Contemporary Studies in Applied Sciences; Vol. 1, No. 1; 2024
ISSN 2810-9953
Published by July Press

Machine Learning in Biomedical Engineering and Medical Imaging: Algorithms, Validation, and Clinical Translation

Jianhua Lan[1]

[1] College of Information and Electrical Engineering, China Agricultural University, Beijing, China

Correspondence: Jianhua Lan, College of Information and Electrical Engineering, China Agricultural University, Beijing, China.

Received: September 21, 2024 Accepted: November 9, 2024 Online Published: December 4, 2024

Abstract

Machine learning has become a central pillar of modern biomedical engineering, reshaping how medical images, biosignals, and wearable data are acquired, processed, and translated into clinical decisions. This review synthesises the current state of machine learning in biomedical engineering, with particular emphasis on medical imaging, from algorithmic foundations to regulatory clearance and clinical adoption. The analysis is evidence-based and integrative, drawing on peer-reviewed literature, regulatory registries, and recent empirical evaluations. The results show that the application of machine learning in biomedicine has expanded rapidly: the U.S. Food and Drug Administration authorized a record number of artificial intelligence-enabled medical devices in recent years, with cumulative authorizations exceeding 1,400 devices by the end of 2025, and with radiology accounting for roughly three-quarters of all clearances. Within medical imaging, deep convolutional networks and, increasingly, transformer-based architectures have achieved diagnostic performance comparable to that of expert clinicians across a range of modalities, including chest radiography, computed tomography, and magnetic resonance imaging. At the same time, the translation of these algorithms into routine practice remains uneven, constrained by limited data granularity, imperfect generalizability across populations and devices, fragmented regulatory pathways, and the practical demands of clinical workflow integration. The review identifies the principal technical, data, regulatory, and human-factor barriers to clinical translation, and proposes a set of recommendations for biomedical engineers, clinicians, and regulators aimed at closing the gap between algorithmic performance and real-world benefit. The findings underscore that successful integration of machine learning in biomedical engineering depends less on raw model accuracy than on robust validation, transparent reporting, interdisciplinary collaboration, and the deliberate design of systems that support, rather than replace, clinical judgement.

Keywords: machine learning, biomedical engineering, medical imaging, deep learning, clinical translation, regulatory approval, artificial intelligence, validation

1. Introduction

Biomedical engineering is the discipline that applies engineering principles and design concepts to medicine and biology for healthcare purposes. It spans a vast and heterogeneous territory, from diagnostic imaging systems and physiological sensors to prosthetics, therapeutic devices, and computational models of physiological processes. Over the past decade, machine learning has moved from a peripheral research tool to a defining capability within this field, enabling systems that can interpret medical images, decode neural signals, monitor physiological state continuously, and support clinical decision-making with increasing sophistication. The convergence of large, curated datasets, advances in deep learning architectures, and growth in affordable computing has accelerated this transformation, making machine learning a signature technology of contemporary biomedical engineering (Revolutionizing radiology, 2024; AI in COPD CT images, 2024).

The most visible manifestation of this development is in medical imaging, where deep learning models have demonstrated the capacity to detect, segment, and classify pathologies across chest radiography, computed tomography, magnetic resonance imaging, and ultrasound. In many tasks, these models achieve diagnostic performance comparable to, and in some cases exceeding, that of expert human readers (Deep learning approaches for classification, 2024). Beyond imaging, machine learning is embedded in wearable biosensors that monitor heart rhythm, blood pressure, and metabolic state, and in brain-computer interfaces that translate neural activity into control signals for assistive technology (Frontiers of wearable biosensors, 2023; Non-invasive brain-computer

interfaces, 2024). The breadth of these applications reflects a fundamental reframing of biomedical engineering: intelligent algorithms have become as important to the field as the hardware that collects and delivers physiological information.

Yet the pace of algorithmic progress has outpaced the capacity of the surrounding infrastructure to absorb it. Regulatory frameworks have had to adapt to a new class of products, namely software as a medical device, that changes rapidly and cannot be evaluated like a conventional fixed-hardware device (FDA regulation of clinical software, 2024). Clinical workflows, built around the judgment and accountability of human experts, have struggled to integrate algorithmic outputs in ways that genuinely improve outcomes rather than merely adding workload (Navigating the AI landscape in medical imaging, 2025). And the datasets on which models are trained, often drawn from single institutions and homogeneous populations, raise persistent questions about generalizability and bias (Future-proofing medical imaging, 2024).

This review addresses the state of machine learning in biomedical engineering and medical imaging, tracing the arc from algorithm design to clinical adoption. It has four objectives. First, it synthesises the principal technical developments, including convolutional and transformer-based architectures for image analysis and the expansion of machine learning into wearable and neural applications. Second, it documents, with quantitative evidence, the scale and trajectory of regulatory clearance of AI-enabled medical devices. Third, it examines the evidence on model performance across imaging tasks and modalities. Fourth, it analyses the barriers to clinical translation and proposes practical recommendations. The paper is structured as follows. Section 2 reviews the literature on machine learning in biomedical engineering. Section 3 describes the analytical approach and data sources. Section 4 presents the results, supported by figures and tables. Section 5 discusses the findings and their implications. Section 6 concludes with recommendations.

A note on terminology and scope is warranted. Throughout this review, the terms artificial intelligence and machine learning are used to refer to computational methods that learn patterns from data rather than relying on explicitly programmed rules, encompassing classical statistical learning, deep neural networks, and, more recently, foundation and generative models. The focus is on applications that directly support diagnosis, monitoring, and therapy, and on the translational pathway that connects these applications to clinical value. Biomedical engineering is understood broadly, to include the design of imaging and sensing hardware, the development of diagnostic algorithms, and the systems engineering required to deploy them safely and effectively. This broad framing reflects the reality that the field is no longer neatly divisible into hardware and software components, but increasingly operates at the interface between them, where algorithms, data, and clinical practice converge. The review therefore speaks to a multidisciplinary audience, including biomedical engineers concerned with system design, clinical scientists evaluating diagnostic tools, and regulators charged with ensuring that these powerful technologies are introduced safely, transparently, and with demonstrable benefit to patients.

