International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(16s) 2026 pp. 350-356
https://doi.org/10.70917/ijcisim-2026-4580
Article

AI-Enabled Smart Healthcare Systems: Innovative Approaches to Disease Prediction, Accurate Diagnosis, and Effective Public Health Management

Vinit Kumar Ramawat[1], G.PRABHAKARAN [2], Pinki Das[3], A. Sasi Kumar [4], Ramesh Kumar[5], Tara Sasanka, C[6]

[1]School of Nursing, Noida International University, Gautam Budh Nagar, Greater Noida, Uttar Pradesh, Vicky.aryan21@gmail.com
[2]Department Of CIVIL ENGINEERING, SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY, TIRUPATI DISTRICT, PUTTUR, ANDHRA PREDESH, gprabhadhana@gmail.com
[3]School of Nursing, School of Nursing Science & Research, Sharda University, Gautam Budh Nagar, Greater Noida, Uttar Pradesh, dpinki55@gmail.com
[4]School of Science and Computer Studies, CMR University, Bangalore askmca@yahoo.com
[5]Independent Researcher, Dhanbad, Jharkhand, rameshbitm@gmail.com

Abstract: Artificial intelligence (AI) has moved from a peripheral research interest to a central component of contemporary healthcare delivery, with deep learning systems now matching or exceeding specialist-level performance on discrete diagnostic tasks spanning dermatology, ophthalmology, radiology, and oncology. This paper reviews the technical foundations and applied evidence base for AI-driven smart healthcare systems across three domains: disease prediction from structured and unstructured clinical data, image-based diagnostic classification, and population-level public health management, including the accelerated adoption of AI tools during the COVID-19 pandemic. The review synthesizes landmark diagnostic-performance studies, including dermatologist-level skin cancer classification, diabetic retinopathy detection, pneumonia detection from chest radiographs, and international breast cancer screening evaluation, alongside the clinical machine learning literature addressing implementation, validation, and algorithmic bias. Comparative tables summarize reported diagnostic performance metrics, data modalities, and clinical domains across the reviewed systems. The paper concludes that AI-driven diagnostic systems have achieved genuine, reproducible performance parity with human specialists on narrow, well-defined tasks, while broader clinical deployment remains constrained by validation, generalizability, and algorithmic-bias challenges that the reviewed literature has only begun to resolve, and identifies prospective real-world validation and equity-focused model auditing as the central future research prospects.
Keywords: Artificial Intelligence, Smart Healthcare, Disease Prediction, Medical Diagnosis, Deep Learning, Public Health, Machine Learning in Medicine, Algorithmic Bias.

1. Introduction

Healthcare systems worldwide face a persistent structural mismatch between the volume of clinical and diagnostic data generated and the human specialist capacity available to interpret it, a mismatch that has motivated more than a decade of research into machine learning systems capable of performing specific diagnostic and predictive tasks at or beyond expert-level accuracy [1]. Obermeyer and Emanuel's early framing of this shift in the *New England Journal of Medicine* argued that machine learning's comparative advantage in medicine lies not in replacing physician judgment broadly but in prediction, specifically the well-defined, data-rich, pattern-recognition tasks such as image classification and risk stratification that constitute a substantial share of diagnostic and screening workflows [2]. This

framing anticipated the subsequent wave of landmark diagnostic-performance studies reviewed in this paper, each of which targets a narrow, well-specified prediction task rather than general clinical reasoning [3], [4], [5].

