Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 18 March 2026
Revised 26 April 2026
Accepted 24 May 2026

Natural Resources for Human Health 2026, Vol. 6 (2s): 466–484

KEYWORDS:
Artificial intelligence, machine learning, deep learning, healthcare, medical imaging, predictive analytics, federated learning, explainability

Artificial Intelligence Driven Healthcare - A Survey of Machine Learning, Deep Learning and Intelligent Predictive Methods

Rajesh L[1]*, P. MeganaSanthoshi[2], Latha Krishna Moorthy[3], Abdullah Farhan J. Alshammari[4], Gadhiraju Tej Varma[5], Amaresh A M[6]

[1]Associate Professor, Department of Information Science and Engineering, Dayananda Sagar Academy of Technology and Management, Bengaluru, India. Email: rajesh-ise@dsatm.edu.in
[2]Associate Professor, KLM college of Engineering for Women, Kadapa, India. Email: pmsanthoshi@gmail.com
[3]Dept. of Biomedical Engineering, SSIT, SSAHE University, Tumakuru, India. Email: lathak@ssit.edu.in
[4]Assistant Professor, Department of CSE, Yanbu Industrial College, Saudi Arabia. Email: shammariafj@rcjy.edu.sa. ORCID: 0009-0006-5215-4724
[5]Assistant Professor, Department of IT, SRKR Engineering College, Bhimavaram, India. Email: tejvarma.varma@gmail.com
[6]Assistant Professor, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India. Email: amaresham@jssstuniv.in ORCID ID: 0000-0001-7499-6558
*Corresponding Author Email: rajesh-ise@dsatm.edu.in

ABSTRACT: Artificial intelligence has moved from a promising research direction to a working component of modern clinical practice, supporting screening, diagnosis, prognosis, treatment planning and hospital operations. This survey reviews the principal families of methods that drive this transformation: classical machine learning models built on engineered features, deep learning architectures that learn representations directly from images, signals and records, and intelligent predictive frameworks that combine learning with attention, graphs, generative modeling, reinforcement and federated optimization. We trace the methodological evolution from support vector machines and ensemble trees to convolutional, recurrent, transformer and graph neural networks, and we examine landmark clinical results in medical imaging, physiological signal analysis and electronic health record modeling. Comparison tables consolidate representative studies across modalities, tasks, architectures and reported outcomes, and mathematical formulations of the core models are provided to make the survey self-contained. We further review privacy-preserving and communication-efficient federated learning, blockchain-assisted data integrity, explainability techniques and the documented risks of bias and adversarial fragility. Finally, we identify open challenges in generalization, validation, deployment and regulation, and outline research directions toward multimodal foundation models and trustworthy clinical decision support. The survey spans foundational algorithms, landmark clinical evaluations and advances through 2026, and is intended as a structured entry point for researchers and practitioners.

Natural Resources for Human Health
eISSN: 2583-1194

1. INTRODUCTION

Healthcare generates data at a scale and variety that no other domain matches: radiological images, digital pathology slides, continuous physiological waveforms, genomic sequences, structured laboratory results and free-text clinical narratives. Turning this data into timely, reliable and equitable clinical decisions is the central promise of artificial intelligence (AI) in medicine [1-2]. Over the past decade, that promise has been substantiated by a series of landmark results in which learning systems matched or approached the performance of trained specialists on narrowly defined tasks, including dermatologist-level skin cancer classification [3], detection of diabetic retinopathy at ophthalmologist-level sensitivity and specificity [4], cardiologist-level arrhythmia detection from single-lead electrocardiograms [5] and radiologist-comparable breast cancer screening [6].

Terminology in this literature is layered. Artificial intelligence denotes the broad goal of machine behavior that would be called intelligent in a human; machine learning (ML) is the subset that improves from data rather than explicit programming; deep learning (DL) is the subset of ML built on many-layered neural networks; and intelligent predictive methods, as used in this survey, covers the growing class of systems that combine learned models with structural priors, optimization strategies or decision-making loops, such as attention, graphs, generative processes, reinforcement and federation. The boundaries matter practically, because the data volume, computational budget, interpretability and regulatory pathway of a solution differ sharply across these layers.