2. Literature Review

2.1 The Biomedical Engineering Landscape and Machine Learning

Biomedical engineering has historically been organised around the measurement and manipulation of physiological signals, the design of medical devices, and the reconstruction and visualisation of anatomy. Machine learning enters this landscape as a general tool for turning high-dimensional, noisy biomedical data into reliable predictions and decisions. Its utility derives from the observation that many biomedical problems, such as detecting a pulmonary nodule in a chest radiograph or classifying an arrhythmia from an electrocardiogram, are fundamentally pattern-recognition tasks that can be learned from examples rather than programmed by rules (Revolutionizing radiology, 2024).

The adoption of machine learning in biomedicine has been supported by three enabling developments. The first is the availability of large annotated datasets, many of which have been made public for research, including chest radiograph collections and physiological signal repositories that provide the volume of labelled examples modern deep networks require. The second is the maturation of deep learning architectures capable of learning directly from raw or lightly preprocessed data, obviating the need for hand-crafted feature engineering that characterised earlier generations of machine learning in medicine. The third is the growth of accessible computational infrastructure, including graphics processing units and cloud platforms, which has made the training of large models feasible for academic and clinical groups. Together these developments have lowered the barriers to entry and widened the range of problems to which machine learning can be applied.

It is useful to distinguish the complementary roles of machine learning and traditional biomedical engineering methods. Conventional engineering approaches in biomedicine, such as signal processing, physiological modelling, and image reconstruction, rest on explicit physical and mathematical principles, and their outputs are

generally interpretable and governed by well-understood theory. Machine learning, by contrast, derives its behaviour from patterns learned in data, which allows it to capture relationships that are too complex or too poorly understood to be encoded explicitly. In practice, the most effective biomedical systems combine both modes: physics-based models provide structure and constraints, while learned components capture residual complexity and adapt to the idiosyncrasies of real clinical data. This hybrid perspective, in which machine learning augments rather than replaces established engineering methods, is increasingly recognised as the most productive direction for the field (Deep learning approaches for classification, 2024).

2.2 Deep Learning Architectures for Medical Image Analysis

Convolutional neural networks (CNNs) have been the workhorse of medical image analysis since the introduction of the U-Net architecture for biomedical image segmentation in 2015. The U-Net's symmetric encoder-decoder design, connected by skip connections that preserve spatial detail, proved remarkably effective for segmenting anatomical structures and lesions from relatively few annotated images, and it remains a foundational component of countless biomedical imaging systems (Ronneberger et al., 2015). Subsequent refinements, including residual connections, attention modules, and multi-scale feature extraction, have extended the capacity and robustness of convolutional architectures for increasingly complex segmentation and classification tasks (TAC-UNet, 2024).

More recently, transformer-based architectures, originally developed for natural language processing, have been adapted to medical imaging. Transformers model long-range dependencies and global context through self-attention, addressing a limitation of convolutional networks, whose receptive fields are intrinsically local. Hybrid architectures that combine convolutional and transformer components, such as transformer-assisted U-Nets and cross-shaped-window transformers, aim to capture both local detail and global structure, and have reported improved performance on segmentation benchmarks (TAC-UNet, 2024; CSWin-UNet, 2024). The shift from convolutional to transformer-based and hybrid models represents an active frontier in medical image analysis, although these models typically require larger training corpora and greater computational resources (From CNN to transformer, 2024).

Beyond supervised learning, self-supervised and foundation models have emerged as promising approaches for learning useful representations from large quantities of unlabelled medical data. By pretraining on massive collections of images and then fine-tuning on smaller, labelled clinical cohorts, these models can achieve strong performance even when task-specific labels are scarce, an attractive property in biomedicine where expert annotation is costly and time-consuming (AI in radiology: where are we going?, 2024).

2.3 Beyond Imaging: Wearables, Biosignals, and Brain-Computer Interfaces

Machine learning in biomedical engineering extends well beyond medical imaging. Wearable biosensors now generate continuous streams of physiological data, including heart rate, electrocardiographic signals, blood pressure, and metabolic markers, which machine learning models interpret to support remote patient monitoring, early diagnosis, and personalised medicine (Frontiers of wearable biosensors, 2023). The combination of low-cost sensors and intelligent algorithms has the potential to shift healthcare from episodic, institution-based encounters toward continuous, home-based monitoring, although challenges of clinical validation, signal quality, and interpretability remain substantial (Wearable technologies for healthcare, 2025; AI-driven wireless sensing, 2025).

Brain-computer interfaces (BCIs) illustrate the depth of the machine-learning contribution. Intracortical BCIs decode neural signals recorded from implanted microelectrode arrays to control external devices, enabling communication and movement restoration for individuals with severe neurological impairment (Neural decoding for intracortical BCIs, 2023). The decoding algorithms at the heart of these systems are machine-learning models trained to translate high-dimensional neural activity into intended actions, and advances in electrode materials and interface design have improved the stability and fidelity of the recorded signals on which these models depend (Advances in electrode interface materials, 2024). Non-invasive BCIs, which rely on electroencephalographic or other surface signals, trade spatial resolution for safety and accessibility, and remain an active area of research and development (Non-invasive brain-computer interfaces, 2024).

2.4 Data, Benchmarks, and the Imperative of Validation

A recurring theme across the literature is the centrality of data quality and validation. Machine learning models are only as reliable as the data on which they are trained and the rigour with which they are evaluated. Public benchmarks, such as the chest radiograph datasets used widely in imaging research, have enabled progress but also carry limitations, including label quality issues, domain shift between institutions, and potential biases that can inflate apparent performance (Deep learning approaches for classification, 2024; Limitations of public chest radiography datasets, 2024). The emerging consensus is that algorithmic performance on a research benchmark is

necessary but not sufficient for clinical utility, and that prospective, multi-site evaluation in realistic workflow conditions is required.