Esteva et al.'s dermatologist-level skin cancer classification study, published in *Nature*, trained a convolutional neural network on 129,450 clinical images spanning 2,032 distinct skin diseases and demonstrated that the resulting classifier matched the diagnostic accuracy of 21 board-certified dermatologists across biopsy-proven test cases, a result widely regarded as the first unambiguous demonstration that deep learning could achieve specialist-level performance on a genuinely difficult, high-stakes medical image classification task [3]. Gulshan et al.'s parallel study, published in *JAMA*, developed and validated a deep learning algorithm for detecting diabetic retinopathy from retinal fundus photographs, reporting sensitivity and specificity exceeding 90 percent against a reference standard established by panels of ophthalmologists, a result with direct public health relevance given diabetic retinopathy's status as a leading cause of preventable blindness in working-age adults globally [4]. Rajpurkar et al.'s CheXNet model extended this diagnostic-parity finding to chest radiograph interpretation, reporting that a 121-layer convolutional network trained on more than 100,000 frontal-view chest X-rays achieved pneumonia-detection performance exceeding the average performance of four practicing radiologists on the same held-out test set [5].

Beyond discrete image-classification tasks, a parallel literature has examined AI's application to structured electronic health record (EHR) data for disease risk prediction and clinical decision support. Rajkomar, Dean, and Kohane's review of machine learning in medicine, published in the *New England Journal of Medicine*, catalogued applications spanning mortality prediction, readmission risk stratification, and treatment-response prediction from EHR data, while emphasizing that the successful translation of strong retrospective model performance into genuine clinical benefit depends on prospective validation within the actual clinical workflow the model is intended to support, a distinction the authors argue much of the early machine-learning-in-medicine literature had not adequately addressed [6]. Chen and Asch's related commentary cautioned against what they termed the "peak of inflated expectations" surrounding clinical machine learning, arguing that strong performance on retrospective, curated datasets frequently fails to generalize to the messier, more heterogeneous data encountered in live clinical deployment [7].

The COVID-19 pandemic produced an unusually concentrated natural experiment in AI-driven public health management, generating a rapid proliferation of predictive and diagnostic models within a compressed timeframe. Wynants et al.'s systematic review and critical appraisal of COVID-19 prediction models identified 232 distinct prediction models published within the pandemic's first year, but found that the substantial majority were at high risk of bias due to non-representative patient selection, inadequate handling of missing data, or insufficient external validation, a finding that directly illustrates the validation gap Rajkomar, Dean, and Kohane and Chen and Asch had identified in the broader clinical machine learning literature [7], [8]. Bullock et al.'s complementary mapping of AI applications against COVID-19 catalogued the pandemic's AI applications across diagnosis, epidemiological surveillance, and drug-repurposing research, while similarly noting that few of the surveyed applications had progressed from research demonstration to validated clinical or public health deployment within the period studied [9].

This paper synthesizes the diagnostic-performance, clinical-implementation, and public-health-application literatures into a single comparative framework, examining reported performance metrics, data modalities, and validation status across the major AI-driven smart healthcare systems reviewed, before turning to the algorithmic-bias and equity considerations that increasingly shape the field's translational trajectory.

Research Objectives

- To review landmark evidence on AI-driven diagnostic performance across dermatology, ophthalmology, radiology, and oncology.
- To synthesize the machine learning literature addressing structured electronic health record-based disease prediction and clinical decision support.
- To examine AI's role in public health management, with particular attention to the COVID-19 pandemic as a concentrated case study.
- To evaluate reported validation gaps and algorithmic-bias concerns across the reviewed clinical machine learning literature.
- To identify future research priorities for prospective validation and equity-focused model auditing in smart healthcare systems.

