic expectations, measurement issues, concentration of benefits coupled with rent dissipation and implementation lags due to workforce retraining needs [26].

3.1. Healthcare-specific challenges

Within healthcare, the productivity paradox is particularly pronounced due to regulatory constraints, market complexity, stringent privacy requirements, and fragmented stakeholder landscapes [25]. These factors have hindered the successful adoption of earlier general-purpose technologies, such as Electronic Health Records (EHRs) [25]. Despite large public investments (e.g., the HITECH Act) [27], EHR productivity outcomes were disappointing, increasing clinician workload and documentation burden [25,28–30]. This experience illustrates how misaligned incentives, poor user-centered design, and insufficient organizational alignment can undermine the productivity potential of innovation.

3.2. Technical and methodological barriers to AI adoption

Beyond these systemic implementation challenges, AI adoption in healthcare faces additional technical and methodological barriers that reinforce the paradox and are centered around three interconnected dimensions: reproducibility gaps, technological complexity, and data quality.

A critical barrier in AI research is the issue of reproducibility [31–33], which is pivotal for scientific validity and clinical translation. As underscored in a recent systematic review by Kolbinger et al. [34], AI research in medicine suffers from significant variability in reporting standards, which compromises transparency, comparability, and ultimately reproducibility. More, the reproducibility crisis stems from AI's intrinsic non-determinism. Machine learning (ML) models depend on stochastic optimization, random initialization, and hardware variability; even changing a single random seed can double reported performance estimates [35]. DL trained with stochastic gradient descent yield different parameters on each run, making reproducibility dependent on precise documentation of numerous hidden parameters and software versions [36,37]. These challenges manifest across complexity dimensions encompassing data model and learning [31–33].

Data quality remains a critical bottleneck for healthcare AI. Biomedical data are often heterogeneous, incomplete, biased, and difficult to access due to privacy regulations [38,39]. These limitations foster overfitting, poor generalizability, and data leakage, inflating reported performance and complicating validation. Integrating multimodal datasets from different sources requires complex preprocessing, which can itself introduce variability and inconsistency.

3.3. Generative AI-specific challenges

The rise of GenAI adds further complexity to the productivity paradox [40]. Unlike discriminative models, generative systems produce novel outputs, raising concerns about accuracy, bias, and accountability.

A key issue is hallucination, where LLMs generate plausible but factually incorrect statements with high confidence [41]. Since these models predict words based on statistical patterns rather than factual reasoning, hallucinations can cause dangerous misinformation about treatments, drug interactions, or diagnoses [41].

LLMs also inherit and amplify biases that reinforce healthcare disparities [42,43]. Common manifestations include confirmation bias, framing bias (where wording affects outcomes), and anchoring bias (where initial impressions dominate reasoning) [44]. Addressing these issues requires techniques like Retrieval-Augmented Generation (RAG) and Supervised Fine-Tuning (SFT) to align model outputs with verified medical evidence [45–47].

Evaluation is further complicated because multiple valid answers can exist for clinical queries, necessitating new assessment frameworks sensitive to clinical nuance and patient safety [48].

3.4. Current solutions and reporting standards

To mitigate these challenges and narrow the productivity gap, several reporting frameworks have been developed. TRIPOD-AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis – AI) and CONSORT-AI (Consolidated Standards of Reporting Trials – AI) set methodological and transparency standards for AI research. They require detailed descriptions of data sources, preprocessing, model development, validation, and performance metrics.

TRIPOD-AI emphasizes transparent reporting of model design and validation, while CONSORT-AI extends similar requirements to clinical trials using AI systems, ensuring clear disclosure of system

Table 1
Summary of key reporting guidelines for studies involving Artificial Intelligence (AI) in healthcare. Each checklist is categorized by its intended study phase, application scope, primary reporting focus, target users, and AI-specific considerations.

Checklist	Scope of Use	Study Phase	Core Focus	Intended Users	AI-Specific Elements
TRIPOD-AI [49]	Prediction models (diagnostic/prognostic)	Model development & validation	Comprehensive documentation of AI-driven prediction tools	Model developers, biostatisticians, journal editors	Addresses handling of structured/unstructured data, model explainability, fairness, calibration, reproducibility
STARD-AI [50]	Diagnostic accuracy studies	In silico or retrospective analysis	Transparent reporting of diagnostic test performance	Clinician-scientists, diagnostic developers	Covers ground truth definition, data leakage risk, model bias, and AI-specific test characteristics
SPIRIT-AI [51]	Protocols for AI-based trials	Trial planning	Ensures methodological rigor before clinical testing	Trialists, regulatory reviewers, funding bodies	Requires pre-specification of AI interaction, data input/output logic, and usability plans
CONSORT-AI [52]	Results of clinical trials with AI	Post-intervention reporting	Clear and complete documentation of trial findings	Trial authors, peer reviewers, guideline developers	Includes description of AI system behavior, human-AI interaction, and error analysis
DECIDE-AI [53]	Early-phase clinical implementation	First real-world use (live)	Evaluation of usability, feasibility, and user-AI dynamics	Implementation researchers, human factors experts	Focuses on human-AI interaction, assisted performance, and iterative design-feedback loops
CLAIM [54]	AI in medical imaging	Model development & testing	Standardized reporting for reproducibility and transparency	Imaging researchers, radiologists, ML engineers	Recommends detailed reporting of data provenance, annotation, internal/external testing, imaging protocols
CHEERS-AI	Health economic evaluations of AI interventions	Economic modeling phase	Transparent and standardized reporting of economic evaluations of AI-based health technologies	Health economists, decision makers, journal editors	AI technique specification, user autonomy assessment, AI learning over time, development/validation details, implementation costs, AI-specific uncertainty
CHART [55]	Chatbot Health Advice studies	Performance evaluation of generative AI-driven chatbots	Transparent reporting of studies evaluating chatbots for clinical evidence and health advice	Clinicians, researchers, editors, peer reviewers, and readers	Model identification, prompt engineering details, query strategy, chatbot-specific performance evaluation, ground truth definition
TRIPOD-LLM [56]	LLM studies in biomedical applications	Model development, tuning, prompt engineering, and evaluation	Comprehensive reporting of LLM-based research across the model lifecycle	Academic/industry researchers, journal editors, peer reviewers, other stakeholders	Modular framework, transparency throughout model lifecycle, human oversight requirements, task-specific performance metrics, prompt engineering details, living document approach
GAMER [57]	All types of medical research (literature reviews, clinical trials, observational studies)	All research phases (study design, data collection, analysis, manuscript writing)	Transparent disclosure of Generative AI use across entire research process	Authors, reviewers, journal editors	Focus on tool declaration, specifications, prompting techniques, verification methods, data privacy, and impact assessment
IJMEDI/CHAMAI [58]	Quality assessment of Medical AI/ML studies	All CRISP-DM phases (problem understanding → deployment)	Methodological rigor, reproducibility, data quality, validation, reporting standards	Authors, reviewers, journal editors, AI developers in healthcare	Data variability, handling missing data, imbalance, ground truth reliability, validation (internal/external), interpretability, fairness, bias, carbon footprint, code/data sharing, deployment readiness

Abbreviations: TRIPOD-AI/LLM: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis Artificial Intelligence/Large Language Models; STARD: Standards for Reporting of Diagnostic Accuracy Studies; SPIRT: Standard Protocol Items: Recommendations for Interventional Trials; CONSORT: Consolidated Standards of Reporting Trials; DECIDE: Developmental and Exploratory Clinical Investigations of Decision Support Systems Driven by Artificial Intelligence; CLAIM: Checklist for Artificial Intelligence in Medical Imaging; CHEERS: Consolidated Health Economic Evaluation Reporting Standards; CHART: Chatbot Assessment Reporting Tool; GAMER: Generative Artificial intelligence tools in MEdical Research; IJMEDI/CHAMAI Checklist for Healthcare Assessment of Medical Artificial Intelligence.

functionality and evaluation criteria. These frameworks, when rigorously applied, standardize reporting, enhance transparency, and support reproducibility (see Table 1 for checklist examples).

New extensions address generative AI challenges: TRIPOD-LLM adapts, emphasizing transparency, human oversight, and task-specific performance, while CHART (Chatbot Assessment Reporting Tool) provides guidelines for evaluating AI-driven healthcare chatbots. However, adherence remains inconsistent, revealing a significant implementation gap between theoretical standards and actual practice [34].

