Global Health & Medicine. 2026; 8(3):166-181. REVIEW
DOI: 10.35772/ghm.2026.01048

Clinical artificial intelligence (AI) in Japan: Regulatory pathways, domain-specific evidence, and its data infrastructure from an international perspective

Kenji Karako[1,*], Wei Tang[1,2]

[1] Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan;
[2] National Center for Global Health and Medicine, Japan Institute for Health Security, Tokyo, Japan.

Abstract: Artificial intelligence (AI) has advanced rapidly across clinical domains, generating both a growing evidence base and dedicated regulatory frameworks for AI-based software as a medical device (SaMD). This review provides a comprehensive assessment of clinical AI across five major domains—diagnostic imaging, gastrointestinal endoscopy, cardiology and remote patient monitoring, diagnosis of infectious diseases, and an AI-ready data infrastructure—examining Japan's regulatory framework, approved device portfolio, and research contributions in an international context. We reviewed literature published between 2019 and 2026, using Japan's regulatory trajectory, approved device portfolio, and domain-specific research output as the primary lens for international comparison and prioritizing prospective studies, multicenter trials, and real-world implementation reports. The state of evidence varies markedly across the domains examined: endoscopy AI has the strongest randomized trial base, while diagnostic imaging AI has seen a systematic decline in real-world performance despite large-scale regulatory approval. Across the three dimensions examined, Japan has a distinctive profile: its strengths are a regulatory and clinical deployment infrastructure—evince by an established program medical device pathway and among the world's highest densities of diagnostic imaging systems and endoscopy volumes—while the data infrastructure lags, constrained by limited open-access resources relative to programs such as The Cancer Imaging Archive and the European Health Data Space. Large language models and generative AI, falling largely outside existing SaMD frameworks, carry the risk of hallucinations and gaps in oversight that healthcare systems in Japan and abroad are only beginning to address. Japan's established program medical device regulatory pathway, high-volume clinical deployment infrastructure, and proven regulatory-approval-to-reimbursement pathway provide a strong foundation for clinical AI adoption; post-approval change management frameworks and clinical accountability mechanisms need to be strengthened, AI-ready data accessibility needs to be expanded, and validated tools need to be embedded within reimbursed clinical workflows to translate this foundation into internationally competitive AI development and deployment.

***Keywords*:** artificial intelligence (AI), machine learning, Japan, software as a medical device, data infrastructure

1. Introduction

Artificial intelligence (AI), and in particular machine learning and deep learning, has rapidly transformed clinical medicine over the past decade. Applications now span diagnostic imaging, electrocardiogram (ECG) interpretation, gastrointestinal endoscopy, diagnosis of infectious diseases, radiation treatment planning, surgical assistance, and remote patient monitoring (*1*). In several narrowly defined tasks—such as detecting diabetic retinopathy from fundus photographs or adenomas during colonoscopy—AI systems have demonstrated diagnostic accuracy comparable to or exceeding that of expert clinicians (*2*). These developments have generated substantial clinical interest and have accelerated the regulatory and health technology assessment processes needed to translate research findings into safe, effective, and reimbursable clinical tools (*3*).

The governance of AI-enabled clinical software has converged on the concept of software as a medical device (SaMD), defined by the International Medical Device Regulators Forum (IMDRF) as software that fulfils a medical purpose independently of hardware (*4*). Major regulatory entities—including the United States Food and Drug Administration (FDA), the European Union (EU), the United Kingdom (UK), Japan, China, Singapore, and South Korea—have each established or are actively developing SaMD-

specific frameworks that address risk classification, clinical evaluation, post-marketing surveillance, and the particular challenge of managing AI model updates over the product lifecycle (*5-10*). In parallel, the research community has developed standardized reporting guidelines—including CONSORT-AI for clinical trials (*11*), TRIPOD+AI for prediction model studies (*12*), and STARD-AI for diagnostic accuracy research (*13*)—to improve the transparency and reproducibility of AI-based clinical evidence (*14*). Despite this regulatory and methodological progress, systematic comparative analysis of clinical AI development across major healthcare systems—jointly examining regulatory maturity, clinical deployment infrastructure, and data governance—remains limited.

Meaningful international comparison of medical AI development requires examining not only regulatory frameworks but also three complementary dimensions: the maturity and transparency of regulatory and reimbursement pathways, the clinical infrastructure available to support deployment, and the data resources required for model training and validation. Japan has classified AI-based software as "program medical devices" under the Pharmaceuticals and Medical Devices (PMD) Act; as of September 2025, 51 of the 172 approved program medical devices carried an AI-utilization designation, spanning imaging, endoscopy, cardiology, and infectious disease applications (Figure 1) (*8,15*). The United States FDA maintains a non-exhaustive public list of over 1,400 AI-enabled medical device authorizations spanning three decades, reflecting a markedly higher absolute volume of cleared or approved AI tools (*16*). With respect to clinical infrastructure, Japan's density of CT, MRI, and PET scanners—184 units per million population, compared to 86 in the United States, 74 in Germany, and 19 in the United Kingdom—indicates a high-volume imaging environment that represents a substantial deployment base for AI-assisted clinical applications (Figure 2) (*17,18*). The third dimension—data infrastructure for AI model development and validation—varies substantially across countries in scale, accessibility, and governance, and is examined in detail in Section 7.

This review examines the current state of clinical AI across five major domains, positioning Japan's regulatory maturity, clinical deployment infrastructure, and data resources within an international comparative framework to assess where Japan stands and what developments are required to sustain progress. We examine evidence published from 2019 through 2026, with an emphasis on studies published since 2021, prioritizing prospective studies, multicenter trials, and real-world implementation reports, supplemented by key regulatory and guideline documents. Following a summary of the global regulatory landscape (Section 2), we address AI in diagnostic imaging (Section 3), endoscopy (Section 4), cardiology and remote monitoring (Section 5), diagnosis of infectious diseases (Section 6), and an AI-ready data infrastructure (Section 7). Cross-cutting challenges—including the methodological and regulatory implications of large language models and generative AI, which represent an emerging category of clinical tools with a distinct evidence and oversight profile—are discussed in Section 8, followed by conclusions in Section 9. Table 1 provides an overview of the landmark clinical studies discussed across these domains.

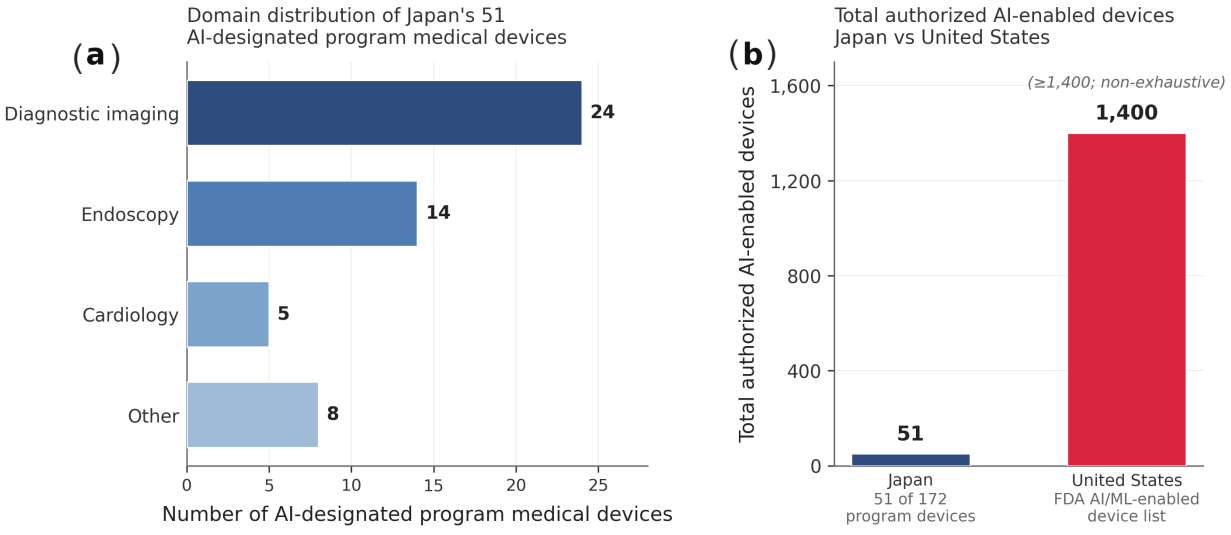

Figure 1. AI-enabled medical device authorizations as of September 2025. (a) Distribution of Japan's 51 designated program medical devices using AI; "Other" encompasses the diagnosis of infectious disease and additional applications not individually enumerated. **(b)** Total AI-enabled devices authorized in Japan (51 of 172 program medical devices) versus the United States FDA AI/ML-enabled device list (≥ 1,400 authorizations; non-exhaustive public list). Data sources: PMDA program medical device list (*15*); MHLW approval status (*33*); FDA AI/ML-enabled device list (*16*).