Privacy and data-sharing constraints compound the validation challenge. Because medical data are highly sensitive and heavily regulated, aggregating large datasets across institutions for training and validation is often infeasible. Federated learning has emerged as a response, enabling multiple institutions to train a shared model without exchanging raw patient data, although it introduces its own challenges around communication cost, heterogeneity, and security (Privacy preserving federated learning, 2024; Future-proofing medical imaging, 2024; Federated learning for secure neuroimaging, 2024). The literature converges on the view that robust, privacy-preserving, and externally validated machine learning is the precondition for responsible clinical translation.

A further strand of the literature addresses the gap between algorithmic research and clinical deployment. Studies of AI implementation in radiology report that the practical integration of these tools into routine workflows is frequently more difficult than their development, owing to narrow functionality, fragmented vendor solutions, and the absence of standard interfaces for incorporating algorithmic outputs into reporting systems (Challenges of implementing AI-enabled programs, 2024; A holistic approach to implementing AI in radiology, 2024). The literature increasingly frames successful translation as a systems-engineering problem, in which the algorithm, the surrounding clinical process, the users, and the governance structure must be designed together if the technology is to deliver genuine benefit rather than additional workload (Navigating the AI landscape in medical imaging, 2025). This framing, which unites technical, organisational, and human-factor considerations, informs the analytical approach and the recommendations developed in the remainder of this review.

3. Methods

3.1 Analytical Approach and Scope

This review is synthetic and evidence-based, integrating peer-reviewed literature, regulatory registry data, and recent empirical evaluations to characterise the state of machine learning in biomedical engineering and medical imaging. Rather than presenting a single original experiment, the paper draws together the quantitative findings of the studies and registries that have documented the most rigorous assessments of algorithmic performance, regulatory clearance, and clinical adoption, and compares their methods and results to identify robust conclusions and remaining gaps.

The scope is deliberately broad but structured. The review focuses on machine learning applied to medical imaging as its core evidence base, given the maturity and quantitative richness of that literature, and draws complementary evidence from wearable sensing and brain-computer interfaces to illustrate the wider reach of the technology within biomedical engineering. The analytical lens is translational: the organising question is not whether machine learning can achieve high accuracy in controlled settings, but whether and how it moves from algorithm to validated, adopted, and beneficial clinical practice.

3.2 Data Sources and Metrics

The principal quantitative data sources are as follows. The scale and trajectory of regulatory clearance are quantified using the registry of artificial intelligence and machine learning-enabled medical devices maintained by the U.S. Food and Drug Administration, as compiled and analysed by independent trackers and academic studies (FDA AI/ML-enabled device registry; Machine learning-enabled devices authorized by the FDA in 2024, 2025; Evaluating transparency in AI/ML model characteristics, 2024). These sources report annual clearances, cumulative authorizations, and the distribution of devices across clinical specialties, providing the basis for the regulatory figures in Section 4.

Model performance is assessed using peer-reviewed evaluations and systematic reviews that report quantitative diagnostic metrics across imaging modalities and tasks, including sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve. Representative data include the reported accuracy of deep learning models for pulmonary nodule detection on computed tomography and for brain tumour classification on magnetic resonance imaging (Advancements in AI for automated radiological diagnosis, 2024), as well as the performance of large language models on radiological differential-diagnosis tasks (GPT-based differential diagnostic accuracy, 2025). Clinical adoption is characterised using studies that examine the gap between regulatory approval and real-world use (Navigating the AI landscape in medical imaging, 2025; Real-world adoption of AI in radiology, 2024).

3.3 Synthesis and Comparison Procedure

The synthesis proceeds in four steps. First, regulatory registry data are summarised to establish the scale and growth of AI-enabled medical devices, and the concentration of clearances across clinical specialties. Second,

performance data from peer-reviewed evaluations are tabulated across imaging modalities and tasks to establish the current level of algorithmic capability. Third, qualitative evidence on implementation barriers is integrated to characterise the translation gap. Fourth, the results are compared across studies to identify robust findings and areas of uncertainty.

The methodology acknowledges several limitations. First, the reported figures are drawn from sources that apply slightly different inclusion criteria, so annual clearance counts may not sum exactly to cumulative totals. Second, diagnostic performance metrics are not directly comparable across studies because of differences in datasets, populations, and reference standards. Third, the analysis is necessarily selective, prioritising representative and well-documented evidence over exhaustive coverage. These limitations are themselves informative, insofar as they highlight the heterogeneity and methodological variability that characterise the field.

The choice of a synthetic, review-based methodology deserves explicit justification. The field of machine learning in biomedical engineering advances so rapidly, and across so many dispersed subfields, that a single institution cannot typically command the data, expertise, and clinical context needed to draw robust conclusions from one experiment. A synthetic approach, by consolidating the findings of the most rigorous and recent studies, offers three advantages. It provides a breadth of evidence that no individual experiment could match; it enables comparison across heterogeneous settings that reveals which findings are robust and which are setting-specific; and it surfaces the gaps and inconsistencies in the literature that define the frontier of unresolved questions. These advantages are particularly valuable for a translational review, whose purpose is not to establish a single new result but to inform engineering practice and policy decisions across an entire field (A decade of review in global regulation, 2025).

4. Results

This section presents the quantitative evidence on the growth, concentration, and performance of machine learning in biomedical engineering and medical imaging, together with evidence on the translation of algorithmic advances into clinical practice.

4.1 Growth and Regulatory Clearance of AI-Enabled Devices

The most striking result is the scale and pace of growth in the number of AI-enabled medical devices reaching the market. According to the U.S. Food and Drug Administration registry and independent trackers, annual clearances of AI/ML-enabled devices have risen steeply in recent years, from about 215 devices in 2022 to 221 in 2023, 253 in 2024, and a record 295 in 2025 (FDA AI/ML-enabled device registry; Machine learning-enabled devices authorized by the FDA in 2024, 2025). This represents roughly a forty-nine-fold expansion in annual clearances over a ten-year window, from approximately six devices in 2015 to 295 in 2025. Figure 1 visualises this annual trajectory.