4. Ethical considerations

AI can transform healthcare by enhancing research, prevention, diagnosis, treatment, and resource allocation. Its adoption, however, must go beyond efficiency and follow the WHO's principles for responsible AI: autonomy, safety, transparency, accountability, equity, and sustainability [1,3]. These principles acquire real value only when fully implemented in research, clinical practice, and patient experience.

A central ethical concern is the transformation of the medical profession. As physicians rely more on AI for diagnostic and therapeutic decisions, there is a risk of diminished professional competence and moral judgment. Excessive dependence on algorithms may lead to "moral de-skilling", weakening clinical authority and eroding patient trust [4]. Yet this shift may also offer opportunities: by delegating technical or repetitive tasks to AI, clinicians can devote more time to holistic care, reinforcing the human and ethical dimensions of medicine.

Patients, increasingly involved through data sharing and continuous monitoring, become more active in prevention and personalized care. However, this evolution may generate pressures to-

ward self-management and behavioral conformity. Moreover, medical data are shaped by social inequalities; underrepresentation or inconsistent coding can embed discriminatory patterns into AI models, thereby reproducing inequities [5]. Ethical AI development must confront these risks by ensuring fairness, privacy, and respect for individual rights.

Trust is fundamental to responsible AI. Authentic trust requires understanding rather than passive reliance. Because AI often operates as a "black box," clinicians need competencies to interpret and explain algorithmic outputs, while patients must be empowered to question decisions and exercise autonomy. Public healthcare systems are essential in promoting equity, education, and shared literacy.

Ultimately, AI ethics in medicine must transition from abstract principles to practical implementation. Safeguarding autonomy, ensuring transparency, fostering accountability, promoting inclusiveness, and pursuing sustainability must guide the entire AI life cycle—from design to validation and clinical deployment—and be incorporated into medical training and patient education. Only by reinforcing these human capacities can AI enhance, rather than undermine, the ethical foundations of medicine: professional responsibility, patient autonomy, and mutual trust [6].

5. AI and clinical education

A fundamental barrier to the integration of AI in hepatology is the persistent gap between technological innovation and clinician preparedness. While AI methods have advanced rapidly, the ability of healthcare professionals to interpret, validate, and communicate algorithmic outputs has not progressed at the same pace [59,60]. As highlighted by Topol, "without a strong foundation in data science and critical thinking, physicians risk becoming mere 'clickers' rather than true interpreters of clinical context" [59].

This disparity risks creating a "two-speed" medicine, in which data scientists build models detached from clinical reality, while clinicians struggle to understand and trust their implications.

5.1. AI literacy and professional development

AI literacy, the capacity to comprehend, evaluate, and appropriately apply AI outputs, should become a core competency for hepatologists. Training must extend beyond technical familiarity to include understanding of uncertainty, bias, and the epistemic limits of AI-generated knowledge [40,41].

Medical curricula should integrate foundational data science, algorithmic reasoning, and ethics, emphasizing interpretive responsibility rather than computational proficiency. Clinicians must remain critical interpreters of data-driven insights, not passive recipients of algorithmic output.

5.2. The role of the third sector and patient associations

Patient organizations and civil society actors, including associations operating within the Third Sector, can play a decisive role in promoting transparency, education, and co-design of AI tools. Their participation ensures that innovation remains aligned with patients' values and societal expectations. Collaborative frameworks such as *EUPATI* (European Patients' Academy on Therapeutic Innovation) demonstrate that shared learning between clinicians, patients, and researchers fosters accountability, inclusivity, and trust [61,62]. National programmatic documents, including the "Piano Nazionale della Prevenzione" and the Guidelines for the Implementation of Digital Health, explicitly emphasize the role of associations in promoting digital literacy in healthcare [62].

5.3. Human–AI collaboration and workflow integration

Human–AI collaboration requires a paradigm shift: AI systems must be integrated into clinical workflows as decision-support companions rather than autonomous actors. The principle of *human-in-the-loop* ensures continuous human validation of algorithmic outputs, preserving medical judgment and patient-centered care [22,63,64].

To foster appropriate reliance, AI interfaces should communicate confidence levels, calibration information, and model uncertainty. Visual explanation tools can enhance situational awareness, allowing clinicians to weigh algorithmic input against contextual factors [65,66].

Automation bias—where clinicians over-trust machine suggestions—must be mitigated through organizational safeguards, transparent model documentation, and iterative user training. Conversely, underutilization due to skepticism can deprive patients of beneficial innovation. Thus, successful integration depends on "trust calibration", balancing human expertise with computational intelligence.

In light of this, AISF could take the lead in promoting a structured alliance with representatives of the third sector for the co-design of educational and informative programs, thus contributing to the conscious dissemination of AI in clinical practice and the construction of a trust-based and co-responsible ecosystem among institutions, clinicians, and citizens.

6. AISF position statements and key principles on AI in hepatology

Recognizing all mentioned AI-linked issues, AISF emphasizes the importance of key recommendations for the application of AI approaches in hepatology (Box 1).

7. Selected application of AI in hepatology

AI has demonstrated tangible potential across nearly all domains of hepatology, from non-invasive diagnosis to transplant management.

7.1. Role of AI in MASLD/MASH

The rising global prevalence of metabolic dysfunction-associated steatotic liver disease (MASLD), now affecting approximately 38 % of the population, has intensified the clinical demand for precise, scalable diagnostic and management strategies [67].

AI technologies, especially ML, DL and neural networks, are increasingly used to address current limitations in MASLD diagnosis and prognostic evaluation. This trend is reflected in the sharp rise in related research: between 2020 and 2025, the number of studies applying AI to MASLD grew over 11-fold compared to earlier years (https://pubmed.ncbi.nlm.nih.gov/?term=artificial±intelligence±NAFLD&sort=date, Pubmed accessed on 24/07/2025).

Indeed, encouraging results in identifying MASLD and its progressive stages, including fibrosis and metabolic dysfunction-associated steatohepatitis (MASH), were obtained by leveraging molecular omics data, structured clinical information, and imaging. For instance, recent studies have applied ML to public liver transcriptome datasets and histological images from liver biopsy to identify key prognostic signatures, yielding good performance [68–71]. However, despite promising outcomes, the identified genes or histological patterns may vary across datasets and techniques, requiring cautious interpretation and external validation.

When applied to clinical data and instrumental analyses, supervised ML algorithms have shown superior accuracy in detecting

Box 1
AISF position statements on Artificial Intelligence in hepatology.

Box 1. AISF Position Statements and Key Principles

AISF recognizes AI as a strategic enabling technology for the future of hepatology, with the potential to improve diagnostic accuracy, prognostic stratification, and care personalization.

AISF considers data quality, standardization, and interoperability to be critical national priorities for enabling safe and effective AI in hepatology and supports the development of shared, harmonized data infrastructures.

AISF encourages the development of shared multimodal data repositories, including clinical, biochemical, imaging, histological, and omics data, to advance scientific research and enable robust, reproducible, and generalizable artificial intelligence applications in hepatology.

AISF identifies prospective, multicenter, and real-world clinical studies as a strategic research priority to validate AI systems and translate them from experimental tools into safe and effective components of routine liver care.

AISF supports the principle that AI models used in hepatology must be evaluated not only for discrimination performance (e.g., AUC, sensitivity), but also for calibration, robustness, transportability, fairness, and explainability.

AISF considers transparency in data provenance, preprocessing pipelines, model architecture, and validation procedures a prerequisite for scientific credibility and clinical translation of AI systems.

AISF supports the principle of *human-in-the-loop* decision-making, whereby AI must function as a clinical decision-support tool and never as a fully autonomous decision maker in hepatology.

AISF states that appropriate reliance and trust calibration should be explicit design goals of AI systems, requiring interfaces that communicate uncertainty, confidence, and limitations to clinicians and patients.

AISF recognizes algorithmic bias as a major threat to equity in liver care and recommends that all AI models be systematically evaluated for differential performance across sex, age, ethnicity, and relevant clinical subgroups.

AISF affirms that no AI system intended for clinical use in hepatology should be adopted without robust external validation across independent populations and healthcare settings.

AISF recognizes Generative AI as a promising yet high-risk class of technologies and supports their use in hepatology only when combined with domain-specific knowledge sources, retrieval-based systems, and human oversight.

AISF supports the incorporation of AI literacy, data science principles, and ethical reasoning into the education and continuous professional development of hepatologists.