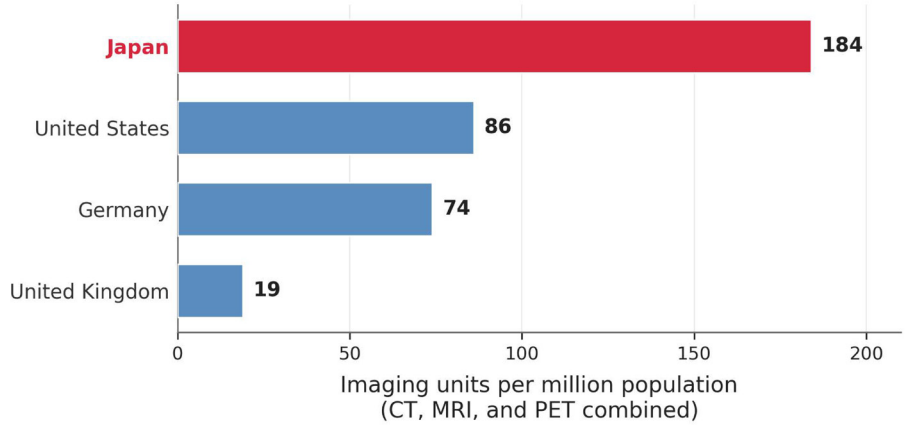

Figure 2. Density of CT, MRI, and PET scanners per million population in selected countries (2023). Japan's substantially higher scanner density compared to other major economies provides a high-volume imaging environment that constitutes a large deployment base for AI-assisted clinical applications. Data source: OECD Health Statistics 2025 (*17,18*).

Table 1. Landmark clinical AI studies by domain discussed in this review

Domain	Study	Design	Endpoint	Key finding
Radiology	McKinney *et al*. 2020 (*20*)	Retrospective validation	Breast cancer detection	Outperformed clinical reads on UK+US test sets; surpassed all 6 radiologists in US reader study
Endoscopy	Wang *et al*. 2019 (*21*)	RCT	ADR	34% *vs* 28% (control)
Endoscopy	Repici *et al*. 2020 *(22)*	RCT	ADR	54.8% *vs* 40.4% (control)
Cardiology (ECG)	Attia *et al*. 2019 (*23*)	Retrospective	AF detection in sinus rhythm	AUC 0.87 (454,789 ECGs)
Cardiology (wearable)	Perez *et al*. 2019 (*24*)	Prospective cohort	PPV for AF notification	84% concurrent AF confirmation (419,297 participants)
Infectious diseases	Okiyama *et al*. 2022 *(25)*	Prospective validation	Accuracy of influenza diagnosis	Comparable to or exceeding antigen rapid test, especially in early illness

Note: ADR, adenoma detection rate; AF, atrial fibrillation; AUC, area under the receiver operating characteristic curve; ECG, electrocardiogram; PPV, positive predictive value; RCT, randomized controlled trial.

2. The global regulatory landscape for medical AI

While major regulatory jurisdictions have converged on the SaMD concept established by the IMDRF, they differ considerably in how they classify, evaluate, and govern AI-based software over its clinical lifecycle—differences that directly determine the pace of and conditions under which AI tools reach clinical deployment.

2.1. The SaMD framework

The IMDRF classifies SaMD risk along two axes—the significance of information provided for clinical decision-making and the healthcare situation in which it is used—yielding four risk categories with progressively stringent requirements; clinical evaluation must demonstrate analytical validity, clinical validity, and clinical utility (*4,26*). The IMDRF N88 document (2025) formalizes shared Good Machine Learning Practice (GMLP) expectations—covering data quality, model transparency, and post-marketing monitoring—that underpin lifecycle management requirements across jurisdictions (*27,28*).

2.2. Regulatory approaches across major jurisdictions

Despite adopting the SaMD concept, major jurisdictions differ considerably in their classification systems, authorization pathways, and AI-specific guidance (Table 2).

The FDA's Predetermined Change Control Plan (PCCP) allows pre-specified model updates without a new submission; the EU AI Act (August 2024) classifies medical AI as high-risk under a dual MDR/AI Act compliance regime (*5,6,29*). The UK's MHRA publishes dedicated software and AI lifecycle guidance, and the National Institute for Health and Care Excellence (NICE) maintains an Evidence Standards Framework that stratifies evidence requirements for digital health technologies

Table 2. Regulatory frameworks for AI-based medical devices across major countries and regions

Country/Region	Regulatory framework	AI-specific characteristics
United States	Class I–III; 510(k)/De Novo/PMA (FDA)	PCCP for pre-specified model updates; Good Machine Learning Practice principles
European Union	MDR/IVDR; Notified Body conformity assessment	AI Act (high-risk classification); dual MDR–AI Act compliance requirement
United Kingdom	UK MDR 2002; MHRA authorization	MHRA SaMD/AI lifecycle guidance; NICE Evidence Standards Framework for NHS adoption
China	NMPA Class II/III registration	AI classification principles (2021); PIPL data governance requirements
Singapore	HSA SaMD lifecycle regulation	Change management guidance; AI in Healthcare Guidelines (AIHGle 2.0)
South Korea	MFDS; Digital Medical Products Act	Dedicated AI device evaluation guidelines covering trial design and validation
Japan	PMD Act; Program Medical Device	Machine learning review guidance from the PMDA; post-approval change management is considered

Note: FDA, Food and Drug Administration; GMLP, Good Machine Learning Practice; HSA, Health Sciences Authority; IVDR, In Vitro Diagnostic Regulation; MDR, Medical Device Regulation; MFDS, Ministry of Food and Drug Safety; MHRA, Medicines and Healthcare products Regulatory Agency; NHS, National Health Service; NICE, National Institute for Health and Care Excellence; NMPA, National Medical Products Administration; PCCP, Predetermined Change Control Plan; PIPL, Personal Information Protection Law; PMD Act, Pharmaceuticals and Medical Devices Act; PMDA, Pharmaceuticals and Medical Devices Agency; PMA, Premarketing Approval; SaMD, software as a medical device; UK MDR, UK Medical Device Regulations.

according to their function and risk, providing the basis for NHS procurement decisions (*7,30*).

In Asia, South Korea enacted the Digital Medical Products Act to provide a dedicated framework for AI-enabled digital health, Singapore's Health Sciences Authority has adopted a lifecycle approach with explicit change management guidance and has issued its AI in Healthcare Guidelines, and China's National Medical Products Administration published classification principles for AI-based medical software in 2021 (*9,10,31*). Japan regulates AI-based software as "program medical devices" under the PMD Act, with the Pharmaceuticals and Medical Devices Agency (PMDA) having published review considerations specific to machine learning-based devices covering risk–benefit assessment, analytical and clinical validation, and post-market change management (*8,32*). As of September 2025, 172 program medical devices held valid marketing authorization in Japan under the PMD Act framework, 51 of which carried an AI-utilization designation spanning imaging, endoscopy, cardiology, and infectious disease applications (Figure 1) (*15*). Regulatory authorization under the PMD Act is a prerequisite for, but distinct from, reimbursement under Japan's national health insurance system; coverage classification is determined separately by the Central Social Insurance Medical Council under the C1 and C2 new-technology categories—a pathway examined in the context of the nodoca system (Section 6.2) and discussed as a policy challenge in Section 8.3.

3. AI in diagnostic imaging

Of the 51 AI-utilization–designated program medical devices approved by the PMDA as of September 2025, 24 fall within the imaging category—the largest single subcategory—with approved indications spanning chest CT and plain radiography (detection of lung nodules, quantification of interstitial lung disease, and identification pneumonia), brain MRI (detection of an intracranial aneurysm), breast ultrasound (characterization of lesions), and musculoskeletal applications (detection of a fracture) (*15,33*). This concentration of approvals reflects both the maturity of the PMDA regulatory pathway for image analysis software and Japan's exceptional imaging infrastructure: with 184 CT, MRI, and PET units per million population—more than twice the density in the United States or Germany—Japan has one of the highest-volume deployment environments for AI-assisted clinical applications worldwide (Figure 2) (*17,18*).

In breast cancer screening, a deep learning model outperformed clinical reads on held-out UK and US test sets—reducing false-positive rates by 1.2% (UK) and 5.7% (US)—surpassing all six radiologists in an independent US reader study, and it reduced second-reader workload by 88% in a simulated UK double-reading workflow (*20*). In chest radiology and CT, AI systems have displayed improved performance across detection of lung nodules, characterization of interstitial lung disease, and triage of acute findings (*34*). In time-critical settings, automated large vessel occlusion (LVO) detection software significantly reduced door-to-groin (DTG) times and the time from CT initiation to the start of endovascular therapy (EVT) in acute stroke care (*35*).

Despite these proof-of-concept demonstrations, systematic analyses have consistently identified a

critical gap between single-facility performance and generalization to external populations: AI models validated at one site frequently exhibit marked performance degradation when applied to images acquired with different scanners, acquisition protocols, or patient demographics (*36*). Japan's imaging devices span a diverse mix of domestic manufacturers alongside international platforms operating under varying acquisition protocols, and training datasets compiled from a single center are unlikely to represent the full range of image characteristics encountered in clinical practice; differences in population-level prevalence and temporal drift from scanner upgrades further erode post-deployment calibration. These challenges underscore the importance of prospective multi-site validation and post-marketing performance monitoring (*12,13*).