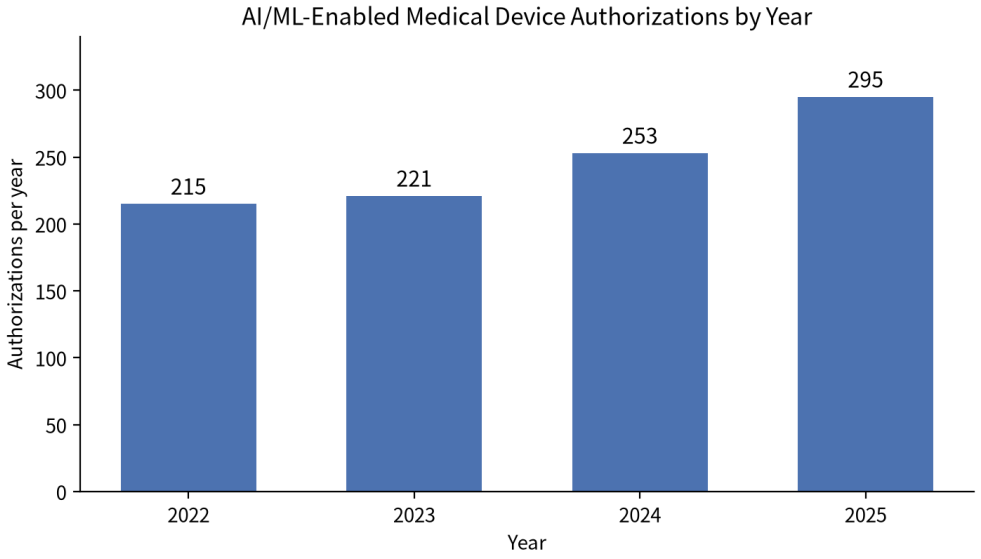


Figure 1. Annual authorizations of AI/ML-enabled medical devices in the United States, 2022-2025

Cumulative authorizations have grown correspondingly. The FDA had approved or cleared more than 600 AI/ML-enabled devices by the end of 2022, over 860 by the end of 2023, roughly 1,016 by the end of 2024, and about 1,451 by the end of 2025 (FDA AI/ML-enabled device registry; Evaluating transparency in AI/ML model characteristics, 2024). Figure 2 charts this cumulative trajectory. The overwhelming majority of devices have been cleared through the 510(k) pathway, relying on substantial equivalence to predicate devices, with a smaller share authorised through the De Novo pathway (Machine learning-enabled devices authorized by the FDA in 2024, 2025).

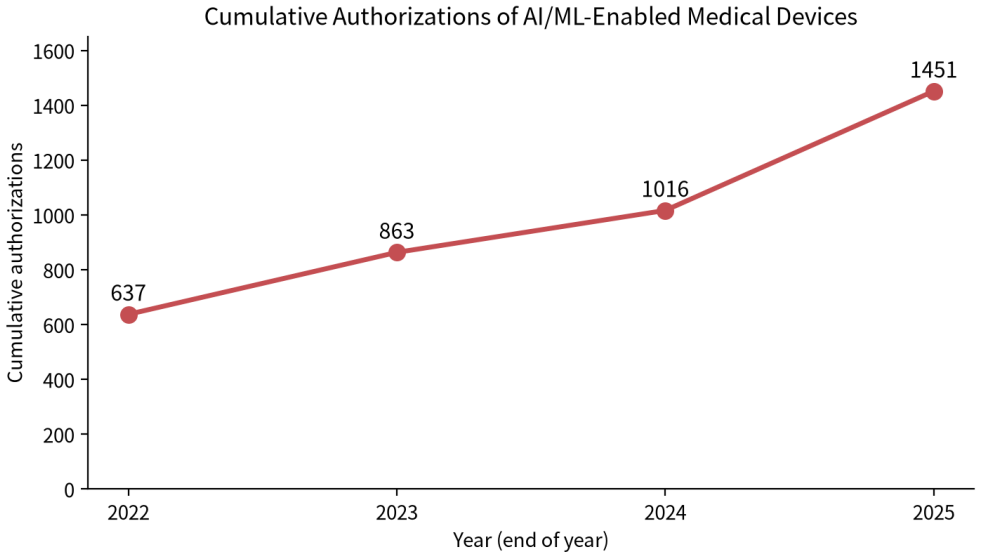


Figure 2. Cumulative authorizations of AI/ML-enabled medical devices in the United States, 2022-2025

4.2 Clinical Specialty Concentration

The distribution of AI-enabled medical devices across clinical specialties is highly concentrated. Radiology is dominant, accounting for roughly three-quarters of all cumulative AI/ML clearances, with 1,104 of 1,451 devices by the end of 2025 falling under the radiology panel (FDA AI/ML-enabled device registry). Other specialties account for much smaller shares. Table 1 summarises the concentration of authorizations across major clinical specialties.

Table 1. Distribution of AI/ML-enabled medical device authorizations by clinical specialty (cumulative through 2025)

Clinical Specialty	**Approximate Authorizations**	**Share of Total**
Radiology	1,104	76.1%
Cardiovascular	~94	6.5%
Neurology	~87	6.0%
Other specialties	~166	11.4%
Total	1,451	100.0%

This concentration reflects the natural fit between machine learning and image-based diagnosis, where large labelled datasets, well-defined detection and classification tasks, and quantitative evaluation metrics are readily available. It also implies that the benefits of machine learning in biomedicine are, at present, distributed unevenly across clinical domains, with imaging applications far ahead of other areas in terms of regulatory maturity and market presence (Revolutionizing radiology, 2024).