AISF encourages multidisciplinary collaboration between clinicians, researchers, data scientists, engineers, computer scientists, roboticists, ethicists, and patient representatives throughout the entire lifecycle of AI systems, from design to clinical implementation.

AISF recognizes patient organizations and Third Sector actors as essential partners in the co-design, evaluation, and responsible dissemination of AI-based tools in hepatology.

AISF supports the alignment of AI development and deployment in hepatology with European and national regulatory frameworks, including the EU Artificial Intelligence Act and the Italian digital health strategy.

hepatic steatosis and MASH, incorporating routine demographic, clinical, and laboratory parameters (including age, sex, serum liver enzymes, glucose). In same case, these models outperformed conventional non-invasive tests and risk scores, such as APRI and FIB-4 [72,73].

Concerning medical imaging, convolutional neural networks (CNNs) have been applied to ultrasound, magnetic resonance (MR) imaging and computed tomography (CT) scans to automate the quantification of hepatic steatosis and fibrosis, while minimizing the inter-variability (i.e. different instruments or protocols) [74–77].

Large Language Models (LLMs) are emerging tools in MASLD/MASH management, offering support in guideline interpretation, clinical reporting, and patient communication. Early studies show that LLMs, especially when combined with retrieval-augmented generation, can provide clinically relevant outputs and approximate the performance of some non-invasive assessments [78]. Broader reviews confirm their potential for disease stratification and monitoring, although robust validation, domain-specific fine-tuning, and high-quality datasets remain essential to ensure reliability, safety, and real-world applicability [79].

Importantly, the multimodal integration of clinical, biological, and imaging data is essential for comprehensive disease characterization, and AI offers the capability to effectively perform such integration, as also shown by recent efforts on MASLD-related adverse outcomes [80].

Despite such encouraging results, many published studies applying AI to MASLD still exhibit substantial limitations that restrict the generalizability and clinical applicability of their findings.

One major challenge lies in the variability in data collection methods, diagnostic criteria, and annotation standards, which prevents comparison or harmonization among studies pursuing similar objectives [2]. Moreover, most models were developed and tested within single-center or retrospective cohorts, lacking external validation across diverse populations and clinical settings.

Model interpretability, particularly critical for DL-based models, essential for the successful clinical integration of AI tools [81]. Understanding how predictions are generated, which variables influence them, and how to optimize models for real-world use is crucial [82].

7.2. Applications of AI in autoimmune and cholestatic liver diseases

Autoimmune and cholestatic liver diseases, including autoimmune hepatitis (AIH), primary biliary cholangitis (PBC), and primary sclerosing cholangitis (PSC), remain difficult to manage due to overlapping features, variable progression, and limited reliable biomarkers. AI can address these challenges by introducing standardized, quantitative, and integrative approaches across histology, imaging, and clinical data.

Whole slide imaging combined with DL is reshaping liver histopathology. In AIH, the AI(H) model showed high accuracy in identifying portal inflammation and necroinflammation [83]. Quantitative histological analyses also have prognostic value: in pediatric autoimmune liver disease, fibrosis burden correlates with biliary involvement and adverse outcomes [84]. Beyond slide-level classification, DL enables disease discrimination. Gerussi et al. [85] demonstrated that whole slide image–based models could differentiate AIH from PBC without manual annotations. The HOTSPoT pipeline offers an open-source, robust framework for segmenting portal tracts and fibrosis in liver biopsies [86]. This approach can be extended to future studies of autoimmune liver diseases. Collectively, these tools enhance diagnostic precision, reduce inter-observer variability, and generate quantitative biomarkers beyond traditional histology. enhancing accuracy, prognostic stratification, and individualized monitoring in PSC. Radiomics and DL applied to MR and MRCP are rapidly advancing the field of PSC, where accurate imaging is central for diagnosis and monitoring. The DeePSC model achieved robust performance in automated PSC diagnosis [87]. Quantitative MRCP-derived scores predict medium-term outcomes [88], while composite indices integrating imaging features

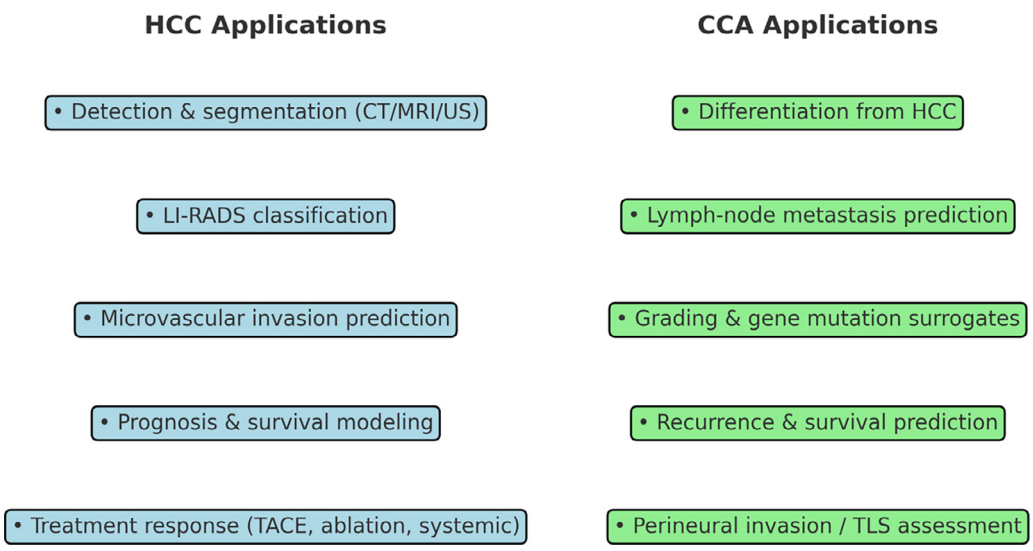

Fig. 1. Overview of main AI applications in HCC versus CCA.

correlate strongly with prognosis [89]. Quantitative MRCP metrics are also associated with disease severity and clinical outcomes, further underscoring their prognostic value [90]. AI-driven MRI analysis has enabled accurate early detection of perihilar cholangiocarcinoma in PSC patients [91]. Together, these approaches provide reproducible, non-invasive biomarkers of fibrosis, inflammation, and malignancy risk. By enhancing diagnostic accuracy and prognostic stratification, they can improve monitoring strategies and support more individualized treatment planning in PSC.

Large language models (LLMs) extend AI applications to clinical reporting and decision support. Daza et al. [92] compared four chatbots in autoimmune liver disease, revealing variability in accuracy but supporting their potential in patient engagement and triage. Colapietro et al. evaluated ChatGPT-4 in AIH, showing it could provide relevant insights but required domain-specific adaptation to avoid hallucinations [93]. Emerging multimodal systems link digital pathology, imaging, and serological data, enabling LLMs to generate standardized reports or answer clinician queries in natural language. Such applications remain early-stage, and dedicated clinical agents are still needed.

7.3. Role of AI in viral hepatitis

Chronic viral hepatitis B and C pose a major public health challenge due to their potential to progress to cirrhosis and increase HCC risk [3,94]. Despite effective antiviral therapies and widespread HBV vaccination-strategies that have markedly improved prognosis-important challenges remain in screening, risk stratification, and personalized management [3,94]. In this context, AI is emerging as a valuable tool, capable of integrating and analyzing large volumes of clinical, biochemical, histological, and radiological data [40,95].

When it comes to screening, AI is proving useful in optimizing processes for identifying individuals with undiagnosed chronic viral infection. The Spanish Intelligen-C strategy, an AI-based decision support system combining retrospective data analysis and an automated screening process for hospitalized or emergency patients, proved effective in increasing detection of HCV-RNA-positive individuals [96]. Moreover, Obaido et al. explored ML approaches to support HBV diagnosis. Using demographic and clinical records from the UC Irvine ML Repository, researchers applied a Shapley Additive exPlanations (SHAP)-based analysis to evaluate multiple classifiers. Among them, the AdaBoost model achieved the best performance, reaching 92 % accuracy in predicting HBV status [97].

Several ML models have been developed in recent years with regard to risk stratification, particularly with the aim of predicting the development of HCC in patients with chronic viral hepatitis [95]. In a large cohort study conducted in Hong Kong, Wong et al. developed and validated ML models useful to predict the risk of HCC in patients with chronic hepatitis B and/or C [98]. These models were trained using clinical and biochemical data from over 120,000 individuals. Among the tested algorithms, ridge regression and random forest models exhibited superior performance in comparison to conventional HCC risk scores, achieving AUROC values of 0.844 and 0.837, respectively [98].