Published real-world deployment evidence in Japan remains limited relative to the volume of regulatory authorizations. A prospective evaluation of AI approved for the detection of an intracranial aneurysm by Ito *et al*. highlights this gap: while diagnostic sensitivity was maintained in routine clinical use, the false-positive burden in unselected cases substantially exceeded that reported in the regulatory validation study (*37,38*). Comparable post-deployment evaluations of AI systems for chest imaging—the most numerous category within Japan's approved imaging portfolio—have not yet been widely published, reflecting a broader structural gap between pre-marketing validation requirements and the post-marketing performance evidence base.

4. AI in endoscopy

Gastrointestinal endoscopy is among the most evidence-rich domains for clinical AI, with a substantial body of randomized controlled trial (RCT) data supporting the integration of AI-based detection and characterization tools into colonoscopy practice. As of September 2025, 14 of the 51 AI-utilization–designated program medical devices approved by the PMDA fall within the endoscopy category—the second largest subcategory—reflecting high procedural volumes and an exceptional gastrointestinal malignancy burden (*15*). The endoscopy setting offers several features that facilitate AI development and evaluation: procedures are video-based and therefore generate large volumes of labelled training data; the primary clinical endpoint—the adenoma detection rate (ADR)—is a measurable surrogate for colorectal cancer prevention; and the procedural context allows direct real-time display of AI output to the endoscopist without requiring a separate reporting infrastructure.

4.1. Computer-aided detection and diagnosis: CADe and CADx

Computer-aided detection (CADe) systems analyze the live video stream and generate real-time alerts when the AI model identifies a region with a suspected polyp or other mucosal lesion, prompting the endoscopist to inspect the flagged area. Computer-aided diagnosis (CADx) systems go further, characterizing a detected lesion—for example, distinguishing a hyperplastic polyp from an adenoma, or predicting the depth of submucosal invasion—to aid in deciding on a strategy to "resect and discard" or "diagnose and leave," thus reducing unnecessary polypectomies and associated costs.

The most extensively studied clinical application is CADe-assisted colonoscopy for detection of colorectal adenoma. Wang *et al*. conducted an RCT in which patients undergoing colonoscopy were allocated to AI-assisted detection or conventional colonoscopy, and they found that the group with AI-assistance had a significantly higher ADR (34% versus 28%), with the system detecting additional adenomas that were missed in the withdrawal phase (*21*). A subsequent multicenter RCT by Repici *et al*. confirmed these findings in an Italian population, demonstrating that real-time AI-aided colonoscopy increased the ADR compared to standard colonoscopy (54.8% versus 40.4%) (*22*). Multiple meta-analyses of RCTs have since confirmed that CADe-assisted colonoscopy improves the ADR as a pooled outcome, with consistent findings across different AI systems and endoscopic settings (*39*).

Despite these efficacy data, CADe implementation raises important questions: the ADR is a surrogate endpoint for interval cancer prevention with no prospective proof that AI-assisted ADR gains reduce the long-term incidence of interval cancer; false-positive alerts add inspection time and risk operator fatigue; and the benefit is heterogeneous across endoscopists, suggesting AI assistance functions primarily as a quality levelling tool rather than a universal performance enhancer (*40*).

High-confidence optical characterization of diminutive polyps (≤ 5 mm) is the prerequisite for the "resect and discard" strategy, in which diminutive adenomas are removed without pathological examination and hyperplastic polyps are left in situ. Achieving the negative predictive value threshold (≥ 90% for high-confidence diagnoses) required by international guidelines requires not only high AI accuracy but also rigorous operator training and standardized image capturing conditions—requirements that have proved difficult to satisfy in routine endoscopy practice outside expert centers (*41*).

4.2. Japan's AI endoscopy portfolio and multicenter trial contributions

The age-standardized rate (ASR) for the incidence of colorectal cancer in Japan is 36.6 per 100,000 population—substantially above rates in the United States (27.0), United Kingdom (30.9), and Germany

(25.7)—while the ASR for the incidence of gastric cancer is 27.6 per 100,000, compared to 4.1 in the United States (*42,43*). This dual malignancy burden translates into high endoscopic procedure volumes, an established endoscopic training culture, and a correspondingly large supply of annotated training data for AI development.

EndoBRAIN (Olympus Corporation / Cybernet Systems) is a CADx system that uses a convolutional neural network to characterize colorectal lesions in real time during a colonoscopy, outputting a probability score for neoplastic versus non-neoplastic tissue (*40,44*). The device received PMDA approval and has been deployed in clinical practice at multiple Japanese facilities. The EndoBRAIN-EYE variant, which integrates CADe functionality with the existing CADx platform, has been evaluated in prospective clinical trials registered in the University Hospital Medical Information Network Clinical Trials Registry (UMIN-CTR), including the EYE-OPENER trial (*45*). These prospective studies are designed to assess whether the combined CADe/CADx platform improves clinically meaningful outcomes in routine endoscopy practice, going beyond the ADR to examine resection rates, procedure duration, and operator workload. Japan's high burden of gastric malignancy has similarly driven the development of AI for gastric endoscopy: PMDA-approved program medical devices in this domain include tools for detection of gastric lesions and for assessment of the depth of early gastric cancer invasion—the latter is a clinically critical distinction that determines eligibility for endoscopic versus surgical resection (*15*). The existence of approved tools across both colorectal and gastric applications reflects how Japan's epidemiological profile has shaped the scope of its regulatory portfolio in AI-assisted endoscopy.

Recognizing that AI systems trained and validated predominantly in single-country populations may not be generalizable across the diverse ethnicities, dietary patterns, and bowel preparation practices of Asian populations, a multicenter trial initiative coordinated by the National Cancer Center Japan was announced in 2024 (*46*). This program aims to evaluate the efficacy of AI-assisted colonoscopy across multiple Asian countries in a prospectively registered, multicenter design, with the explicit goal of generating evidence of external validity applicable to the broader Asia-Pacific region. The initiative is particularly significant because the majority of high-quality RCTs underpinning AI-assisted colonoscopy have been conducted in predominantly East Asian or European populations with limited cross-population generalizability; Japan's capacity to coordinate multinational Asian trials positions it as a key contributor to closing this gap in external validity.

5. AI in cardiology and remote patient monitoring

Cardiovascular disease remains the leading cause of death globally, and the cardiological domain offers particularly rich sources of structured longitudinal data—electrocardiograms (ECGs), echocardiograms, implantable device telemetry, and wearable biosignals—that are well-suited to machine learning. Of the 51 AI-utilization–designated program medical devices approved in Japan as of September 2025, five fall within cardiology—the smallest subcategory by volume, reflecting the complexity of cardiac AI development relative to image analysis tasks—and yet Japan's high cardiovascular disease burden, aging population, and extensive cardiac registry infrastructure represent a deployment environment with considerable potential (*15,18*).

5.1. AI for ECG interpretation and cardiac risk stratification

The standard 12-lead ECG is the most widely performed cardiac examination worldwide, and its digital format makes it immediately tractable for deep learning analysis. A foundational study by Attia *et al*. at the Mayo Clinic trained a convolutional neural network on 454,789 ECGs from 180,922 patients and demonstrated that the model could identify patients who had AF but were currently in sinus rhythm at the time of recording, with an area under the receiver operating characteristic curve (AUC) of 0.87 in a validation set of 36,280 patients (*23*). This finding implied that the AI system detected subtle ECG features reflecting atrial remodeling that precede the clinical manifestation of atrial fibrillation (AF)—a latent clue not evident in conventional ECG interpretation. A companion study by the same group used a similar architecture to detect left ventricular (LV) systolic dysfunction, achieving an AUC of 0.93 at identifying patients with an ejection fraction of 35% or less, suggesting that population-level screening for LV dysfunction using routine ECGs may be feasible (*47*).

These results have stimulated substantial interest in research, but several barriers to clinical translation remain. Most published models are trained and validated on retrospective data from single facilities, and external validation across different ECG recording systems, electrode placement standards, and demographic populations has revealed a significant heterogeneity in performance (*48*). The clinical actionability of a positive AI alert also depends critically on the availability of a downstream cardiological assessment: a model that correctly identifies patients at risk of future AF is of clinical value only if those patients can be promptly referred for confirmatory testing and, where indicated, started on anticoagulation therapy. Without a structured care pathway linked to AI output, the population-level benefit of screening remains theoretical. Early prospective evidence from a Japanese cohort study suggests that AI-ECG screening can identify AF with a higher discriminative ability than conventional risk scoring, offering a practical approach to population-based

screening in older adults (*49*). Prospective multicenter validation studies of AI-ECG tools specifically in Japanese populations remain limited relative to the volume of approved products, and the existing cardiac registry infrastructure—notably the Japan Registry of All Cardiac and Vascular Diseases (JROAD, covering 1,500 facilities as of 2023) and the Japan Percutaneous Coronary Intervention Registry (J-PCI, with over 2.4 million cumulative cases)—represents an underutilized platform with which to close this gap in external validity (*50,51*) (discussed further in Section 7).