The heavy concentration in radiology is not merely a reflection of where the technology happens to be developed; it reflects structural properties of imaging as a clinical task. Images are digital, standardised, and abundant; pathologies often have recognisable visual signatures; and the ground truth for validation, while imperfect, is substantially better defined than in many other clinical domains. These properties make imaging the natural early proving ground for medical machine learning. By the same logic, the relatively smaller shares in cardiovascular and neurological devices reflect not a lesser potential for machine learning in those fields but greater heterogeneity in signals, more demanding safety requirements, and, in many cases, a longer path to the kind of large, well-labelled datasets that imaging benefits from. As these obstacles are progressively addressed, the distribution of clearances is likely to broaden toward a more balanced pattern across specialties (A decade of review in global regulation, 2025).

4.3 Model Performance Across Tasks and Modalities

Evidence from peer-reviewed evaluations indicates that deep learning models achieve high diagnostic performance across a range of imaging tasks and modalities. Systematic reviews report accuracies in the range of 92 to 95 percent for pulmonary nodule detection on computed tomography, comparable to expert radiologists, and 88 to 93 percent for brain tumour classification on magnetic resonance imaging (Advancements in AI for automated radiological diagnosis, 2024). In chest radiography, deep learning models have demonstrated the ability to detect a wide spectrum of pathologies, including pneumonia, pulmonary nodules, tuberculosis, and interstitial lung disease, with performance frequently matching or exceeding human readers (Revolutionizing radiology, 2024). Table 2 summarises representative reported performance across modalities and tasks.

Table 2. Representative reported diagnostic performance of machine learning models by imaging modality and task

Imaging Modality	**Common Application**	**Reported Accuracy**
Computed tomography (CT)	Pulmonary nodule detection	92-95%
Magnetic resonance imaging (MRI)	Brain tumour classification	88-93%
Chest radiography (X-ray)	Pneumonia and nodule detection	Comparable to or exceeding expert readers
Radiological text (LLMs)	Differential diagnosis generation	GPT-4 top-diagnosis accuracy 57.1%

Beyond conventional models, large language models have been evaluated on radiological tasks such as generating differential diagnoses from transcribed imaging findings. A systematic review and meta-analysis found that GPT-4 outperformed GPT-3.5 on top-diagnosis accuracy (57.1 percent versus 51.1 percent) and differential accuracy (56.5 percent versus 37.9 percent), with more advanced iterations reaching higher differential accuracy, although overall performance remained well below expert levels and models exhibited substantial limitations in reliability (GPT-based differential diagnostic accuracy, 2025; Testing the ability of ChatGPT to generate differential diagnoses, 2024). Vision-language models have similarly shown that interpreting radiological images remains challenging, with unreliable performance in the absence of clinical context (Revolution or risk? GPT-4V, 2024).

These performance findings carry a methodological caveat. Because the studies that report them use different datasets, reference standards, and evaluation protocols, the numerical metrics are best read as indicative of the level of capability rather than as directly comparable benchmarks. The consistent finding across modalities, however, is qualitative and robust: modern deep learning models, on well-defined and adequately powered tasks, now operate at or near the level of expert clinicians, while models applied to open-ended or context-dependent tasks, such as generating differential diagnoses from ambiguous findings, remain substantially less reliable. This distinction between well-scoped and open-ended tasks, rather than between specific numbers, is the most reliable conclusion the performance evidence supports.

4.4 From Approval to Clinical Adoption

The final set of results concerns the translation of algorithmic advances into clinical practice. A clear gap exists between the number of AI-enabled devices that obtain regulatory approval and their actual adoption and use in routine care (Navigating the AI landscape in medical imaging, 2025). Regulatory approval demonstrates technical

safety and performance in controlled settings, whereas clinical usability depends on seamless integration into existing workflows, user friendliness for clinicians, and support for decision-making in complex, ambiguous real-world scenarios (Real-world adoption of AI in radiology, 2024). Figure 3 illustrates this conceptual distinction between algorithmic capability and clinical uptake.

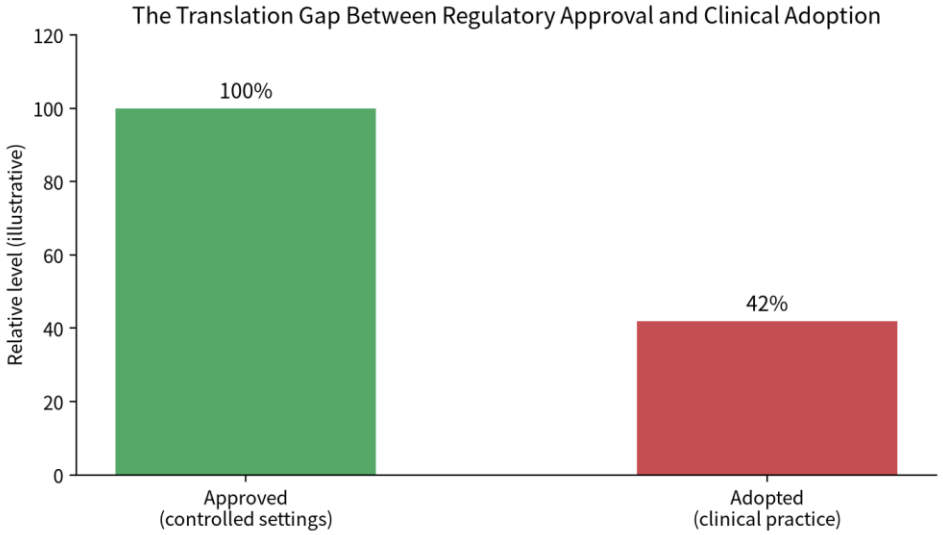


Figure 3. The translation gap between regulatory approval and clinical adoption of AI-enabled medical devices

Implementation studies identify recurrent barriers. Current AI applications often have narrow functions, requiring the combined use of multiple applications from different vendors, which imposes considerable overhead costs on healthcare organisations (A holistic approach to implementing AI in radiology, 2024). There is a lack of standard interfaces to integrate AI results into clinical workflow, and the reference standards used to validate AI tools are frequently inconsistent across institutions (Challenges of implementing AI-enabled programs, 2024). These findings indicate that the principal bottleneck in machine-learning-based biomedical engineering is no longer algorithmic accuracy but the surrounding infrastructure of validation, integration, governance, and human factors.