Recent advances in large language models (LLMs) have created new opportunities for supporting clinical decision-making [40]. Giuffrè et al. evaluated OpenAI's GPT-4 Turbo in recommending direct-acting antiviral (DAA) regimens and answering open-ended questions about HCV management [99]. They tested retrieval-augmented generation (RAG) and supervised fine-tuning (SFT) using the 2020 EASL HCV guidelines as external knowledge or fine-tuning data. Retrieval strategies incorporating broader contextual information significantly improved model performance. Models with RAG showed greater accuracy (up to 91.7 %) and clarity (91.7 %) than baseline GPT-4 Turbo (36.6 % accuracy, 46.6 % clarity). Similarly, SFT-enhanced GPT-4 achieved 71.7 % accuracy and 88.3 % clarity [99]. These findings highlight the potential of LLMs, particularly when enriched with domain-specific expertise, to assist clinicians in customizing antiviral therapy, ensuring adherence to protocols, and improving care consistency. Though preliminary, such tools are expected to become increasingly relevant in individualized viral hepatitis management.

7.4. Role of AI in drug-induced liver hepatitis

Drug-induced liver injury (DILI) is a complex and potentially severe adverse reaction to drugs, herbal products, or dietary supplements that can mimic the clinical presentation of other liver diseases. In severe cases, DILI may progress to acute liver failure [100]. However, the lack of distinct diagnostic biomarkers still hinders its diagnosis. Given the related public health risks, new methodologies are needed to improve DILI management [101]. Thanks to electronic health records, larger datasets are increasingly available, highlighting the potential of AI for risk stratification and prognos-

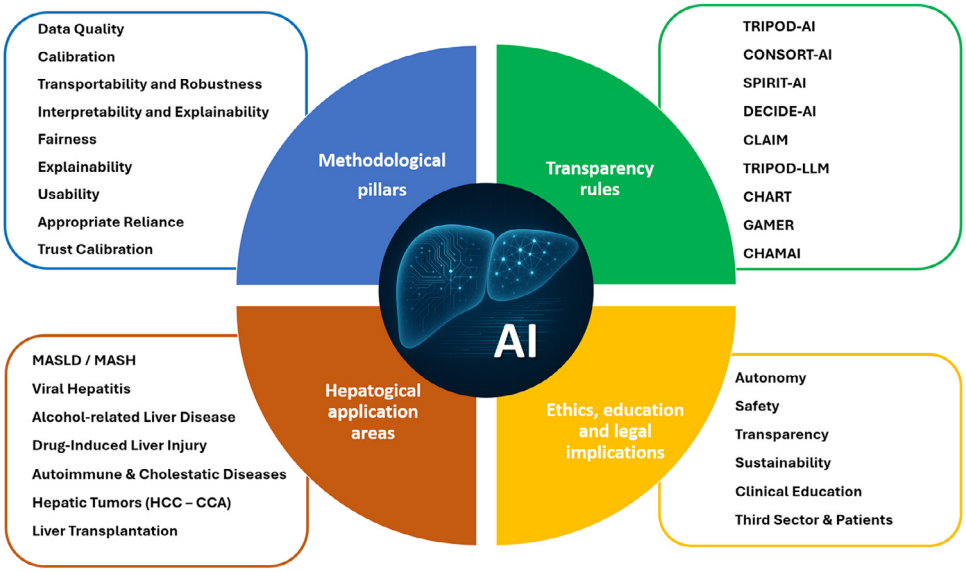

Fig. 2. Graphical abstract.

tic evaluation of DILI [102]. However, AI use in this setting remains limited.

In terms of risk stratification, several AI-based models have been developed in recent years using clinical, biochemical, genomic, or molecular data. A DL model capable of distinguishing DILI at a genomic level was also generated [103]. Trained on gene expression data from DILI-positive and -negative samples obtained from public datasets, the algorithm showed excellent predictive performance, with 97.1 % accuracy, 97.4 % sensitivity, and an AUC of 0.989 [103]. Furthermore, Yu et al. explored salivary metabolomics as a non-invasive diagnostic tool for DILI [104]. After identifying differentially expressed metabolites in DILI patients versus healthy controls, they applied machine ML build predictive models. This approach identified five salivary metabolites that distinguished DILI patients from controls, with AUC values ranging from 0.733 to 1.00 [104].

In the context of prognostic evaluation, a study based on data from the Spanish DILI Registry used a bootstrap-enhanced least absolute shrinkage operator (Bolasso) to select variables, followed by multivariable logistic regression to predict poor outcomes in 912 patients experiencing their first episode of idiosyncratic DILI [105]. The resulting model incorporated prior drug allergies, R-ratio-based pattern of liver injury, female sex, AST, total bilirubin, and platelet count. It exhibited excellent predictive performance, with an AUC of 0.887 in the Spanish DILI Registry and 0.932 upon external validation in the LATINDILI Network for predicting liver-related death or transplantation [105].

The potential application of Natural Language Processing (NLP) in the context of DILI has also recently been explored, particularly to support the assessment of causality in DILI through by automatically classifying regulatory documents and scientific literature [106,107]. However, although promising, the evidence is still preliminary and requires further development.

7.5. Role of AI in alcoholic liver disease

Alcoholic liver disease (ALD) is a leading global cause of morbidity and mortality, encompassing conditions from simple steatosis to alcoholic hepatitis, cirrhosis, and acute-on-chronic liver failure [108]. In Italy, its prevalence is notably high at 16.1 %, the highest among the European countries assessed [109]. Prevention strategies span multiple levels: reducing alcohol consumption in the general population (primary prevention), preventing progression in those with early-stage ALD (secondary prevention), and minimizing complications and mortality in advanced stages such as cirrhosis or alcohol-related hepatitis (tertiary prevention).

AI-based approaches may offer significant potential to revolutionize the clinical management of ALD, enabling more sensitive, non-invasive, and personalized methods for detection and monitoring, while also supporting timely interventions to enhance patient outcomes [110]. A psychosocial model predicting post-transplant harmful alcohol use showed strong internal performance (AUC = 0.930) but declined markedly on external validation (AUC = 0.692), underscoring limitations with subjective data [111]. Mezzacappa et al. [112] identified proteomic ML panels in 596 subjects (360 biopsied) that accurately detected fibrosis (AUC = 0.92) and inflammation (AUC = 0.87), also predicting clinical outcomes (C-index = 0.90; 0.79). A global study across 23 centers showed ML models predicting 30-day mortality (AUC = 0.811) and 90-day mortality (AUC = 0.799), outperforming traditional scores [113], though broader validation remains necessary.

AI models integrating clinical and omics data show strong diagnostic and prognostic potential as non-invasive alternatives, but evidence remains preliminary, with limited validation and ethical concerns requiring broader multicenter standardization before clinical adoption.

7.6. AI in hepatic tumors: current applications in hepatocellular carcinoma and cholangiocarcinoma

AI, encompassing ML, DL, and radiomics, is reshaping the diagnosis and management of hepatic tumors, particularly hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA). AI applications span early detection, characterization, prognostication, pathology, and prediction of treatment response.

Radiomics pipelines extract quantitative descriptors from CT and MRI to aid detection, characterization, and staging of HCC and CCA. Meta-analyses show that CT-based AI/radiomics achieves moderate pooled accuracy, while MRI-based radiomics demonstrates promising but heterogeneous results requiring standardization and external validation [114–116]. DL has also enabled automated LI-RADS classification and lesion segmentation, improving the diagnostic workflow [117].

Whole-slide image (WSI) analysis has emerged as a powerful AI application. Multi-instance learning models can predict microvascular invasion (MVI), a crucial prognostic feature in HCC, directly from hematoxylin and eosin slides with high accuracy. These models provide interpretable heatmaps and may reduce inter-observer variability [118,119]. Imaging-based AI surrogates of MVI also show potential to guide surgical and transplant strategies [120,121].

AI models integrating imaging and clinical data predict outcomes such as overall survival, recurrence-free survival, and response to locoregional therapies (e.g., surgical ablation, TACE, radioembolization). A systematic review reported pooled AUROCs above 0.80, supporting their clinical value while stressing the need for multi-center, prospective validation [122]. These models may allow individualized treatment allocation and closer follow-up for high-risk patients.