5.2. Wearable devices and AF screening

The Apple Heart Study, conducted in the United States between 2017 and 2018, enrolled 419,297 participants who wore an Apple Watch and consented to receive irregular pulse notifications triggered by an algorithm detecting inter-beat interval irregularity (*24*). Among participants who received a notification, 84% had AF confirmed with a simultaneous ECG patch worn at the time of notification, and 34% had AF identified in a subsequent 90-day monitoring period. While the positive predictive value for concurrent AF was high, the study highlighted a fundamental challenge of population screening: persistent AF was not ultimately confirmed in the majority of notification recipients during follow-up, requiring medical evaluation for a large number of individuals whose downstream management trajectory remained uncertain.

The Fitbit Heart Study subsequently evaluated PPG-based irregular rhythm detection on a different consumer platform in a US cohort and reported a high positive predictive value (98.2%) for AF when an irregular rhythm notification was issued during concurrent ECG monitoring (*52*). Taken together, these landmark studies established that consumer wearables can achieve clinically relevant sensitivity in AF detection, but they also demonstrated that the translation of a wearable alert into a meaningful clinical outcome requires a carefully designed confirmation and care pathway. High notification rates in low-prevalence populations generate large numbers of notifications requiring medical follow-up, potentially overwhelming cardiology services and exposing patients to the anxiety and risks associated with unnecessary examination.

Calibrating the detection algorithm threshold—balancing sensitivity against the false-positive notification rate—is a design decision with direct implications for healthcare system workload and patient experience that must be addressed before population-scale wearable AF screening programs can be implemented responsibly (*53*). In Japan, home-use program medical devices incorporating AF detection notifications and sleep apnea detection alerts have received regulatory approval and are commercially available (*15,33*), but systematic evaluation of their real-world impact on clinical outcomes and healthcare utilization in the Japanese context has not yet been widely reported. The contrast with the Apple Heart Study—which enrolled 419,297 participants in a single prospective study—illustrates the scale at which the leading wearable AF screening programs have been evaluated internationally; prospective wearable studies of comparable scale have not yet been reported in Japanese cohorts, despite the direct relevance of AF screening to Japan's aging population. Japan's demographic profile—the highest proportion of adults age 65 or older among OECD countries—creates a strong demand for AI-assisted home monitoring (*18,19*). CureApp HT (a smartphone-based hypertension management device) began to be covered by national insurance in 2022 after demonstrating cost-effectiveness, illustrating the pathway through which evidence-based digital health tools can qualify for reimbursement in Japan (*54*).

6. AI in the diagnosis of infectious diseases

The diagnosis of infectious diseases presents a distinctive set of challenges for AI: the pathogen in question, disease prevalence, and the clinical presentation of infection vary with season, geographic region, circulating strain, and host immunity—factors that shift the data distribution that a model encounters after deployment relative to the distribution on which it was trained. AI applications in this domain span three broad functional categories: image-based diagnosis (interpreting chest radiographs or CT scans for pneumonia and related conditions), augmentation of a rapid test (automating or checking the quality of the reading of lateral flow or other point-of-care tests), and multimodal clinical integration (combining symptom profiles, vital signs, and laboratory or imaging results to generate a differential diagnosis probability—as exemplified by the nodoca system for influenza diagnosis discussed in Section 6.2). Japan holds a distinctive position in this domain: despite a smaller portfolio of AI tools approved for diagnosis of approved infectious diseases compared to imaging or endoscopy, it has produced one of the most operationally complete regulatory-approval-to-reimbursement trajectories for an AI-based clinical decision-making support system in the Asia-Pacific region, through the nodoca system for diagnosis of influenza, and it has since expanded approved AI applications to include culture-free identification of species causing bacterial infections (*15*).

6.1. The COVID-19 experience and methodological lessons

Studies proliferated rapidly during the COVID-19 pandemic, with many reporting a high diagnostic performance in detecting pneumonia due to SARS-CoV-2 on chest CT or plain radiography using deep learning models. A systematic methodological review

by Roberts *et al*. evaluated 415 published machine learning studies involving COVID-19 detection and prognosis based on chest imaging, and it concluded that none met the requirements for clinical use (*55*). The most commonly identified problems included dataset contamination (patients appearing in both training and test sets), use of retrospective convenience samples with unrepresentative negative controls, a lack of external validation, and failure to account for the clinical context in which the model would be deployed. This analysis provided a cautionary benchmark for the field, illustrating how the urgency of a public health crisis can accelerate publication volume while simultaneously eroding methodological standards.

A key challenge specific to AI diagnosis of infectious diseases is the distributional shift driven by changing prior probability: a model trained during a COVID-19 wave will encounter a very different disease prevalence in an inter-epidemic period, substantially altering positive and negative predictive values even if model sensitivity and specificity are unchanged (*3*). This structural vulnerability means that AI systems to diagnose infectious diseases require more frequent performance re-evaluation and retraining than systems applied to stable disease populations, with post-marketing monitoring designed around seasonal and epidemic periodicity.

6.2. nodoca: AI-assisted diagnosis of influenza in Japan

nodoca (Aillis Inc., Tokyo) is a program medical device that acquires images of the posterior pharynx using a dedicated imaging device and it integrates this visual information with structured clinical data—including patient age, symptom onset, body temperature, and vaccination history—to estimate the probability of influenza and generate a differential diagnosis to assist the clinician in his or her assessment. Okiyama *et al*. reported that the deep learning model trained on pharyngeal images and structured clinical records achieved an influenza detection accuracy comparable to, and in some subgroups exceeding, conventional rapid antigen tests—particularly in the early phase of illness when viral loads may fall below standard kit sensitivity (*25*).

The PMDA reviewed and approved nodoca as a program medical device, with a regulatory review report addressing its intended clinical use—an aid to diagnose influenza—together with evidence of its analytical and clinical validation, the performance boundaries of the model, and post-marketing surveillance requirements (*56*). In December 2022, the Central Social Insurance Medical Council granted nodoca a C2 reimbursement classification, evaluating it as a new technology fee rather than as a standard device covered by insurance (*57*). This sequence—prospective clinical validation, PMDA regulatory authorization, and Chuikyo's approval of its reimbursement—constitutes the most complete and transparent example of the regulatory–reimbursement chain for an AI-based clinical decision-making support system in Japan to date, and its serves as a reference pathway for developers of subsequent AI diagnostic tools.

Implementation challenges include image quality consistency in busy primary care settings, the need for annual recalibration as dominant circulating strains and vaccine components change across influenza seasons, and unresolved assignment of liability when an AI-assisted diagnostic decision leads to an adverse outcome (*3*).

6.3. Broader applications and future directions

Beyond influenza and COVID-19, AI's applications in the diagnosis of infectious diseases are expanding into detection of tuberculosis on chest radiography (*58*) and prediction of sepsis risk based on clinical and laboratory data. Sepsis represents a disease burden on a scale that warrants AI-based detection tools: global estimates indicate approximately 48.9 million cases and 11 million deaths annually, accounting for roughly 20% of all deaths worldwide (*59,60*). In 2024, the Sepsis ImmunoScore—an AI tool predicting sepsis severity based on host immune biomarkers—received FDA De Novo authorization, establishing a regulatory reference point for AI to evaluate sepsis that Japan has not yet matched under the PMDA framework (*61*). In Japan, the portfolio of AI approved for diagnosis of infectious diseases does extend beyond influenza: PMDA-authorized program medical devices include AI software for estimating bacterial species causing urinary tract infections without culture and a microbial classification support program—tools that complement nodoca's multimodal approach and signal an expanding scope for culture-free AI-based diagnosis of infectious diseases in Japan (*15*). Systematic clinical validation data regarding these newer approved tools have not yet been widely published, representing an evidence gap comparable to that observed in the imaging domain.

Adherence to STARD-AI and TRIPOD+AI reporting standards (*12,13*) is especially critical to AI for diagnosis of infectious diseases given the domain's structural susceptibility to distributional shift—a challenge that post-marketing surveillance frameworks in Japan and abroad are only beginning to systematically address.

7. An AI-ready data infrastructure: International initiatives and Japan's position

The performance of AI models in clinical medicine is bounded not only by algorithmic design but by the quality, scale, and interoperability of the data on which they are trained and validated. As evidence has accumulated that models trained at single facilities regularly fail to generalize across sites, the field has

converged on a consensus that a large-scale, multimodal, and externally accessible data infrastructure is a necessary precondition for robust AI development. This section surveys the principal international initiatives for building an AI-ready medical data infrastructure, characterizes Japan's current position, and comparatively analyzes strengths, gaps, and trajectories. Japan's position in this landscape exemplifies a structural paradox that recurs across the domains reviewed in this paper: the country generates some of the world's largest volumes of clinical and administrative health-related data through its universal insurance system, and yet the infrastructure for converting this data into AI-ready, externally accessible research platforms lags materially behind that of comparable OECD nations—a gap that is the central analytical focus of this section.