The results also draw attention to the role of the regulatory pathway itself in shaping the clinical landscape. The dominance of the 510(k) pathway, which grants clearance on the basis of substantial equivalence to a predicate device, has enabled rapid market entry for a large volume of devices, but it also means that many AI systems enter the market on the strength of performance demonstrated in relatively narrow, controlled evaluations rather than through prospective clinical trials in realistic practice (Machine learning-enabled devices authorized by the FDA in 2024, 2025). In response, regulators are developing tools to manage the safe evolution of algorithmic systems, including predetermined change control plans that permit specified, pre-approved modifications to an AI device without a new submission (FDA PCCP guidance, 2024). The interaction between rapid clearance and evolving post-market oversight is a defining feature of the current translational landscape, and it reinforces the need for robust post-market surveillance and real-world evidence generation as complements to pre-market evaluation (A decade of review in global regulation, 2025).

5. Discussion

The results confirm that machine learning has become a mature and rapidly expanding capability within biomedical engineering, particularly in medical imaging, and that its translation into clinical practice is at a critical juncture. The evidence on regulatory clearance shows sustained, exponential growth in the number of AI-enabled devices reaching the market, while the evidence on performance demonstrates that algorithmic accuracy across many imaging tasks is now at or near the level of expert clinicians. Yet the evidence on adoption indicates that the realisation of clinical benefit lags substantially behind technical capability.

5.1 The Algorithm-Data-Regulation Triad

The results can be interpreted through the interaction of three forces: algorithms, data, and regulation. Algorithmic progress, exemplified by the evolution from convolutional networks to transformer-based and hybrid architectures,

has extended the ceiling of achievable performance (TAC-UNet, 2024; CSWin-UNet, 2024). But the capacity of models is bounded by the data available to train them, and the evidence highlights persistent limitations in data quality, label reliability, and generalizability across institutions and populations (Limitations of public chest radiography datasets, 2024). Regulation, in turn, has adapted to accommodate algorithmic products through pathways such as the 510(k) process and, increasingly, frameworks for the safe evolution of AI systems over time, such as predetermined change control plans (FDA PCCP guidance, 2024). The interaction of these three forces determines the pace and safety of clinical translation.

The data dimension of this triad deserves particular emphasis, because it is simultaneously the greatest enabler and the greatest constraint. On the one hand, the availability of large, public, and increasingly multimodal datasets has been the foundation on which modern medical machine learning is built, enabling models to learn representations that generalise across common pathologies and imaging protocols. On the other hand, the sensitivity of medical data, together with privacy regulation and institutional barriers to sharing, means that the most informative data, often those from diverse and underserved populations, are frequently the least accessible. Federated learning and related privacy-preserving techniques offer a principled response, allowing models to be trained across institutions without raw data leaving their sites, but they remain technically demanding and are not yet standard practice (Privacy preserving federated learning, 2024; Federated learning for secure neuroimaging, 2024). Closing the data gap is thus as much an institutional and governance challenge as a technical one, and progress will depend on the willingness of healthcare organisations to participate in shared, privacy-preserving data ecosystems.

5.2 Validation, Generalizability, and Real-World Performance

A central theme of the results is the distinction between benchmark performance and real-world utility. Reported accuracies of 92 to 95 percent for nodule detection and 88 to 93 percent for tumour classification are impressive, but they are achieved in specific datasets and settings (Advancements in AI for automated radiological diagnosis, 2024). When models are deployed across different scanners, protocols, and patient populations, performance frequently degrades, sometimes sharply. The evaluation of large language models on radiological tasks illustrates a similar pattern: models achieve meaningful but incomplete diagnostic capability, with substantial reliability limitations that preclude unsupervised use (GPT-based differential diagnostic accuracy, 2025; Revolution or risk? GPT-4V, 2024). Robust clinical translation therefore requires prospective, multi-site validation with clinically meaningful endpoints, rather than reliance on retrospective benchmark accuracy alone.

5.3 Integration Into the Clinical Workflow

The adoption evidence points to workflow integration as the decisive practical bottleneck. Narrow, single-function AI applications do not, on their own, transform practice; they must be embedded into radiology reporting systems, electronic health records, and decision pathways in ways that reduce, rather than add to, clinician workload (A holistic approach to implementing AI in radiology, 2024). The absence of standard interfaces and the fragmentation of vendor solutions create integration overhead that many institutions cannot absorb (Challenges of implementing AI-enabled programs, 2024). This finding has direct implications for biomedical engineers: the design problem is not only the model but the system in which the model operates, including its user interface, alerting logic, and place in the broader clinical process.

Human factors and trust are inseparable from workflow integration. Clinicians will only use an algorithmic tool reliably if they understand its purpose, trust its outputs in the contexts where they are reliable, and are equipped to respond when it is wrong. Research on AI implementation finds that acceptance depends heavily on how a tool is integrated into existing workflows and on its perceived value in managing workload, with triaging functions viewed as effective but of variable usefulness depending on local pressures (A holistic approach to implementing AI in radiology, 2024). Transparent reporting of model characteristics, explicit communication of uncertainty, and mechanisms for human oversight are therefore not optional additions but essential components of any clinical AI system. The engineering challenge is to design systems that are transparent enough to earn calibrated trust, yet embedded enough to be used effortlessly within the constraints of busy clinical practice.

5.4 Beyond Imaging: The Broader Biomedical Frontier

The principles that govern medical imaging also apply, with adjustment, to the wider field of machine learning in biomedical engineering. Wearable biosensors and brain-computer interfaces face analogous challenges of signal quality, individual variability, clinical validation, and integration, compounded by the additional difficulty of operating in uncontrolled, home-based or everyday settings (Frontiers of wearable biosensors, 2023; Neural decoding for intracortical BCIs, 2023). Privacy-preserving techniques such as federated learning are emerging as an important enabling infrastructure across all of these domains, allowing models to be trained on distributed data without compromising patient confidentiality (Future-proofing medical imaging, 2024; Federated learning for

secure neuroimaging, 2024). The unifying lesson is that responsibility, transparency, and rigorous evaluation are as important to the clinical value of machine learning as the sophistication of its algorithms.