In CCA, AI is being applied to differential diagnosis (distinguishing intrahepatic CCA from HCC), lymph-node metastasis prediction, tumor grading, genomic surrogate prediction, and prognosis. Radiomics-based models have been shown to outperform clinical-feature-only models for intrahepatic CCA detection and to predict microvascular invasion, mutational status, and recurrence [123,124]. Nevertheless, methodological quality remains limited, with insufficient external validation and retrospective study designs [125].

Despite strong performance metrics, barriers to clinical translation remain. Data heterogeneity, lack of standardization, limited reproducibility, and inconsistent adherence to reporting standards actually prevent clinical implementation. Prospective, multicenter studies with harmonized acquisition protocols, robust external validation, and explainable AI frameworks will be essential to achieve clinical adoption in the field of HCC and CCA (Fig. 1).

7.7. Role of AI in transplant medicine

AI has long influenced transplantation, showing strong potential in pre- and post-liver transplant care. By analyzing vast datasets, it uncovers hidden relationships among demographic, clinical, genetic, and imaging variables that traditional methods often miss [126,127].

It is well recognized that the demand for liver transplantation continues to expand with advanced aging of donor and recipient, recipient complex comorbidities and new indications such as early transplantation for severe acute alcohol-related hepatitis acute on chronic liver failure, MASLD/MASH, and transplant oncology [114,115].

Transplant decision-making faces significant challenges, including the risk of delaying surgery beyond the optimal window, which can lead to futile transplantation.

A semi-structured interview study with 53 multidisciplinary liver transplant providers was applied. At the end, 6 themes were raised to be important identified six priorities for the design of fair AI-based clinical decision support for liver transplant listing decisions:

1) transparency in the creators behind the AI.
2) understanding how the AI uses data to support recommendations.
3) acknowledgement that AI could mitigate emotions and biases.
4) AI as a member of the transplant team, not a replacement.
5) identifying patient resource needs.
6) including the patient's role.

When the multidisciplinary liver transplant providers were asked about their perception about AI-based clinical decision, they reported they would feel more comfortable if the tool was made by academic institutions, suggested playing an active part in the development of the tool early and concluded that they were cautiously optimistic about the potential for AI to improve clinical outcomes for patients [128].

Although the use of AI in transplant medicine raises justified concerns, it plays a crucial role in refining donor-recipient evaluation through detailed, data-driven analyses. One of the most promising areas involves the histological assessment of donated organs [122].

In Italy, hepatologists and pathologists from the University of Padua and Oxford University are leading the PNRR project *"Bringing Medicine Digitalization into the Italian Solid Organ Transplant Network"* The initiative applies ML to integrate donor and recipient data—including anthropometric parameters, laboratory results, graft histology, and surgical variables such as ischemia time—to predict post-transplant outcomes (graft and patient survival, complications). The integration by convolutional neural networks (CNNs) of histologic, biochemical, and clinical data supports accurate outcome prediction.

One of the earliest studies by the Bologna group and collaborators showed that artificial neural networks outperformed MELD scores in predicting mortality among patients with end-stage liver disease [129]. Nearly two decades later, AI has been integrated into the Gender-Equity Model for Liver Allocation (GEMA-AI) to improve waiting list prioritization. Using the same variables as existing models, GEMA-AI outperformed GEMA-Na, MELD 3.0, and MELD-Na, showing greater benefits for women. Its nonlinear design reclassified about 6 % of patients within 90 days, potentially preventing one death per 59 candidates [130].

Could AI assist in liver allocation? This is a topic to be approached with caution. Currently, the ethics in the field of AI and transplant remain unexplored, and the view of the public unknown [131].

Liver transplant allocation is highly complex, influenced by donor and recipient characteristics, disease etiology, comorbidities, and socio-ethical factors, often leading to non-uniform prioritization across centers.

7.8. Future directions and recommendations

In healthcare, a coordinated strategy is required for addressing both technical and systemic barriers. This includes greater transparency in training and validation processes, open data and code sharing, and access to computational resources to enable independent verification.

Authorities should establish flexible, risk-based rules ensuring patient safety, ethical integrity, and transparency while fostering innovation. Policies promoting data interoperability, standardized reporting, and incentives for compliance with guidelines can improve reproducibility and accelerate real-world implementation.

In summary, resolving the productivity paradox demands a multifaceted approach: enforcing standardized methodologies, strengthening reproducibility, improving data quality, fostering organizational adaptation, and modernizing regulatory oversight. Without these complementary transformations, AI's promise of enhanced productivity, clinical outcomes, and efficiency will remain largely theoretical. Conversely, through transparent, reproducible, and ethically grounded AI integration, healthcare can achieve meaningful gains in both productivity and patient care.

8. Conclusion

AI in Hepatology may enhance diagnostic and therapeutic accuracy while allowing clinicians to devote more time to relational and empathetic aspects of care. By automating repetitive tasks and improving the management of chronic conditions, AI can optimize resources and reduce medical errors (Fig. 2). Its effective in-

tegration, however, requires ethical safeguards, transparency, and continuous professional training to ensure responsible, safe, and patient-centered use.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