7.1. International initiatives for AI-ready medical data

Maintained by the National Cancer Institute, the Cancer Genome Atlas (TCGA) provides multi-omic and clinical data for over 20,000 samples across 33 cancer types and has underpinned a substantial fraction of published computational oncology and pathology AI research (*62*). The Cancer Imaging Archive (TCIA) complements the TCGA with over 30.9 million medical images from 37,568 subjects across a wide range of modalities and clinical contexts, making it the world's largest openly accessible repository of annotated medical imaging data (*63*). The NIH Bridge2AI program and the All of Us Research Program extend this infrastructure with explicit AI-readiness mandates and large-scale multimodal cohorts designed to support prospective validation on a population scale (*64,65*).

In Europe, the European Health Data Space (EHDS), which entered into force in March 2025, establishes a cross-member harmonized framework for the secondary use of electronic health data, including provisions for federated analysis without requiring data to leave national borders (*66*). The EHDS is expected to create a governance infrastructure for large-scale multicenter AI training and validation across EU member states—a capability structurally absent from European medical AI research to date. The United Kingdom's NHS AI Lab developed shared medical imaging platforms, though independent evaluation has highlighted implementation challenges in translating infrastructure investments into active use in research (*67*).

In Asia, South Korea has adopted a centralized curation strategy: the National Information Society Agency (NIA) manages the AI Hub, a nationally maintained repository of annotated datasets—including medical imaging, clinical records, and biosignal collections—made available to domestic and international researchers under defined access conditions (*68*). The South Korean model, in which government-funded agencies take direct responsibility for dataset quality and annotation, offers an instructive counterpoint to the federated governance designs favored in Europe and the market-facilitated repository model in the United States. In the Asian context, South Korea also represents the most direct structural comparator to Japan: both countries operate universal health insurance systems that generate large-scale administrative claims data, and yet South Korea has invested in a nationally coordinated annotation infrastructure for AI research that Japan has not yet replicated at a comparable scale—a contrast that frames the analysis of Japan's data position in Section 7.2. Key attributes of these international initiatives as well as Japan's principal infrastructure programs are compared in Table 3.

7.2. Japan's medical AI data infrastructure

Japan possesses several large-scale clinical databases that provide a foundation for medical AI research, though these have developed primarily for disease surveillance and pharmacovigilance rather than as AI-ready research platforms. The National Database of Health Insurance Claims and Specific Health Checkups (NDB) covers virtually the entire Japanese population through the universal health insurance system; while the scale of the claims data is unmatched, a systematic review of NDB-based research has identified access difficulties, issues with patient identification, and the lack of links to imaging or biomarkers as constraints on its utility for AI model development (*69*). The Medical Information Database Network (MID-NET), operated by the PMDA and linking hospital information systems at participating facilities to cover approximately 8.7 million patients—roughly 7% of Japan's population, in contrast to the NDB's near-universal coverage—was designed principally for post-marketing drug safety surveillance but represents a resource for AI-based research on pharmacovigilance and clinical outcomes (*70*).

Domain-specific registries further strengthen Japan's data position. The National Clinical Database (NCD), a surgical outcomes registry operated jointly by relevant surgical societies, had accumulated 28.48 million cumulative surgical cases as of 2023 (*71*). The JROAD, covering 1,500 participating facilities as of 2023, and the Japan Percutaneous Coronary Intervention Registry (J-PCI), with over 2.4 million cumulative cases, collectively represent one of the most complete national procedure-level cardiovascular datasets among OECD countries (*50,51*). These registries have supported observational and epidemiological research but have not yet been systematically capitalized upon as training or validation platforms for clinical AI models—a gap attributable to access governance designed for epidemiological use cases rather than AI development, the lack of links to imaging and molecular data, and the lack of any regulatory or institutional mandate requiring AI model validation to be conducted using these

Table 3. Principal international and Japanese initiatives on a data infrastructure for medical AI

Country	Initiative	Data type	Scale	Access model
United States	TCIA	Medical imaging	30.9M images, 37,568 subjects	Open access
United States	All of Us	EHR + genomic + wearable	> 873,000 participants	Controlled access
United States	Bridge2AI	Multi-modal (AI-ready)	$130M program	Controlled access
Europe	EHDS (2025)	EHR (federated)	Cross-border EU	Federated
United Kingdom	NHS AI Lab	Imaging + pathology	National platforms	Institutional
South Korea	AI Hub (NIA)	Annotated datasets	National repository	Controlled access
Japan	NDB	Claims + checkups	Near-universal coverage	Restricted
Japan	J-MID (JRS/AMED)	Medical imaging	543M images, 10 univ. hospitals	Controlled access
Japan	NCD/ ROAD/J-PCI	Surgical/cardiovascular	28.48M/1,500 inst./2.4M cases	Research access

Note: AMED, Japan Agency for Medical Research and Development; EHDS, European Health Data Space; EHR, electronic health record; J-MID, Japan Medical Imaging Database; J-PCI, Japan Percutaneous Coronary Intervention Registry; JROAD, Japan Registry of All Cardiac and Vascular Diseases; JRS, Japan Radiological Society; NCD, National Clinical Database; NDB, National Database of Health Insurance Claims and Specific Health Checkups; NHS, National Health Service; NIA, National Information Society Agency; TCIA, The Cancer Imaging Archive.

resources.

In the medical imaging domain, a high imaging volume and fragmented institutional data ownership means that Japan generates large quantities of diagnostic images and yet lacks the access infrastructure for competitive AI model development and multi-institutional validation—a gap identified by Ueda *et al*. and which prompted the J-MID initiative (*72*). The Japan-Medical Image Database (J-MID), a Japan Radiological Society (JRS) initiative supported by the Japan Agency for Medical Research and Development (AMED), addresses this gap by aggregating anonymized CT and MRI images with diagnostic reports from 10 major university hospitals *via* the SINET academic network; as of May 2024, J-MID had assembled over 543 million images (1.68 million cases), constituting what the project describes as an unparalleled repository of real-world radiological data in Japan (*73*).

The Next-generation Medical Infrastructure Act, promulgated in May 2023 and later amended, provides a new legal category of pseudonymized medical information, enabling certified data operators to link and process health records from multiple sources for research purposes under a unified governance regime (*74*). Progress on interoperability standards has been slower: the adoption of Japan-specific HL7 FHIR profiles for the exchange of medication and clinical data is still in at an early stage of implementation, which constrains the technical basis for cross-institutional data linkage that more mature federated research infrastructures require (*75*).

7.3. Comparative analysis: Strengths, gaps, and convergence

The NDB, MID-NET, and domain-specific registries represent population-wide or procedure-level datasets of a scale uncommon outside the United States, but access for research use is administratively difficult and the datasets lack the multimodal linkage—integrating imaging, genomic, and longitudinal clinical records—that characterizes the most capable international AI platforms.

The design philosophies of these initiatives differ materially: the United States combines open-access repositories with a federated infrastructure for sensitive records; the EHDS prioritizes federated-first data sovereignty; and Japan's framework under the Next-generation Medical Infrastructure Act is architecturally closer to the federated model, though a certified data operator ecosystem is still being established. Federated learning can approach centralized training performance when data distributions are similar across sites, but heterogeneous data quality and non-IID distributions remain active research challenges (*76*).

The competitive trajectory for Japan's medical AI will be materially shaped by three converging developments: the pace at which the Next-generation Medical Infrastructure Act framework generates research-accessible pseudonymized data pools, with only a handful of certified data operators having been designated as of early 2026; the extent to which the J-MID expands access beyond its current restricted institutional-partner model to support externally validated research comparable to that enabled by the TCIA, given that access remains limited to institutional partners despite the database having assembled over 543 million images from 10 major university hospitals; and the breadth of HL7 FHIR adoption enabling cross-institutional data linkage, currently in an early stage of implementation across Japan's hospital information systems. The clinical registries described above—the NCD, JROAD, and J-PCI— already provide a procedure-level evidence base that few countries can match and could become platforms for prospective AI validation studies if linked to imaging and molecular data under the governance framework the Next-generation Medical Infrastructure Act enables. Realizing this potential will require coordinated investment in data standardization, access infrastructure, and governance that goes beyond individual institutional initiatives and that matches the ambition of the Bridge2AI and EHDS programs described earlier.

8. Challenges and future directions

This section synthesizes the cross-cutting challenges that recur across the clinical domains reviewed earlier—other than data infrastructure and governance, addressed in Section 7—and it identifies the policy, methodological, and research developments needed to facilitate the responsible advancement of clinical AI.

8.1. Bridging the gap between performance metrics and clinical outcomes

A fundamental tension in clinical AI research is the reliance on intermediate performance metrics—the AUC, sensitivity, Dice similarity coefficient, and adenoma detection rate—as proxies for the patient-level outcomes that determine clinical value. The literature on AI-assisted imaging related to COVID-19, in which hundreds of studies reported high AUC values for retrospective data while none met the requirements for clinical deployment, illustrates the extent of this gap (*55*).