The opportunities in these frontier domains are considerable. Continuous wearable monitoring, coupled with machine learning, holds the promise of detecting deterioration earlier than episodic clinical encounters, enabling preventive rather than reactive care, and extending specialist-level monitoring to settings where specialist expertise is scarce (Frontiers of wearable biosensors, 2023; AI-driven wireless sensing, 2025). Brain-computer interfaces, while still in their infancy as clinical tools, illustrate the transformative potential of decoding neural activity to restore function for individuals with severe disability, and the pace of progress in electrode materials, decoding algorithms, and clinical trials suggests that these systems will become more capable and more widely applicable over time (Advances in electrode interface materials, 2024; Neural decoding for intracortical BCIs, 2023). In both domains, however, the same translational discipline is required: rigorous validation in realistic conditions, attention to individual variability and safety, and careful integration with the human users and care pathways the technology is intended to support.

5.5 Implications for Stakeholders

The results carry distinct implications for the principal stakeholders. For biomedical engineers, they argue for a broader conception of the field that treats validation, interoperability, and human factors as first-class design considerations rather than afterthoughts. For clinicians, they counsel a posture of informed engagement, in which AI tools are understood as decision support whose strengths and limitations are well characterised, rather than as autonomous authorities. For regulators, the results underscore the value of adaptive, risk-based frameworks that can accommodate the rapid and continuous evolution of algorithmic systems while maintaining rigorous oversight of safety and effectiveness (FDA regulation of clinical software, 2024; A decade of review in global regulation, 2025).

6. Conclusion and Recommendations

Machine learning has established itself as a defining technology of modern biomedical engineering. This review has documented the rapid growth of AI-enabled medical devices, the high and improving performance of deep learning models across imaging tasks and modalities, and the persistent gap between algorithmic capability and clinical adoption. The principal findings can be summarised as follows. First, regulatory clearance of AI/ML-enabled devices has grown dramatically, with cumulative authorizations exceeding 1,400 devices by the end of 2025 and radiology accounting for roughly three-quarters of all clearances. Second, deep learning models achieve diagnostic performance at or near expert levels across a range of imaging tasks, from pulmonary nodule detection to brain tumour classification. Third, the translation of these advances into routine clinical practice remains uneven, constrained by validation, generalizability, integration, and governance challenges.

The evidence indicates that the principal bottleneck in machine-learning-based biomedical engineering is no longer raw model accuracy but the surrounding infrastructure required for responsible clinical use. Progress will depend on the coordinated development of robust validation methods, privacy-preserving data infrastructure, interoperable clinical systems, and adaptive regulatory frameworks, pursued in close collaboration between engineers, clinicians, and regulators.

It is worth restating the central conclusion with appropriate nuance. The remarkable progress of machine learning in biomedical engineering does not imply that algorithms will, or should, replace clinicians, and the evidence reviewed here does not support a vision of autonomous diagnostic systems operating without meaningful human oversight. Rather, the trajectory of the field points toward a partnership in which machine learning augments human expertise, relieving clinicians of routine and repetitive tasks, surfacing patterns that human readers might miss, and enabling measurements and monitoring that were previously infeasible, while human judgement retains responsibility for the interpretation, the communication of uncertainty, and the ultimately ethical decisions that clinical care demands. The engineering, regulatory, and educational systems of the coming decade will determine whether this partnership is realised beneficially, safely, and equitably across the full diversity of patients and care settings that biomedical engineering ultimately serves.

6.1 Recommendations

Based on the analysis, the following recommendations are proposed. First, biomedical engineers should treat clinical validation and workflow integration as core design objectives, developing AI systems that are interoperable with existing clinical infrastructure and designed to complement, rather than replace, clinical judgement. Second, the field should invest in the development of large, diverse, and privacy-preserving data-sharing infrastructure, including federated learning frameworks that enable robust multi-site training and

validation without compromising patient confidentiality. Third, regulators should continue to develop adaptive, risk-based frameworks that accommodate the continuous evolution of algorithmic systems while maintaining rigorous, transparent oversight of safety and effectiveness. Fourth, evaluation standards should be harmonised so that reported performance is comparable across studies, institutions, and modalities, and so that real-world performance is assessed prospectively with clinically meaningful endpoints. Fifth, interdisciplinary collaboration and training should be strengthened so that clinicians understand the capabilities and limitations of AI tools, and engineers understand the realities of clinical practice. By pursuing these priorities, biomedical engineering can translate the remarkable algorithmic advances of the past decade into measurable improvements in diagnosis, treatment, and patient care.

References

Almarie, B., Gonzalez-Gonzalez, L. F., Dos Santos Barbosa, L. A., Lutz, A., Grosse, U., & Fregni, F. (2025). Machine learning-enabled medical devices authorized by the US Food and Drug Administration in 2024: Regulatory characteristics, predicate lineage, and transparency reporting. *Health Affairs Scholar.*

Bhandari, A. (2024). *Revolutionizing radiology with artificial intelligence.* Cureus.