2. Literature Review

2.1 Image-Based Diagnostic Classification

Esteva et al.'s dermatology classifier, trained end-to-end on clinical images without hand-engineered feature extraction, demonstrated performance on par with board-certified dermatologists across both the classification of malignant versus benign lesions and the finer-grained classification of specific skin cancer subtypes, and the authors specifically highlighted the model's potential for deployment via consumer smartphone cameras as a low-cost triage tool in settings lacking specialist dermatological access [3]. McKinney et al.'s subsequent international evaluation of an AI system for breast cancer screening, published in *Nature*, trained and validated a deep learning mammography-interpretation system across large, geographically distinct datasets from the United Kingdom and the United States, reporting an absolute reduction in both false positives and false negatives relative to the original radiologist assessments, and further demonstrated in a reader study that combining the AI system's assessment with a single human reader matched the accuracy of the standard double-reading process used in the UK screening program, suggesting a potential workflow-efficiency benefit alongside the accuracy improvement [10]. Litjens et al.'s broad survey of deep learning in medical image analysis, reviewing more than 300 contributions across the field, found consistent performance gains from convolutional architectures relative to earlier feature-engineered approaches across nearly every medical imaging subspecialty examined, while noting that the degree of performance improvement varied considerably by imaging modality and by the availability of large, well-annotated training datasets specific to each clinical task [11].

2.2 Structured Data-Driven Disease Prediction

Rajkomar, Dean, and Kohane's review distinguished image-based diagnostic classification, the domain of the studies reviewed in Section 2.1, from structured EHR-based prediction, which draws on longitudinal, multi-modal patient records, including laboratory values, vital signs, medication history, and free-text clinical notes, to predict outcomes such as inpatient mortality, 30-day readmission, and prolonged length of stay [6]. Miotto, Wang, Wang, Jiang, and Dudley's comprehensive review of deep learning for healthcare specifically addressed the challenge of representing heterogeneous, irregularly sampled EHR data in a form suitable for deep learning, cataloguing approaches including recurrent neural networks for temporal sequence modeling and autoencoder-based representation learning for reducing the extremely high dimensionality of raw EHR feature spaces, and concluded that while EHR-based deep learning models frequently outperform traditional logistic-regression-based clinical risk scores on retrospective validation, the interpretability of these more complex models remains a persistent barrier to clinical adoption relative to simpler, more transparent scoring systems clinicians already trust [12]. Ching et al.'s broader review of deep learning applications across biology and medicine situated EHR-based clinical prediction within this larger landscape, noting that the field's rapid technical progress has consistently outpaced the parallel development of rigorous, prospective clinical validation standards [13].

2.3 Implementation, Validation, and the Translational Gap

Char, Shah, and Magnus's commentary on implementing machine learning in healthcare identified a set of ethical and practical challenges specific to clinical deployment, including the risk that a model trained on historical clinical data will reproduce and perpetuate the biases embedded in that historical data, and the related risk that a model's predictions may become self-fulfilling once clinicians begin to act on them, altering the very outcome distribution the model was trained to predict [14]. Chen and Asch's cautionary analysis argued that the gap between strong retrospective performance and genuine clinical benefit is frequently underappreciated in the published clinical machine learning literature, since retrospective validation on curated, relatively clean datasets does not reliably predict performance when a model encounters the full heterogeneity, missing data, and distributional shift characteristic of live clinical deployment across different hospital systems [7]. Topol's widely cited perspective on the convergence of human and artificial intelligence in medicine, published in *Nature Medicine*, argued that AI's most durable contribution to clinical practice would likely come not from replacing physician judgment but from offloading routine, pattern-recognition-heavy diagnostic tasks, freeing clinician time and attention for the more complex, relational, and judgment-intensive aspects of patient care, a framing consistent with Obermeyer and Emanuel's earlier prediction-focused characterization of AI's comparative advantage [1], [2].

2.4 Algorithmic Bias in Clinical AI

Obermeyer, Powers, Vogeli, and Mullainathan's widely cited analysis, published in *Science*, examined a commercial risk-prediction algorithm used to identify patients for enrollment in high-risk care management programs across a large U.S. health system, and found that the algorithm systematically underestimated the health needs of

Black patients relative to white patients with comparable underlying illness, a bias traced specifically to the algorithm's use of historical healthcare cost, rather than direct health status, as its prediction target, since Black patients in the training data had historically incurred lower healthcare costs for a given level of underlying illness due to unequal access to care [15]. This finding is directly relevant to the validation and generalizability concerns raised by Chen and Asch, and Char, Shah, and Magnus, since it demonstrates that even a model with strong aggregate predictive accuracy can embed a clinically consequential and demographically patterned bias that aggregate performance metrics alone fail to reveal [7], [14], [15]. Davenport and Kalakota's broader review of AI's potential in healthcare situated this bias concern within a wider discussion of regulatory and workforce readiness, arguing that algorithmic bias auditing, alongside clinician AI literacy training, represents a necessary, and currently underdeveloped, component of responsible clinical AI deployment [16].