External validity is a related and persistent problem. Models trained at a single facility frequently exhibit a degradation in performance when applied to different scanners, patient populations, or clinical settings—as documented across multiple domains including the Japanese literature on AI detection of aneurysms (*38*). Prospective, multi-site evaluation studies are the methodological standard required to support regulatory authorization and health technology assessment (*12*).

International reporting frameworks—CONSORT-AI and SPIRIT-AI for clinical trials (*11,77*), TRIPOD+AI for prediction models (*12*), and STARD-AI for diagnostic accuracy studies (*13*) —provide the infrastructure for consistent, auditable evidence generation. Consistent application will raise the evidentiary floor for regulatory submissions and health technology assessments, though journal enforcement remains uneven.

8.2. Model updates, distribution shift, and post-marketing surveillance

AI systems deployed in clinical practice are subject to a distributional shift: changes in the patient population, clinical workflow, imaging equipment, or disease prevalence alter the input distribution relative to training data, gradually or abruptly diminishing model performance. AI to diagnose infectious diseases exemplifies this vulnerability through seasonal drift: the nodoca system to diagnose influenza requires re-evaluation as circulating strains and vaccine components change annually. A temporal drift in radiology, driven by scanner upgrades, acquisition protocol changes, or evolving diagnostic criteria, poses analogous challenges in that domain.

The United States FDA has addressed this challenge through the PCCP framework, which allows manufacturers to specify in advance the types of model updates—retraining on expanded data, threshold adjustments, or feature additions—that may be implemented without requiring a new regulatory submission, provided that the changes fall within pre-agreed performance bounds (*5*). In the European Union, the AI Act creates a "double compliance" burden for medical AI manufacturers—simultaneous MDR/IVDR conformity assessment alongside high-risk transparency and monitoring obligations pursuant to the AI Act (*6,29*).

In Japan, the regulatory review process for program medical devices that utilize machine learning has explicitly identified the management of continuous or post-approval performance changes as a key area requiring further policy development (*8,32*). Articulating a domestic framework for change management that is compatible with PCCP principles and enables seamless post-marketing surveillance through Japan's single-payer claims data infrastructure represents a priority policy task for the PMDA and the Ministry of Health, Labour, and Welfare (*78*). Accountability for adverse outcomes attributable to AI-assisted decisions also yet to be resolved: Japan's program medical device framework defines pre-marketing performance requirements but does not yet specify post-marketing liability norms for AI-assisted clinical decisions.

8.3. Reimbursement design and healthcare system integration

Regulatory authorization is a necessary but not sufficient condition for clinical AI adoption: the design of the reimbursement mechanism—whether AI-assisted procedures incur an additional fee, are assigned a separate billing code, or are subsumed into existing diagnostic fees—directly determines the economic incentive for healthcare providers to adopt and maintain AI systems. Japan's C1/C2 reimbursement classification system for new medical devices and technologies, exemplified by nodoca's C2 classification as a new technology fee, provides a formal pathway for health technology assessment of AI-based clinical decision-making support tools (*57*). However, the criteria for classification decisions, the evidence requirements for reclassification as standard care under insurance, and the mechanisms for incorporating evidence of real-world effectiveness into reimbursement still need to be fleshed out further for AI-specific applications.

Beyond reimbursement classification, healthcare system integration presents practical challenges that Japan has yet to systematically address. Alert fatigue from high false-positive rates and the workforce implications of AI integration represent cross-cutting policy challenges that the existing C1/C2 classification framework does not yet address (*1*).

8.4. Large language models and generative AI in clinical practice

Large language models (LLMs) and generative AI systems present a capability and risk profile distinct from the task-specific diagnostic models reviewed in the preceding sections. Benchmark evaluations have demonstrated that general-purpose LLMs—including GPT-4, which exceeded the passing threshold for the United States Medical Licensing Examination (USMLE) by over 20 points without domain-specific fine-tuning—possess substantial medical knowledge and reasoning ability (*79*). These results have prompted clinical interest in LLM applications including structured drafting of radiology reports and assisting with a differential diagnosis in electronic health records (*80*).

A 2025 systematic review of 15 studies that evaluated LLM-generated radiology reports found that while automated metrics and assessments by radiologists generally indicated acceptable linguistic quality, evidence from prospective implementation studies indicated that there were limited measurable improvements in clinical workflow efficiency or diagnostic accuracy (*81*). The principal risk specific to generative AI in clinical documentation is hallucination—the generation of plausible but factually incorrect content—which in clinical contexts can introduce erroneous findings, incorrect drug names, or fabricated examination results into a medical record in ways that are difficult to detect without systematic quality assurance processes. Unlike task-specific diagnostic AI, which typically produces a structured output (a probability score or bounding box) amenable to threshold-based review, LLM outputs are free-form text that have subtle and context-dependent errors.

A further challenge is regulatory classification. Many clinical LLM applications—and particularly those supporting documentation, summarization, and administrative workflow—do not meet the SaMD definition, as their primary function is not the direct generation of diagnostic or treatment outputs, placing them outside the PMDA program medical device framework in Japan and analogous regulatory perimeters in other jurisdictions. This regulatory gap means that LLM-based clinical tools can be deployed without pre-marketing clinical evaluation, post-marketing surveillance requirements, or change management obligations that apply to approved AI medical devices, creating asymmetric oversight that may underestimate population-level risk as adoption scales.

In Japan, the additional challenge of linguistic specificity is acute. Clinical documentation in Japanese involves a mix of kanji, kana, and medical abbreviations that are tokenized and semantically represented in models trained predominantly on English-language corpora, leading to substantial degradation relative to their native-language performance. Recent work has demonstrated that fine-tuned LLMs can achieve cross-hospital generalizability at recognition of diseases in clinical notes in Japanese, but deployment at scale across the heterogeneous documentation practices of Japan's hospital information systems remains an active research challenge (*82*). Developing and validating Japanese-language medical LLMs on pseudonymized data pools enabled by the Next-generation Medical Infrastructure Act represents a natural extension of the data infrastructure agenda discussed in Section 7. Progress in this area could have direct implications for clinical productivity and documentation quality across Japan's hospital system.

9. Conclusion

The evidence base across the five clinical domains examined in this review differs markedly. AI to assist diagnostic radiology has received regulatory approval at scale—accounting for 24 of Japan's 51 AI-designated program medical devices—but real-world deployment has revealed systematic gaps between controlled validation performance and operational accuracy, compounded in Japan by fragmented institutional data ownership that constrains the multi-institutional external validation needed to close that gap. AI to assist endoscopy has accumulated the most robust clinical trial evidence, driven in part by Japan's dual burden of colorectal and gastric malignancy that has produced 14 PMDA-approved devices and prompted Asian initiatives to conduct multicenter validation to expand the generalizability of existing RCT evidence—derived from predominantly East Asian and European populations—across the broader Asia-Pacific region. AI to assist cardiology has demonstrated the potential for ECG analysis and consumer wearables to facilitate population-level screening for arrhythmia while also exposing care pathway and reimbursement challenges that remain largely unaddressed in Japan despite a strong demographic demand from a population with the highest proportion of adults age 65 or older among OECD member countries. The infectious disease domain has produced Japan's most complete regulatory-approval-to-reimbursement pathway for a clinical AI tool, as exemplified by the nodoca system for diagnosis of influenza; it offers a reference for subsequent submissions involving AI-based clinical decision-making support. Comparative analysis of the AI data infrastructure has revealed that Japan's substantial clinical registry base—spanning cardiovascular procedures, surgical outcomes, and pharmacovigilance—is matched by few OECD countries at the procedure level, and yet a structural gap in accessible medical imaging data and multimodal linkage leaves Japan behind the data accessibility that programs such as TCIA and the EHDS framework afford international researchers. LLMs and generative AI systems, meanwhile, have introduced a category of clinical tool that largely falls outside existing SaMD regulatory frameworks, creating oversight asymmetries that healthcare systems in Japan and abroad

are only beginning to address.

Japan has a distinct position across all three comparative dimensions. In terms of its regulatory infrastructure, Japan's early legislative recognition of SaMD, the lists of approved program medical devices published by PMDA, and the complete regulatory-approval-to-reimbursement pathway demonstrated by nodoca represent a framework that is more operationally developed than that in many comparator countries. In clinical terms of the deployment infrastructure, Japan's density of diagnostic imaging systems and high endoscopy volume—reflected in the concentration of AI approvals within the imaging and endoscopy categories—provide a high-volume deployment base that few healthcare systems can match. In terms of the data infrastructure, however, a structural asymmetry persists: Japan's substantial clinical registries clash with limited data accessibility for external researchers and a medical imaging database in its early stages, in contrast to the open access of the TCIA or the cross-border federated governance of the EHDS. Sustained progress needs to be made across all three dimensions examined in this review—strengthening regulatory frameworks for post-approval model change management and clinical accountability, expanding the accessibility and scale of the AI-ready data infrastructure, and embedding validated AI tools into reimbursed clinical workflows—in order to realize the potential of clinical AI as a durable contributor to healthcare quality, efficiency, and equity in Japan and abroad.