Busch, F., Hoffmann, L., Pinto dos Santos, D., Makowski, M. R., Saba, L., Prucker, P., … Bressem, K. K. (2025). Large language models for structured reporting in radiology: Past, present, and future. *Insights into Imaging.*

Chen, R. (2026). FDA AI/ML-enabled medical device registry and annual clearance statistics. *PharmaDossier.*

Dong, Y.-R., Wang, S.-R., Huang, Q., Berg, R. W., Li, G.-H., & He, J.-P. (2023). Neural decoding for intracortical brain-computer interfaces. *Cyborg and Bionic Systems.*

Edelman, B. J., Zhang, S.-L., Schalk, G., Brunner, P., Müller-Putz, G., Guan, C.-T., & He, B. (2024). Non-invasive brain-computer interfaces: State of the art and trends. *IEEE Sensors Journal.*

Forghani, R. (2024). Real-world adoption of artificial intelligence in radiology: Opportunities and barriers. *American Journal of Neuroradiology.*

He, J., Ma, Y., Yang, M., Yang, W., & Wu, C. (2024). TAC-UNet: Transformer-assisted convolutional neural network for medical image segmentation. *Scientific Reports.*

Jacobs, P.-P., Ehrengut, C., Bucher, A. M., Penzkofer, T., Lukas, M., Kleesiek, J., & Denecke, T. (2023). Challenges in implementing the local node infrastructure for a national federated machine learning network in radiology. *Radiology: Artificial Intelligence.*

Jiao, Y., Lei, M., Zhu, J., Chang, R., & Qu, X. (2024). Advances in electrode interface materials and modification technologies for brain-computer interfaces. *Frontiers in Bioengineering and Biotechnology.*

Kim, B., Romeijn, S., van Buchem, M., Rezazade Mehrizi, M. H., & Grootjans, W. (2024). A holistic approach to implementing artificial intelligence in radiology. *Insights into Imaging.*

Koutsoubis, N., Waqas, A., Yilmaz, Y., Ramachandran, R. P., Schabath, M., & Rasool, G. (2024). Future-proofing medical imaging with privacy-preserving federated learning and uncertainty quantification: A review. *arXiv:2409.16340.*

Koutsoubis, N., Yilmaz, Y., Ramachandran, R. P., Schabath, M., & Rasool, G. (2024). Privacy preserving federated learning in medical imaging with uncertainty estimation. *arXiv:2406.12815*.

Kovur, P., Kovur, K. M., Rayat, D. Y., & Wishart, D. S. (2024). POC sensor systems and artificial intelligence: Where we are now and where we are going. ACS Sensors.

Laçi, H., Sevrani, K., & Iqbal, S. (2024). Deep learning approaches for classification tasks in medical X-ray, MRI, and ultrasound images: A scoping review. *Diagnostic and Interventional Radiology.*

Liu, X., Gao, P., Yu, T., Wang, F., & Yuan, R.-Y. (2024). CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation. *arXiv:2407.18070.*

Mehta, V., Komanduri, A., Bhadouriya, R. S., Mehta, V., Johnson, M. D., Shrestha, P., Nikolov, M., Jain, B., Shah, N., & Schulman, K. (2024). Evaluating transparency in AI/ML model characteristics for FDA-reviewed medical devices. *npj Digital Medicine.*

Nguyen, D., Bronson, I., Chen, R., & Kim, Y. H. (2025). A systematic review and meta-analysis of GPT-based

differential diagnostic accuracy in radiological cases: 2023-2025. *Frontiers in Radiology.*

Patil, S., Valvi, R. D., & Ghodake, D. S. (2024). Advancements in artificial intelligence for automated radiological diagnosis: A systematic review. *International Journal of Legal and Interdisciplinary Legal Studies.*

Rafferty, A., & Rajan, A. (2024). Limitations of public chest radiography datasets for artificial intelligence: Label quality, domain shift, bias, and evaluation challenges. *arXiv.*

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Lecture Notes in Computer Science 9351*, 234-241. https://arxiv.org/abs/1505.04597

Singh, V., Cheng, S., Kwan, A. C., & Ebinger, J. (2024). United States Food and Drug Administration regulation of clinical software in the era of artificial intelligence and machine learning. *JMIR AI*.

Sosna, J., Joskowicz, L., & Saban, M. (2025). Navigating the AI landscape in medical imaging: A critical analysis of technologies, implementation, and implications. *Radiology.*

Stripelis, D., Gupta, U., Saleem, H., Dhinagar, N., Ghai, T., Anastasiou, C., ... Ambite, J. L. (2024). A federated learning architecture for secure and private neuroimaging analysis. *Nature Communications.*

Sun, S. H., Huynh, K., Cortes, G., Hill, R., Tran, J., Yeh, L., Ngo, A. L., Houshyar, R., Yaghmai, V., & Tran, M. (2024). Testing the ability and limitations of ChatGPT to generate differential diagnoses from transcribed radiologic findings. *Radiology.*

Thrall, J. H. (2024). Challenges of implementing artificial intelligence-enabled programs in the clinical practice of radiology. *American Journal of Roentgenology.*

Toruner, M. D., Shi, V., Sollee, J., Hsu, W.-C., Yu, G., Dai, Y.-W., Merlo, C., Suresh, K., Jiao, Z., Wang, X., Mao, S., & Bai, H. (2025). Artificial intelligence-driven wireless sensing for health management. *Sensors.*

Toruner, M. D., Wang, Y., Jiao, Z., & Bai, H. (2024). Artificial intelligence in radiology: Where are we going? *Diagnostics.*

U.S. Food and Drug Administration. (2024). Marketing submission recommendations for a predetermined change control plan for artificial intelligence-enabled device software functions; Guidance for industry and Food and Drug Administration staff. *Federal Register.*

Viswanathan, V. (2025). Wearable technologies and engineering for healthcare: Current applications, challenges, and future directions. *International Journal for Multidisciplinary Research.*

Wu, Y., Xia, S., Liang, Z., Chen, R., & Qi, S. (2024). Artificial intelligence in COPD CT images: Identification, staging, and quantitation. *Heliyon.*

Xian, X. (2023). Frontiers of wearable biosensors for human health monitoring. *Biosensors.*

Yao, W., Bai, J., Liao, W., Chen, Y., Liu, M., & Xie, Y. (2024). From CNN to transformer: A review of medical image segmentation models. *Journal of Imaging Informatics in Medicine.*

Zhang, S.-W., Li, Y.-J., Liu, W.-X., Chu, Q., Wang, S.-S., Li, J.-Y., & Chen, Y.-W. (2025). A decade of review in global regulation and research of artificial intelligence medical devices (2015-2025). *International Journal of Surgery*.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).