2.5 AI in Public Health Management and Pandemic Response

Vaishya, Javaid, Khan, and Haleem's review of AI applications for COVID-19 pandemic management catalogued applications spanning early outbreak detection from unstructured data sources, epidemiological forecasting, and CT-scan-based diagnostic triage, and characterized the pandemic as an accelerant that compressed years of typical AI-healthcare adoption timelines into a period of months, driven by the urgent operational need for rapid triage and resource-allocation tools under conditions of severe healthcare system strain [17]. Wynants et al.'s systematic review and critical appraisal of the resulting proliferation of COVID-19 prediction models found that despite this rapid proliferation, the overwhelming majority of published models exhibited a high risk of bias, most commonly due to non-representative patient selection and inadequate external validation, and the review's authors explicitly recommended that none of the reviewed models be adopted for clinical use without substantial further validation, a striking conclusion given the models' rapid and widespread research uptake during the same period [8]. Ienca and Vayena's commentary on the responsible use of digital data during the pandemic raised parallel concerns regarding the public health surveillance applications of AI, specifically contact-tracing and mobility-data-based outbreak modeling, arguing that the urgency of pandemic response created pressure to deploy insufficiently validated or inadequately privacy-protective digital surveillance tools, a tension between speed and rigor that recurs throughout the pandemic-era AI healthcare literature [18].

2.6 Research Gap

While the diagnostic-performance literature reviewed in Section 2.1 has established robust, reproducible evidence of specialist-level AI performance on narrowly defined image-classification tasks, and the structured-data and public-health literatures reviewed in Sections 2.2 and 2.5 have documented a substantial and largely unresolved validation gap between retrospective model performance and genuine clinical or public health benefit, comparatively few studies directly integrate diagnostic-performance evidence with the algorithmic-bias auditing methodology Obermeyer, Powers, Vogeli, and Mullainathan applied to structured risk-prediction models, leaving open the question of whether the strong aggregate performance reported for image-based diagnostic classifiers conceals comparable demographically patterned bias [3], [4], [5], [15]. This review addresses this gap by organizing the reviewed evidence within a comparative framework spanning data modality, reported performance, and validation and bias-auditing status.

3. Methodology

This study adopts a qualitative, descriptive, and comparative research design based on synthesis of peer-reviewed clinical machine learning, medical imaging, and public health informatics literature addressing AI-driven disease prediction, diagnosis, and public health management. Reviewed studies were compared along four axes: data modality (medical imaging, structured EHR data, or public health surveillance data); clinical domain; reported performance metric and comparator (human specialist, existing clinical score, or prior model generation); and validation status (retrospective single-site, retrospective multi-site/external, or prospective clinical evaluation).

4. Results and Discussion

4.1 Reported Diagnostic Performance Across Landmark Studies

Table 1 summarizes the reported performance metrics of the landmark image-based diagnostic studies reviewed, alongside their comparator and data scale.