Funding: This work was supported by JSPS KAKENHI Grant Number JP26K21655.

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Organisation for Economic Co-operation and Development. Artificial Intelligence and the health workforce. *https://www.oecd.org/en/publications/artificial-intelligence-and-the-health-workforce_9a31d8af-en.html* (accessed April 13, 2026).
2. Liu X, Faes L, Kale AU, *et al*. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. Lancet Digit Health. 2019; 1:e271-e297.
3. World Health Organization. Regulatory considerations on artificial intelligence for health. *https://www.who.int/publications/i/item/9789240078871* (accessed April 13, 2026).
4. International Medical Device Regulators Forum. Software as a medical device: Possible framework for risk categorization and corresponding considerations. *https://www.imdrf.org/documents/software-medical-device-possible-framework-risk-categorization-and-corresponding-considerations* (accessed April 13, 2026).
5. U.S. Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions. *https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence* (accessed April 13, 2026).
6. European Parliament and Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024. *https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng* (accessed April 13, 2026).
7. Medicines and Healthcare products Regulatory Agency. Software and artificial intelligence (AI) as a medical device. *https://www.gov.uk/government/publications/software-and-artificial-intelligence-ai-as-a-medical-device/software-and-artificial-intelligence-ai-as-a-medical-device* (accessed April 13, 2026).
8. Pharmaceuticals and Medical Devices Agency. Regarding the review of medical devices utilizing machine learning. *https://www.pmda.go.jp/files/000265866.pdf* (accessed April 13, 2026). (in Japanese)
9. National Medical Products Administration. NMPA Announcement on guidance for the classification defining of AI-based medical software products. *https://english.nmpa.gov.cn/2021-07/08/c_660267.htm* (accessed April 13, 2026).
10. Health Sciences Authority. Regulatory guidelines for software medical devices including machine learning-enabled medical devices – A Life Cycle Approach. *https://www.hsa.gov.sg/docs/default-source/hprg-mdb/guidance-documents-for-medical-devices/gl-04-r4-regulatory-guidelines-for-software-medical-devices---a-life-cycle-approach-(2025-dec)-pub.pdf* (accessed April 13, 2026).
11. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. BMJ. 2020; 370:m3164.
12. 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; 385:e078378.
13. Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SR, Golub RM, Denniston AK, Shetty S, Moher D, Bossuyt PM, Darzi A, Ashrafian H, STARD-AI Steering Committee. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025; 31:3283-3289.
14. Karako K, Gao J. Artificial intelligence (AI)-assisted ultrasound in clinical trials: Endpoint automation, decentralized monitoring, and regulatory readiness. Drug Discov Ther. 2026; 20:91-103.
15. Pharmaceuticals and Medical Devices Agency. List of approved program medical devices. *https://www.pmda.go.jp/review-services/drug-reviews/about-reviews/devices/0052.html* (accessed April 13, 2026). (in Japanese)
16. U.S. Food and Drug Administration. Artificial intelligence-Enabled Medical Devices. *https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices* (accessed April 13, 2026).
17. Organisation for Economic Co-operation and Development. Health at a Glance 2025: OECD indicators. *https://www.oecd.org/en/publications/health-at-a-glance-2025_8f9e3f98-en.html* (accessed April 13, 2026).
18. Organisation for Economic Co-operation and

Development. Health at a Glance 2025: Japan. *https://www.oecd.org/en/publications/health-at-a-glance-2025_15a55280-en/japan_319bfc39-en.html* (accessed April 13, 2026).
19. Karako K. Integration of wearable devices and deep learning: New possibilities for health management and disease prevention. Biosci Trends. 2024; 18:201-205.
20. McKinney SM, Sieniek M, Godbole V, *et al*. International evaluation of an AI system for breast cancer screening. Nature. 2020; 577:89-94.
21. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, Lei L, Li L, Guo Z, Lei S, Xiong F, Wang H, Song Y, Pan Y, Zhou G. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): A double-blind randomised study. Lancet Gastroenterol Hepatol. 2020; 5:343-351.
22. Repici A, Badalamenti M, Maselli R, *et al*. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology. 2020; 159:512-520.
23. Attia Z, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, Carter RE, Yao X, Rabinstein AA, Erickson BJ, Kapa S, Friedman PA. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. Lancet. 2019; 394:861-867.
24. Perez MV, Mahaffey KW, Hedlin H, *et al*. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med. 2019; 381:1909-1917.
25. Okiyama S, Fukuda M, Sode M, Takahashi W, Ikeda M, Kato H, Tsugawa Y, Iwagami M. Examining the use of an artificial intelligence model to diagnose influenza: Development and validation study. J Med Internet Res. 2022; 24:e38751.
26. International Medical Device Regulators Forum. Software as a Medical Device (SaMD): Clinical evaluation. *https://www.imdrf.org/documents/software-medical-device-samd-clinical-evaluation* (accessed April 13, 2026).
27. U.S. Food and Drug Administration and Health Canada and Medicines and Healthcare Products Regulatory Agency. Good machine learning practice for medical device development: Guiding principles. *https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles* (accessed April 13, 2026).
28. International Medical Device Regulators Forum. Good machine learning practice for medical device development: Guiding principles (IMDRF/AIML WG/N88 FINAL:2025). *https://www.imdrf.org/documents/good-machine-learning-practice-medical-device-development-guiding-principles* (accessed April 13, 2026).
29. Medical Device Coordination Group. MDCG 2025-6 Interplay between the Medical Devices Regulation (MDR) & In vitro Diagnostic Medical Devices Regulation (IVDR) and the Artificial Intelligence Act (AIA). *https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf* (accessed April 13, 2026).
30. National Institute for Health and Care Excellence. Evidence standards framework for digital health technologies. *https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies* (accessed April 13, 2026).
31. Ministry of Food and Drug Safety, Republic of Korea. Regulatory update on medical devices in the Republic of Korea. *https://www.imdrf.org/sites/default/files/2025-09/Korea%20%28MFDS%20MC%29.pdf* (accessed April 13, 2026).
32. Ministry of Health, Labour, and Welfare, Japan. Responses to the regulatory treatment of medical devices using AI under the PMD Act. *https://www.mhlw.go.jp/content/10601000/000361102.pdf* (accessed April 13, 2026). (in Japanese)
33. Ministry of Health, Labour, and Welfare, Japan. Approval status of AI-enabled medical device programs. p.12. *https://www.mhlw.go.jp/content/11124500/001329696.pdf* (accessed April 13, 2026). (in Japanese)
34. Cheng PM, Montagnon E, Yamashita R, Pan I, Cadrin-Chênevert A, Perdigón Romero F, Chartrand G, Kadoury S. Deep learning: An update for radiologists. Radiographics. 2021; 41:1427-1445.
35. Martinez-Gutierrez JC, Kim Y, Salazar-Marioni S, *et al*. Automated large vessel occlusion detection software and thrombectomy treatment times: A cluster randomized clinical trial. JAMA Neurol. 2023; 80:1182-1190.
36. Yu AC, Mohajer B, Eng J. External validation of deep learning algorithms for radiologic diagnosis: A systematic review. Radiol Artif Intell. 2022; 4:e210064.
37. Ishihara M, Shiiba M, Maruno H, Kato M, Ohmoto-Sekine Y, Antoine C, Ouchi Y. Detection of intracranial aneurysms using deep learning-based CAD system: Usefulness of the scores of CNN's final layer for distinguishing between aneurysm and infundibular dilatation. Jpn J Radiol. 2023; 41:131-141.
38. Ito R, Asai R, Nakamichi R, Nakane T, Taoka T, Naganawa S. Evaluation of approved AI-based brain aneurysm detection software in clinical practice: Comparison with radiologist assessment and image re-review. Magn Reson Med Sci. 2026; 25.
39. Makar J, Abdelmalak J, Con D, Hafeez B, Garg M. Use of artificial intelligence improves colonoscopy performance in adenoma detection: A systematic review and meta-analysis. Gastrointest Endosc. 2025; 101:68-81.
40. Misawa M, Kudo S, Mori Y. Implementation of artificial intelligence in colonoscopy practice in Japan. JMA J. 2025; 8:60-63.
41. Houwen BBSL, Hazewinkel Y, Giotis I, Vleugels JLA, Mostafavi NS, van Putten P, Fockens P, Dekker E. Computer-aided diagnosis for optical diagnosis of diminutive colorectal polyps including sessile serrated lesions: A real-time comparison with screening endoscopists. Endoscopy. 2023; 55:756-765.
42. World Cancer Research Fund. Colorectal cancer statistics. *https://www.wcrf.org/cancer-trends/colorectal-cancer-statistics/* (accessed April 13, 2026).
43. World Cancer Research Fund. Stomach cancer statistics. *https://www.wcrf.org/cancer-trends/stomach-cancer-statistics/* (accessed April 13, 2026).
44. Kudo SE, Misawa M, Mori Y, *et al*. Artificial intelligence-assisted system improves endoscopic identification of colorectal neoplasms. Clin Gastroenterol Hepatol. 2020; 18:1874-1881.
45. University Hospital Medical Information Network Clinical Trials Registry. Clinical trial registration: EndoBrain multicenter study (UMIN000047839). *https://center6.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000055293* (accessed April 13, 2026). (in