[1] Ethics and governance of artificial intelligence for health. World Health Organization; 2024.
[2] Campagner A, Biganzoli EM, Balsano C, et al. Modeling unknowns: a vision for uncertainty-aware machine learning in healthcare. Int J Med Inf 2025;203:106014. doi:10.1016/j.ijmedinf.2025.106014.
[3] Easterbrook PJ, Luhmann N, Bajis S, et al. WHO 2024 hepatitis B guidelines: an opportunity to transform care. Lancet Gastroenterol Hepatol 2024;9:493–5. doi:10.1016/S2468-1253(24)00089-X.
[4] Natali C, Marconi L, Dias Duran LD, et al. AI-induced deskilling in medicine: a mixed-method review and research agenda for healthcare and beyond. Artif Intell Rev 2025;58:356. doi:10.1007/s10462-025-11352-1.
[5] Cross JL, Choma MA, Onofrey JA. Bias in medical AI: implications for clinical decision-making. PLOS Digit Health 2024;3:e0000651. doi:10.1371/journal.pdig.0000651.
[6] Teo ZL, Thirunavukarasu AJ, Elangovan K, et al. Generative artificial intelligence in medicine. Nat Med 2025;31:3270–82. doi:10.1038/s41591-025-03983-2.
[7] Balsano C, Alisi A, Brunetto MR, et al. The application of artificial intelligence in hepatology: a systematic review. Dig Liver Dis 2022;54:299–308. doi:10.1016/j.dld.2021.06.011.
[8] Balsano C, Burra P, Duvoux C, et al. Artificial intelligence and liver: opportunities and barriers. Dig Liver Dis 2023;55:1455–61. doi:10.1016/j.dld.2023.08.048.
[9] Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health 2021;3:e745–50. doi:10.1016/S2589-7500(21)00208-9.
[10] Cabitza F, Campagner A, Soares F, et al. The importance of being external. Methodological insights for the external validation of machine learning models in medicine. Comput Methods Programs Biomed 2021;208:106288. doi:10.1016/j.cmpb.2021.106288.
[11] Staartjes VE, Kernbach JM. Letter to the editor. Importance of calibration assessment in machine learning–based predictive analytics. J Neurosurg Spine 2020;32:985–7. doi:10.3171/2019.12.SPINE191503.
[12] Bella A, Ferri C, Hernández-Orallo J, Ramírez-Quintana MJ. Calibration of machine learning models. In: Handbook of research on machine learning applications and trends. IGI Global; 2010. p. 128–46. doi:10.4018/978-1-60566-766-9.ch006.
[13] Arrieta-Ibarra I, Gujral P, Tannen J, Tygert M, Xu C. Metrics of calibration for probabilistic predictions. J Mach Learn Res 2022;23:1–54.
[14] Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal–external, and external validation. J Clin Epidemiol 2016;69:245–7. doi:10.1016/j.jclinepi.2015.04.005.
[15] Marcinkevičs R, Vogt JE. Interpretable and explainable machine learning: a methods-centric overview with concrete examples. WIREs Data Min Knowl Discov 2023;13. doi:10.1002/widm.1493.
[16] Cynthia R. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206–15.
[17] Sérgio J. How can I choose an explainer? An application-grounded evaluation of post-hoc explanations. The ACM Conference on Fairness, Accountability, and Transparency; 2021.
[18] Famiglini L, Campagner A, Barandas M, et al. Evidence-based XAI: an empirical approach to design more effective and explainable decision support systems. Comput Biol Med 2024;170:108042. doi:10.1016/j.compbiomed.2024.108042.
[19] Chen RJ, Wang JJ, Williamson DFK, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. Nat Biomed Eng 2023;7:719–42. doi:10.1038/s41551-023-01056-8.
[20] Binns R. On the apparent conflict between individual and group fairness. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. ACM; 2020. p. 514–24. doi:10.1145/3351095.3372864.
[21] Castelnovo A, Crupi R, Greco G, et al. A clarification of the nuances in the fairness metrics landscape. Sci Rep 2022;12:4209. doi:10.1038/s41598-022-07939-1.
[22] Cutillo CM, Sharma KR, Foschini L, et al. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med 2020;3:47. doi:10.1038/s41746-020-0254-2.
[23] Cabitza F, Campagner A, Angius R, et al. AI shall have no dominion: on how to measure technology dominance in AI-supported human decision-making. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. ACM; 2023. p. 1–20. doi:10.1145/3544548.3581095.
[24] Brynjolfsson E. The productivity paradox of information technology. Commun ACM 1993;36:66–77. doi:10.1145/163298.163309.
[25] Wachter RM, Brynjolfsson E. Will generative artificial intelligence deliver on its promise in health care? JAMA 2024;331:65. doi:10.1001/jama.2023.25054.
[26] Brynjolfsson E., Rock D., Syverson C. Artificial Intelligence and the modern productivity paradox: a clash of expectations and statistics. Cambridge, MA: 2017. https://doi.org/10.3386/w24001.
[27] HITECH act of 2009, Pub. L. 111-5, 123 stat.227 (Feb. 17, 2009). Available Online: https://www.HhsGov/Sites/Default/Files/Ocr/Privacy/Hipaa/Understanding/Coveredentities/HitechactPdf n.d.
[28] Trout KE, Chen L-W, Wilson FA, et al. The impact of meaningful use and electronic health records on hospital patient safety. Int J Env Res Public Health 2022;19:12525. doi:10.3390/ijerph191912525.
[29] Dixit RA, Boxley CL, Samuel S, et al. Electronic health record use issues and diagnostic error: a scoping review and framework. J Patient Saf 2023;19:e25–30. doi:10.1097/PTS.0000000000001081.
[30] Hswen Y, Voelker R. Electronic health records failed to make clinicians' lives easier—will AI technology succeed? JAMA 2023;330:1509. doi:10.1001/jama.2023.19138.
[31] Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. JAMA 2020;323:305. doi:10.1001/jama.2019.20866.
[32] Han H. Challenges of reproducible AI in biomedical data science. BMC Med Genom 2025;18:8. doi:10.1186/s12920-024-02072-6.
[33] Ciobanu-Caraus O, Aicher A, Kernbach JM, Regli L, et al. A critical moment in machine learning in medicine: on reproducible and interpretable learning. Acta Neurochir (Wien) 2024;166:14. doi:10.1007/s00701-024-05892-8.
[34] Kolbinger FR, Veldhuizen GP, Zhu J, et al. Reporting guidelines in medical artificial intelligence: a systematic review and meta-analysis. Commun Med 2024;4:71. doi:10.1038/s43856-024-00492-0.
[35] Henderson P, Islam R, Bachman P, et al. Deep reinforcement learning that matters; 2019.
[36] Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018;319:1317. doi:10.1001/jama.2017.18391.
[37] Hinton G. Deep learning—a technology with the potential to transform health care. JAMA 2018;320:1101. doi:10.1001/jama.2018.11100.
[38] Sriram V, Conard AM, Rosenberg I, et al. Addressing biomedical data challenges and opportunities to inform a large-scale data lifecycle for enhanced data sharing, interoperability, analysis, and collaboration across stakeholders. Sci Rep 2025;15:6291. doi:10.1038/s41598-025-90453-x.
[39] Mohammed S, Budach L, Feuerpfeil M, et al. The effects of data quality on machine learning performance on tabular data 2025. Inform Syst 2025:132. doi:10.1016/j.is.2025.102549.
[40] Giuffrè M, Kresevic S, Pugliese N, et al. Optimizing large language models in digestive disease: strategies and challenges to improve clinical outcomes. Liver Int 2024;44:2114–24. doi:10.1111/liv.15974.
[41] Giuffrè M, You K, Shung DL. Evaluating ChatGPT in medical contexts: the imperative to guard against hallucinations and partial accuracies. Clin Gastroenterol Hepatol 2024;22:1145–6. doi:10.1016/j.cgh.2023.09.035.
[42] Poulain Raphael FHBR. Bias patterns in the application of LLMs for clinical decision support: a comprehensive study. ArXiv 2024.
[43] Omiye JA, Lester JC, Spichak S, et al. Large language models propagate race-based medicine. NPJ Digit Med 2023;6:195. doi:10.1038/s41746-023-00939-z.
[44] Mahajan A, Obermeyer Z, Daneshjou R, et al. Cognitive bias in clinical large language models. NPJ Digit Med 2025;8:428. doi:10.1038/s41746-025-01790-0.
[45] Kresevic S, Giuffrè M, Ajcevic M, et al. Optimization of hepatological clinical guidelines interpretation by large language models: a retrieval augmented generation-based framework. NPJ Digit Med 2024;7:102. doi:10.1038/s41746-024-01091-y.
[46] Giuffrè M, You K, Pang Z, et al. Expert of experts verification and alignment (EVAL) framework for large language models safety in gastroenterology. NPJ Digit Med 2025;8:242. doi:10.1038/s41746-025-01589-z.
[47] Giuffrè M, Distefano A, Kresevic S, et al. Guideline-enhanced large language models outperform physician-test takers on EASL campus quizzes multiple choice questions. JHEP Rep 2025;7:101523. doi:10.1016/j.jhepr.2025.101523.
[48] Croxford E, Gao Y, Pellegrino N, et al. Current and future state of evaluation of large language models for medical summarization tasks. Npj Health Syst 2025;2:6. doi:10.1038/s44401-024-00011-2.
[49] Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024:e078378. doi:10.1136/bmj-2023-078378.
[50] Sounderajah V, Ashrafian H, Golub RM, et al. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021;11:e047709. doi:10.1136/bmjopen-2020-047709.
[51] Cruz Rivera S, Liu X, Chan A-W, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. Nat Med 2020;26:1351–63. doi:10.1038/s41591-020-1037-7.