Table 1. Reported Performance of Landmark AI Diagnostic Systems

Study	Clinical Domain	Data Scale	Reported Performance	Comparator
Esteva et al. [3]	Dermatology (skin cancer classification)	129,450 clinical images, 2,032 diseases	Diagnostic accuracy on par with specialists	21 board-certified dermatologists
Gulshan et al. [4]	Ophthalmology (diabetic retinopathy)	~128,000 retinal images (development set)	Sensitivity/specificity both exceeding 90%	Panels of ophthalmologists (reference standard)
Rajpurkar et al. [5]	Radiology (pneumonia detection, chest X-ray)	100,000+ frontal chest X-rays	Exceeded average performance of comparator group	Four practicing radiologists
McKinney et al. [10]	Oncology (breast cancer screening, mammography)	Large UK and US screening datasets	Reduced both false positives and false negatives vs. original reads	Radiologists (UK double-reading standard)

4.2 Data Modality, Validation Status, and Translational Readiness

Table 2 compares the reviewed literature by data modality and reported validation rigor, distinguishing the image-based diagnostic studies from the structured-data and public-health prediction literature.

Table 2. Data Modality and Validation Status Across Reviewed Literature

Domain	Representative Studies	Predominant Data Modality	Reported Validation Status
Image-based diagnosis	Esteva et al. [3]; Gulshan et al. [4]; Rajpurkar et al. [5]; McKinney et al. [10]	Medical imaging (clinical photos, fundus images, X-rays, mammograms)	Multi-site external validation reported in most landmark studies
Structured EHR-based prediction	Rajkomar, Dean & Kohane [6]; Miotto et al. [12]	Structured/longitudinal clinical records	Predominantly retrospective; prospective validation uncommon
Risk-stratification algorithms	Obermeyer et al. [15]	Historical cost and utilization data	Retrospective; bias identified via post-hoc demographic audit
COVID-19 prediction models	Wynants et al. [8]; Vaishya et al. [17]	Mixed (imaging, clinical, epidemiological)	High risk of bias in majority (232 models reviewed); external validation rare
Public health surveillance	Ienca & Vayena [18]; Bullock et al. [9]	Mobility, contact-tracing, epidemiological data	Deployment often outpaced formal validation during pandemic

4.3 Discussion

Taken together, the evidence in Tables 1 and 2 supports a domain-dependent, rather than uniform, assessment of AI readiness across smart healthcare applications. The image-based diagnostic literature reviewed in Section 2.1 and summarized in Table 1 represents the field's most mature and reproducibly validated achievement, with Esteva et al., Gulshan et al., Rajpurkar et al., and McKinney et al. each reporting specialist-comparable or specialist-exceeding performance across independent clinical domains and, in several cases, across geographically distinct validation datasets [3], [4], [5], [10]. This maturity contrasts sharply with the structured-data prediction and public-health-surveillance literature summarized in Table 2, where Rajkomar, Dean, and Kohane, Miotto et al., and Wynants et al. each document a persistent gap between retrospective performance and demonstrated prospective clinical or public health benefit, a gap Chen and Asch attribute specifically to the distributional differences between curated retrospective datasets and the heterogeneous conditions of live deployment [6], [7], [8], [12]. Obermeyer, Powers, Vogeli, and Mullainathan's bias-auditing finding adds a further, largely orthogonal dimension to this readiness

assessment: a model can achieve strong aggregate retrospective performance, as the risk-prediction algorithm they examined did, while simultaneously embedding a clinically consequential demographic bias invisible to standard aggregate performance metrics, a concern that Table 1's landmark diagnostic studies have generally not been subjected to with comparable rigor [15]. The COVID-19 pandemic evidence reviewed in Section 2.5 illustrates both dynamics simultaneously and under compressed timescales: Vaishya, Javaid, Khan, and Haleem document rapid, widespread AI-tool proliferation, while Wynants et al.'s critical appraisal finds that this proliferation substantially outpaced adequate validation, and Ienca and Vayena raise parallel concerns specifically regarding the privacy and validation adequacy of pandemic-era public health surveillance applications [8], [9], [17], [18].