Japanese)
46. National Cancer Center Japan. An Asian multicenter clinical trial evaluating efficacy of computer-aided detection for colonoscopy in colorectal cancer screening. *https://www.ncc.go.jp/en/information/press_release/2024/0111/index.html* (accessed April 13, 2026).
47. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, Pellikka PA, Enriquez-Sarano M, Noseworthy PA, Munger TM, Asirvatham SJ, Scott CG, Carter RE, Friedman PA. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. Nat Med. 2019; 25:70-74.
48. Yagi R, Goto S, Katsumata Y, MacRae CA, Deo RC. Importance of external validation and subgroup analysis of artificial intelligence in the detection of low ejection fraction from electrocardiograms. Eur Heart J Digit Health. 2022; 3:654-657.
49. Masumura M, Ohno A, Yoshinaga H, Sasaki T, Yamauchi Y, Hachiya H, Takahashi A, Imai Y, Fujita H, Ihara K, Ebana Y, Tanaka T, Furukawa T, Sasano T. AI-ECG for early detection of atrial fibrillation: First-year results from a stroke prevention study in Shimizu, Japan. J Arrhythm. 2025; 41:e70132.
50. Japanese Circulation Society. JROAD (Japan Registry of All Cardiac and Vascular Diseases) 2024 Annual Report. *https://www.j-circ.or.jp/jittai_chosa/about/report/* (accessed April 13, 2026).
51. Japanese Association for Cardiovascular Intervention and Therapeutics. J-PCI Registry 2024 Annual Report. *https://www.cvit.jp/docs/registry/annual-report/j-pci/2024.pdf* (accessed April 13, 2026).
52. Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, Pagoto S, McConnell MV, Pantelopoulos A, Foulkes AS. Detection of atrial fibrillation in a large population using wearable devices: The Fitbit Heart Study. Circulation. 2022; 146:1415-1424.
53. Simonson JK, Anderson M, Polacek C, Klump E, Haque SN. Characterizing real-world implementation of consumer wearables for the detection of undiagnosed atrial fibrillation in clinical practice: Targeted literature review. JMIR Cardio. 2023; 7:e47292.
54. Nomura A. Digital health, digital medicine, and digital therapeutics in cardiology: Current evidence and future perspective in Japan. Hypertens Res. 2023; 46:2126-2134.
55. Roberts M, Driggs D, Thorpe M, *et al*. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell. 2021; 3:199-217.
56. Pharmaceuticals and Medical Devices Agency. Review report: Nodoca (software medical device to aid in the differential diagnosis of influenza). *https://www.pmda.go.jp/files/000248783.pdf* (accessed April 13, 2026). (in Japanese)
57. Central Social Insurance Medical Council. Health insurance coverage for medical devices (December 2022): Nodoca. *https://www.mhlw.go.jp/content/12404000/000989561.pdf* (accessed April 13, 2026). (in Japanese)
58. Han ZL, Zhang YY, Li J, Gao S, Liu W, Yang WJ, Xing ZH. A systematic review and meta-analysis of artificial intelligence software for tuberculosis diagnosis using chest X-ray imaging. J Thorac Dis. 2025; 17:3223-3237.
59. Rudd KE, Johnson SC, Agesa KM, *et al*. Global, regional, and national sepsis incidence and mortality, 1990-2017: Analysis for the Global Burden of Disease Study 2017. Lancet. 2020; 395:200-211.
60. World Health Organization. Sepsis. *https://www.who.int/news-room/fact-sheets/detail/sepsis* (accessed April 13, 2026).
61. Bhargava A, López-Espina C, Schmalz L, *et al*. FDA-authorized AI/ML tool for sepsis prediction: Development and validation. NEJM AI. 2024; 1:AIoa2400867.
62. National Cancer Institute. The Cancer Genome Atlas Program (TCGA). *https://www.cancer.gov/ccg/research/genome-sequencing/tcga* (accessed April 13, 2026).
63. Prior F, Smith K, Sharma A, Kirby J, Tarbox L, Clark K, Bennett W, Nolan T, Freymann J. The public cancer radiology imaging collections of The Cancer Imaging Archive. Sci Data. 2017; 4:170124.
64. National Institutes of Health. BRIDGE2AI – Propelling biomedical research with artificial intelligence. *https://bridge2ai.org/* (accessed April 13, 2026).
65. NIH All of Us Research Program. The "All of Us" research program. *https://allofus.nih.gov/* (accessed April 13, 2026).
66. European Commission. European Health Data Space Regulation (EHDS). *https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-regulation-ehds_en* (accessed April 13, 2026).
67. Cresswell K, Williams R, Dungey S, Anderson S, Bernabeu MO, Mozaffar H, Yang X, Sai V, Bea S, Eason S. A mixed methods formative evaluation of the United Kingdom National Health Service Artificial Intelligence Lab. npj Digit Med. 2025; 8:448.
68. National Information Society Agency. AI Hub: Integrated AI training data platform. *https://www.aihub.or.kr/* (accessed April 13, 2026).
69. Suto M, Iba A, Sugiyama T, Kodama T, Takegami M, Taguchi R, Niino M, Koizumi R, Kashiwagi K, Imai K, Ihana-Sugiyama N, Ichinose Y, Takehara K, Iso H. Literature review of studies using the National Database of the Health Insurance Claims of Japan (NDB): Limitations and strategies in using the NDB for research. JMA J. 2024; 7:10-20.
70. Pharmaceuticals and Medical Devices Agency. Understanding MID-NET®: Its features and efforts to promote its use in academia. *https://www.pmda.go.jp/files/000267996.pdf* (accessed April 13, 2026). (in Japanese)
71. Yamamoto T, Takahashi A, Yoshizumi T, *et al*. 2023 National clinical database annual report by the Japan surgical society. Surgery Today. 2025; 55:295-334.
72. Ueda D, Walston S, Takita H, Mitsuyama Y, Miki Y. The critical need for an open medical imaging database in Japan: Implications for global health and AI development. Jpn J Radiol. 2025; 43:537-541.
73. Akashi T, Kumamaru KK, Wada A, Hashimoto M, Hirata K, Hayakawa Y, Sano K, Kamagata K, Hagiwara A, Ikenouchi Y, Aoki S. Japan-Medical Image Database (J-MID): Medical big data supporting data science. Juntendo Med J. 2025; 71:166-172.
74. Ministry of Health, Labour, and Welfare, Japan. Regarding the amended Next-generation Medical Infrastructure Act. *https://www.mhlw.go.jp/content/10808000/001166476.pdf* (accessed April 13, 2026). (in Japanese)
75. Kobayashi S, Kimura M, Kodama Y, Takada A, Nagashima S, Kawazoe Y, Ohe K. Development of Japan-specific HL7 FHIR medication-related profiles. J Med Syst. 2025; 49:46.
76. Teo ZL, Jin L, Li S, Miao D, Zhang X, Ng WY, Tan TF, Lee DM, Chua KJ, Heng J, Liu Y, Goh RSM, Ting DSW.

Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. Cell Rep Med. 2024; 5:101419.
77. Cruz Rivera S, Liu X, Chan AW, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: The SPIRIT-AI extension. BMJ. 2020; 370:m3210.
78. Ministry of Health, Labour, and Welfare, Japan. Plan to promote investment in reduced Labour: Medical care. *https://www.mhlw.go.jp/content/10801000/001565769.pdf* (accessed April 13, 2026). (in Japanese)
79. Nori H, King N, McKinney SM, Carignan D, Horvitz E. Capabilities of GPT-4 on medical challenge problems. arXiv preprint arXiv:2303.13375 (2023).
80. Karako K, Song P, Chen Y, Tang W. New possibilities for medical support systems utilizing artificial intelligence (AI) and data platforms. Biosci Trends. 2023 Jul 11; 17(3):186-189.
81. Artsi Y, Klang E, Collins JD, Glicksberg BS, Nadkarni GN, Korfiatis P, Sorin V. Large language models in radiology reporting — A systematic review of performance, limitations, and clinical implications. Intell Based Med. 2025; 12:100287.
82. Shimizu S, Nishiyama T, Nagai H, Wakamiya S, Aramaki E. Toward cross-hospital deployment of natural language processing systems: Model development and validation of fine-tuned large language models for disease name recognition in Japanese. JMIR Med Inform. 2025; 13:e76773.

Received March 27, 2026; Revised May 4, 2026; Accepted May 9, 2026.

Released online in J-STAGE as advance publication May 13, 2026.

**Address correspondence to*:
Kenji Karako, Department of Surgery, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8655, Japan.
E-mail: tri.leafs@gmail.com