[52] Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020;26:1364–74. doi:10.1038/s41591-020-1034-x.
[53] Vasey B, Novak A, Ather S, et al. DECIDE-AI: a new reporting guideline and its relevance to artificial intelligence studies in radiology. Clin Radiol 2023;78:130–6. doi:10.1016/j.crad.2022.09.131.
[54] Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for Artificial Intelligence in medical imaging (CLAIM): 2024 update. Radiol Artif Intell 2024;6. doi:10.1148/ryai.240300.
[55] Huo B, Collins G, Chartash D, et al. Reporting guideline for Chatbot health advice studies: the CHART statement. BMC Med 2025;23:447. doi:10.1186/s12916-025-04274-w.
[56] Gallifant J, Afshar M, Ameen S, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med 2025;31:60–9. doi:10.1038/s41591-024-03425-5.
[57] Luo X, Tham YC, Giuffrè M, et al. Reporting guideline for the use of generative artificial intelligence tools in medical research: the GAMER statement. BMJ Evid Based Med 2025. doi:10.1136/bmjebm-2025-113825.
[58] Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics. Int J Med Inf 2021;153:104510. doi:10.1016/j.ijmedinf.2021.104510.
[59] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56. doi:10.1038/s41591-018-0300-7.
[60] Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019;9. doi:10.7189/jogh.09.020318.
[61] Rahal C, Mohan J. The role of the third sector in public health service provision: evidence from 25,338 heterogeneous procurement datasets. J R Stat Soc Stat Soc 2025;188:1085–106. doi:10.1093/jrsssa/qnae092.
[62] Ministero della Salute. Piano nazionale della prevenzione 2020-2025. CcmNetworkIt 2020;2020:174.
[63] Naiseh M, Simkute A, Zieni B, et al. C-XAI: a conceptual framework for designing XAI tools that support trust calibration. J Responsible Technol 2024;17:100076. doi:10.1016/j.jrt.2024.100076.
[64] Mehrotra S, Mehrotra S, Degachi C, Vereschak O, Jonker CM, Tielman ML. A systematic review on fostering appropriate trust in human-AI interaction. ACM J Responsible Comput 2023;26:1–45.
[65] Lundberg Scott M, L S-I. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017. p. 4768–77.
[66] Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med 2021;4:4. doi:10.1038/s41746-020-00367-3.
[67] Wong VWS, Ekstedt M, Wong GLH, et al. Changing epidemiology, global trends and implications for outcomes of NAFLD. J Hepatol 2023;79:842–52. doi:10.1016/j.jhep.2023.04.036.
[68] Lu H, Mao Z, Zheng M, et al. Identification of hub gene for the pathogenic mechanism and diagnosis of MASLD by enhanced bioinformatics analysis and machine learning. PLoS One 2025;20:1–16. doi:10.1371/journal.pone.0324972.
[69] Zhang W, Lu W, Jiao Y, et al. Identifying disease progression biomarkers in metabolic associated steatotic liver disease (MASLD) through weighted gene co-expression network analysis and machine learning. J Transl Med 2025;23. doi:10.1186/s12967-025-06490-7.
[70] Yuan HY, Tong XF, Ren YY, et al. AI-based digital pathology provides newer insights into lifestyle intervention-induced fibrosis regression in MASLD: an exploratory study. Liver Int 2024;44:2572–82. doi:10.1111/liv.16025.
[71] Pulaski H, Harrison SA, Mehta SS, et al. Clinical validation of an AI-based pathology tool for scoring of metabolic dysfunction-associated steatohepatitis. Nat Med 2025;31:315–22. doi:10.1038/s41591-024-03301-2.
[72] Liu M, Jiang L, Yang J, et al. Development and validation of a machine learning-based model for prediction of liver fibrosis and MASH. J Clin Gastroenterol 2025;00:1–9. doi:10.1097/MCG.0000000000002166.
[73] Charu V, Liang JW, Mannalithara A, et al. Benchmarking clinical risk prediction algorithms with ensemble machine learning for the noninvasive diagnosis of liver fibrosis in NAFLD. Hepatology 2024;80:1184–95. doi:10.1097/HEP.0000000000000908.
[74] Pickhardt PJ, Blake GM, Graffy PM, et al. Liver steatosis categorization on contrast-enhanced CT using a fully automated deep learning volumetric segmentation tool: evaluation in 1204 healthy adults using unenhanced CT as a reference standard. Am J Roentgenol 2021;217:359–67. doi:10.2214/AJR.20.24415.
[75] Meneses JP, Qadir A, Surendran N, et al. Unbiased and reproducible liver MRI-PDFF estimation using a scan protocol-informed deep learning method. Eur Radiol 2024;35:2843–54. doi:10.1007/s00330-024-11164-x.
[76] Yousefzamani M, Babapour Mofrad F. Deep learning without borders: recent advances in ultrasound image classification for liver diseases diagnosis. Expert Rev Med Devices 2025;22:827–43. doi:10.1080/17434440.2025.2514764.
[77] Meneses JP, Arrieta C, della Maggiora G, et al. Liver PDFF estimation using a multi-decoder water-fat separation neural network with a reduced number of echoes. Eur Radiol 2023;33:6557–68. doi:10.1007/s00330-023-09576-2.
[78] Wu W, Guo Y, Li Q, Jia C. Exploring the potential of large language models in identifying metabolic dysfunction-associated steatotic liver disease: a comparative study of non-invasive tests and artificial intelligence-generated responses. Liver Int 2025;45. doi:10.1111/liv.16112.
[79] Pugliese N, Bertazzoni A, Hassan C, et al. Revolutionizing MASLD: how artificial intelligence is shaping the future of liver care. Cancers (Basel) 2025;17:722. doi:10.3390/cancers17050722.
[80] Kosick HM-K, McIntosh C, Bera C, et al. Machine learning models using non-invasive tests & B-mode ultrasound to predict liver-related outcomes in metabolic dysfunction-associated steatotic liver disease. Sci Rep 2025;15:24579. doi:10.1038/s41598-025-09288-1.
[81] Lai M, Chen IY, Lai JC, et al. Generative AI in hepatology: transforming multimodal patient-generated data into actionable insights. Hepatol Commun 2025;9. doi:10.1097/HC9.0000000000000683.
[82] Uche-Anya E, Anyane-Yeboa A, Berzin TM, et al. Artificial intelligence in gastroenterology and hepatology: how to advance clinical practice while ensuring health equity. Gut 2022;71:1909–15. doi:10.1136/gutjnl-2021-326271.
[83] Ercan C, Kordy K, Knuuttila A, et al. A deep-learning-based model for assessment of autoimmune hepatitis from histology: AI(H). Virchows Arch 2024;485:1095–105. doi:10.1007/s00428-024-03841-5.
[84] Khendek L, Castro-Rojas C, Nelson C, et al. Quantitative fibrosis identifies biliary tract involvement and is associated with outcomes in pediatric autoimmune liver disease. Hepatol Commun 2025;9. doi:10.1097/HC9.0000000000000594.
[85] Gerussi A, Saldanha OL, Cazzaniga G, et al. Deep learning helps discriminate between autoimmune hepatitis and primary biliary cholangitis. JHEP Rep 2025;7:101198. doi:10.1016/j.jhepr.2024.101198.
[86] Cazzaniga G, L'Imperio V, Bonoldi E, et al. Automating liver biopsy segmentation with a robust, open-source tool for pathology research: the HOTSPoT model. NPJ Digit Med 2025;8:455. doi:10.1038/s41746-025-01870-1.
[87] Ragab H, Westhaeusser F, Ernst A, et al. DeePSC: a deep learning model for automated diagnosis of primary sclerosing cholangitis at two-dimensional MR cholangiopancreatography. Radiol Artif Intell 2023;5. doi:10.1148/ryai.220160.
[88] Cristoferi L, Porta M, Bernasconi DP, et al. A quantitative MRCP-derived score for medium-term outcome prediction in primary sclerosing cholangitis. Dig Liver Dis 2023;55:373–80. doi:10.1016/j.dld.2022.10.015.
[89] Vuppalanchi R, Are V, Telford A, et al. A composite score using quantitative magnetic resonance cholangiopancreatography predicts clinical outcomes in primary sclerosing cholangitis. JHEP Rep 2023;5:100834. doi:10.1016/j.jhepr.2023.100834.
[90] Cazzagon N, El Mouhadi S, Vanderbecq Q, et al. Quantitative magnetic resonance cholangiopancreatography metrics are associated with disease severity and outcomes in people with primary sclerosing cholangitis. JHEP Rep 2022;4:100577. doi:10.1016/j.jhepr.2022.100577.
[91] Singh Y, Eaton JE, Venkatesh SK, et al. Deep learning analysis of MRI accurately detects early-stage perihilar cholangiocarcinoma in patients with primary sclerosing cholangitis. Hepatology 2025. doi:10.1097/HEP.0000000000001314.
[92] Daza J, Bezerra LS, Santamaría L, et al. Evaluation of four chatbots in autoimmune liver disease: a comparative analysis. Ann Hepatol 2025;30:101537. doi:10.1016/j.aohep.2024.101537.
[93] Colapietro F, Piovani D, Pugliese N, et al. Is ChatGPT-4 a reliable tool in autoimmune hepatitis? Am J Gastroenterol 2025;120:914–19. doi:10.14309/ajg.0000000000003179.
[94] Blach S, Terrault NA, Tacke F, et al. Global change in hepatitis C virus prevalence and cascade of care between 2015 and 2020: a modelling study. Lancet Gastroenterol Hepatol 2022;7:396–415. doi:10.1016/S2468-1253(21)00472-6.
[95] Su T-H, Kao J-H. Role of artificial intelligence in the management of chronic hepatitis B infection. Clin Liver Dis (Hob) 2024;23. doi:10.1097/CLD.0000000000000164.
[96] Castro Urda JL, Álvarez M, Cantero H, et al. The efficiency of artificial intelligence for management and clinical decision-making in the identification of patients with hidden HCV infection (Intelligen-C strategy). Gastroenterol Hepatol 2025;48:502362. doi:10.1016/j.gastrohep.2025.502362.
[97] Obaido G, Ogbuokiri B, Swart TG, et al. An interpretable machine learning approach for hepatitis B diagnosis. Appl Sci 2022;12:11127. doi:10.3390/app122111127.
[98] Wong GL-H, Hui VW-K, Tan Q, et al. Novel machine learning models outperform risk scores in predicting hepatocellular carcinoma in patients with chronic viral hepatitis. JHEP Rep 2022;4:100441. doi:10.1016/j.jhepr.2022.100441.
[99] Giuffrè M, Pugliese N, Kresevic S, et al. From guidelines to real-time conversation: expert-validated retrieval-augmented and fine-tuned <scp>GPT</scp>-4 for hepatitis C management. Liver Int 2025;45. doi:10.1111/liv.70349.
[100] Andrade RJ, Aithal GP, Björnsson ES, et al. EASL clinical practice guidelines: drug-induced liver injury. J Hepatol 2019;70:1222–61. doi:10.1016/j.jhep.2019.02.014.
[101] Li X, Tang J, Mao Y. Incidence and risk factors of drug-induced liver injury. Liver Int 2022;42:1999–2014. doi:10.1111/liv.15262.
[102] Niu H, Alvarez-Alvarez I, Chen M. Artificial intelligence: an emerging tool for studying drug-induced liver injury. Liver Int 2025;45. doi:10.1111/liv.70038.
[103] Feng C, Chen H, Yuan X, et al. Gene expression data based deep learning model for accurate prediction of drug-induced liver injury in advance. J Chem Inf Model 2019;59:3240–50. doi:10.1021/acs.jcim.9b00143.
[104] Yu S-M, Zheng H-C, Wang S-C, et al. Salivary metabolites are promising noninvasive biomarkers of drug-induced liver injury. World J Gastroenterol 2024;30:2454–66. doi:10.3748/wjg.v30.i18.2454.
[105] Niu H, Solis-Muñoz P, García-Cortés M, et al. Prior drug allergies are associated with worse outcome in patients with idiosyncratic drug-induced liver