5. Conclusion

This paper has reviewed the technical foundations and applied evidence base for AI-driven smart healthcare systems across disease prediction, diagnosis, and public health management. The evidence indicates that image-based diagnostic AI has achieved genuine, reproducible, and in several cases externally validated performance parity with human specialists across dermatology, ophthalmology, radiology, and oncology, representing the field's most mature translational achievement. Structured EHR-based disease prediction and public-health-surveillance applications, by contrast, remain substantially constrained by a persistent gap between strong retrospective performance and demonstrated prospective clinical or public health benefit, a gap the COVID-19 pandemic's compressed adoption timeline illustrated with particular clarity. Algorithmic bias represents a further, largely distinct readiness dimension that aggregate performance metrics alone do not capture, as demonstrated directly by the demographically patterned bias identified in a widely deployed clinical risk-prediction algorithm. Collectively, this evidence supports a domain-specific rather than uniform assessment of smart healthcare AI's current translational readiness, with narrow, well-validated diagnostic classification considerably further advanced than broader predictive and surveillance applications.

6. Future Work

Future research should prioritize prospective, randomized, or workflow-embedded clinical evaluation of AI diagnostic and predictive tools, extending beyond the retrospective validation that dominates the current literature, since Chen and Asch, and Rajkomar, Dean, and Kohane both identify this translational gap as the field's central unresolved methodological limitation. Further work is needed to extend the demographic bias-auditing methodology Obermeyer, Powers, Vogeli, and Mullainathan applied to a single risk-prediction algorithm systematically across the landmark image-based diagnostic systems reviewed in Table 1, to establish whether the strong aggregate performance these systems report conceals comparable demographic performance disparities. Continued research should also develop standardized, pre-registered validation protocols specifically for pandemic-response and public health AI tools, addressing the speed-versus-rigor tension Ienca and Vayena, and Wynants et al. both identify, so that future public health emergencies do not reproduce the large-scale, inadequately validated model proliferation documented during COVID-19. Finally, future work should pursue improved interpretability methods for structured EHR-based deep learning models, since Miotto, Wang, Wang, Jiang, and Dudley identify interpretability, rather than raw predictive accuracy, as the more persistent barrier to clinical adoption of these more complex model architectures relative to simpler, more transparent risk scores clinicians already trust.

References

1. E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.
2. Z. Obermeyer and E. J. Emanuel, "Predicting the future — Big data, machine learning, and clinical medicine," New England Journal of Medicine, vol. 375, no. 13, pp. 1216–1219, 2016.
3. A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.
4. V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
5. P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.
6. A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019.
7. J. H. Chen and S. M. Asch, "Machine learning and prediction in medicine — Beyond the peak of inflated expectations," New England Journal of Medicine, vol. 376, no. 26, pp. 2507–2509, 2017.
8. L. Wynants et al., "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," BMJ, vol. 369, m1328, 2020.

9. J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," Journal of Artificial Intelligence Research, vol. 69, pp. 807–845, 2020.
10. S. M. McKinney et al., "International evaluation of an AI system for breast cancer screening," Nature, vol. 577, no. 7788, pp. 89–94, 2020.
11. G. Litjens et al., "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60–88, 2017.
12. R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," Briefings in Bioinformatics, vol. 19, no. 6, pp. 1236–1246, 2018.
13. T. Ching et al., "Opportunities and obstacles for deep learning in biology and medicine," Journal of the Royal Society Interface, vol. 15, no. 141, 20170387, 2018.
14. D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care — Addressing ethical challenges," New England Journal of Medicine, vol. 378, no. 11, pp. 981–983, 2018.
15. Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," Science, vol. 366, no. 6464, pp. 447–453, 2019.
16. T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," Future Healthcare Journal, vol. 6, no. 2, pp. 94–98, 2019.
17. R. Vaishya, M. Javaid, I. H. Khan, and A. Haleem, "Artificial intelligence (AI) applications for COVID-19 pandemic," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 14, no. 4, pp. 337–339, 2020.
18. M. Ienca and E. Vayena, "On the responsible use of digital data to tackle the COVID-19 pandemic," Nature Medicine, vol. 26, no. 4, pp. 463–464, 2020.