injury: a machine learning approach for risk stratification. Pharmacol Res 2024;199:107030. doi:10.1016/j.phrs.2023.107030.
[106] Wu Y, Liu Z, Wu L, et al. BERT-based natural language processing of drug labeling documents: a case study for classifying drug-induced liver injury risk. Front Artif Intell 2021;4. doi:10.3389/frai.2021.729834.
[107] Chen M, Wu Y, Wingerd B, et al. Automatic text classification of drug-induced liver injury using document-term matrix and XGBoost. Front Artif Intell 2024;7. doi:10.3389/frai.2024.1401810.
[108] Bataller R, Arab JP, Shah VH. Alcohol-associated hepatitis. N Engl J Med 2022;387:2436–48. doi:10.1056/NEJMra2207599.
[109] Amonker S, Houshmand A, Hinkson A, et al. Prevalence of alcohol-associated liver disease: a systematic review and meta-analysis. Hepatol Commun 2023;7. doi:10.1097/HC9.0000000000000133.
[110] Narayanan P, Wu T, Shah VH, et al. Insights into ALD and AUD diagnosis and prognosis: exploring AI and multimodal data streams. Hepatology 2024;80:1480–94. doi:10.1097/HEP.0000000000000929.
[111] Lee BP, Roth N, Rao P, et al. Artificial intelligence to identify harmful alcohol use after early liver transplant for alcohol-associated hepatitis. Am J Transplant 2022;22:1834–41. doi:10.1111/ajt.17059.
[112] Mezzacappa C, Bhat M. Proteomic panels for alcohol-associated liver disease: accurate, but different enough from existing clinical tests? Gastroenterology 2023;165:1576. doi:10.1053/j.gastro.2023.09.002.
[113] Dunn W, Li Y, Singal AK, et al. An artificial intelligence-generated model predicts 90-day survival in alcohol-associated hepatitis: a global cohort study. Hepatology 2024;80:1196–211. doi:10.1097/HEP.0000000000000883.
[114] Sweta K, Dkhar W, Kadavigere R, et al. Diagnostic accuracy of computed tomography (CT)-based radiomics and artificial intelligence (AI) models in hepatocellular carcinoma: a systematic review and meta-analysis. Clin Radiol 2025;89:107042. doi:10.1016/j.crad.2025.107042.
[115] Brancato V, Cerrone M, Garbino N, et al. Current status of magnetic resonance imaging radiomics in hepatocellular carcinoma: a quantitative review with radiomics quality score. World J Gastroenterol 2024;30:381–417. doi:10.3748/wjg.v30.i4.381.
[116] Miranda J, Horvat N, Fonseca GM, et al. Current status and future perspectives of radiomics in hepatocellular carcinoma. World J Gastroenterol 2023;29:43–60. doi:10.3748/wjg.v29.i1.43.
[117] Stollmayer R, Güven S, Heidt CM, et al. LI-RADS-based hepatocellular carcinoma risk mapping using contrast-enhanced MRI and self-configuring deep learning. Cancer Imaging 2025;25:36. doi:10.1186/s40644-025-00844-6.
[118] Chen Q, Xiao H, Gu Y, et al. Deep learning for evaluation of microvascular invasion in hepatocellular carcinoma from tumor areas of histology images. Hepatol Int 2022;16:590–602. doi:10.1007/s12072-022-10323-w.
[119] Zhang X, Yu X, Liang W, et al. Deep learning-based accurate diagnosis and quantitative evaluation of microvascular invasion in hepatocellular carcinoma on whole-slide histopathology images. Cancer Med 2024;13. doi:10.1002/cam4.7104.
[120] Song D, Wang Y, Wang W, et al. Using deep learning to predict microvascular invasion in hepatocellular carcinoma based on dynamic contrast-enhanced MRI combined with clinical parameters. J Cancer Res Clin Oncol 2021;147:3757–67. doi:10.1007/s00432-021-03617-3.
[121] Zhang W, Guo Q, Zhu Y, et al. Cross-institutional evaluation of deep learning and radiomics models in predicting microvascular invasion in hepatocellular carcinoma: validity, robustness, and ultrasound modality efficacy comparison. Cancer Imaging 2024;24:142. doi:10.1186/s40644-024-00790-9.
[122] Kiani I, Razeghian I, Valizadeh P, et al. Performance of artificial intelligence models in predicting responsiveness of hepatocellular carcinoma to transarterial chemoembolization (TACE): a systematic review and meta-analysis. J Am Coll Radiol 2025. doi:10.1016/j.jacr.2025.08.028.
[123] Zerunian M, Polidori T, Palmeri F, et al. Artificial intelligence and radiomics in cholangiocarcinoma: a comprehensive review. Diagnostics 2025;15:148. doi:10.3390/diagnostics15020148.
[124] Xu L, Chen Z, Zhu D, Wang Y. The application status of radiomics-based machine learning in intrahepatic cholangiocarcinoma: systematic review and meta-analysis. J Med Internet Res 2025;27:e69906. doi:10.2196/69906.
[125] Cannella R, Vernuccio F, Klontzas ME, et al. Systematic review with radiomics quality score of cholangiocarcinoma: an EuSoMII radiomics auditing group initiative. Insights Imaging 2023;14:21. doi:10.1186/s13244-023-01365-1.
[126] Tran J, Sharma D, Gotlieb N, et al. Application of machine learning in liver transplantation: a review. Hepatol Int 2022;16:495–508. doi:10.1007/s12072-021-10291-7.
[127] Germani G, Angrisani D, Addolorato G, et al. Liver transplantation for severe alcoholic hepatitis: a multicenter Italian study. Am J Transplant 2022;22:1191–200. doi:10.1111/ajt.16936.
[128] Strauss AT, Sidoti CN, Sung HC, et al. Artificial intelligence-based clinical decision support for liver transplant evaluation and considerations about fairness: a qualitative study. Hepatol Commun 2023;7. doi:10.1097/HC9.0000000000000239.
[129] Cucchetti A, Vivarelli M, Heaton ND, et al. Artificial neural network is superior to MELD in predicting mortality of patients with end-stage liver disease. Gut 2007;56:253–8. doi:10.1136/gut.2005.084434.
[130] Gómez-Orellana AM, Rodríguez-Perálvarez ML, Guijo-Rubio D, et al. Gender-equity model for liver allocation using artificial intelligence (GEMA-AI) for waiting list liver transplant prioritization. Clin Gastroenterol Hepatol 2025. doi:10.1016/j.cgh.2024.12.010.
[131] Craxì L, Cottone PM, Sacchini D, et al. The Equitable Benefit approach to guide the assessment of medical and psychosocial factors in liver transplant candidacy. Liver Int 2024;44:2263–72. doi:10.1111/liv.16018.