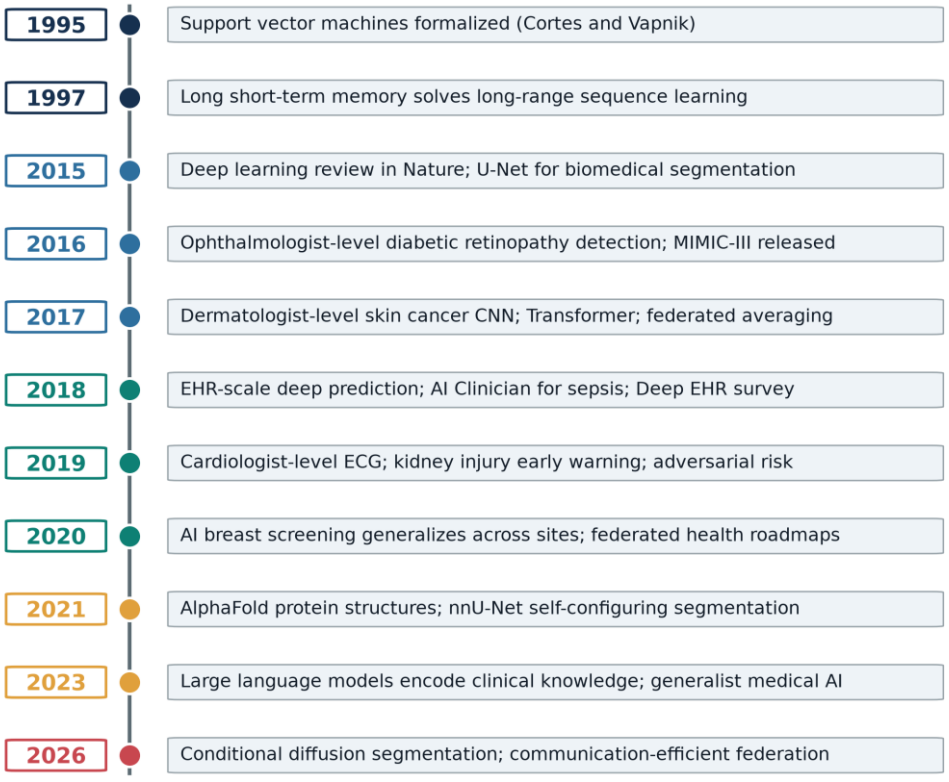

Figure 1. Milestones of artificial intelligence in healthcare.

Three forces converged to make the recent results possible. The first is algorithmic: deep learning replaced hand-crafted feature engineering with hierarchical representation learning, allowing a single architecture family to be reused across modalities [7-8]. The second is data availability: curated repositories such as the MIMIC-III critical care database [9] and large annotated imaging cohorts gave researchers realistic benchmarks on which methods could be compared. The third is computational: graphics processing units and, more recently, edge accelerators reduced training times from months to hours and enabled inference on portable clinical devices.

At the same time, the field has matured beyond single-task classification. Modern intelligent predictive methods integrate attention mechanisms that weigh clinically salient inputs [10], graph neural networks that model anatomical and functional connectivity [11-12], generative and diffusion models that synthesize and segment complex anatomy [13-14], reinforcement learning agents that recommend treatment policies [15] and federated optimization schemes that train across hospitals without moving patient data [16-17]. Large language models that encode clinical knowledge [18] and proposals for generalist medical foundation models [19] indicate the direction of the next decade. Figure 1 places these developments on a timeline.

Despite this progress, systematic reviews caution that many reported comparisons between algorithms and clinicians are methodologically fragile, with few externally validated or prospectively tested systems [20]. Documented failures of fairness [21] and robustness [22] have sharpened attention on trustworthy deployment rather than raw benchmark accuracy. A survey of the field therefore has to cover not only what the models are and what they achieve, but also how they are validated, explained, secured and scaled.

This paper contributes such a survey. Its specific contributions are as follows. First, it organizes the methodological landscape into classical machine learning, deep learning architectures and intelligent predictive frameworks, giving compact mathematical formulations for the models that recur across the literature. Second, it reviews representative applications in three clinical data regimes, namely medical imaging, physiological signals and electronic health records, and consolidates them in comparison tables that report modality, task, architecture and outcome. Third, it reviews the infrastructure of trustworthy medical AI, covering federated learning, privacy, security, integrity verification and explainability. Fourth, it distills open challenges and research directions. The remainder of the paper follows the organization shown in Figure 2.

2. SURVEY SCOPE AND ORGANIZATION

The survey covers peer-reviewed journal articles, landmark conference papers and high-impact clinical evaluations published between 1995 and 2026, selected to span the methodological spectrum from foundational algorithms [23-25] to recent conditional diffusion segmentation [14] and communication-efficient federated learning [26]. Studies were included when they either introduced a method that later became standard practice in health applications or reported a clinical evaluation that shaped subsequent research. Surveys of individual subfields, such as deep learning in medical image analysis [27], deep learning for electronic health records [28] and automated diabetic retinopathy screening [29], are cited as entry points for readers who need greater depth in one area.

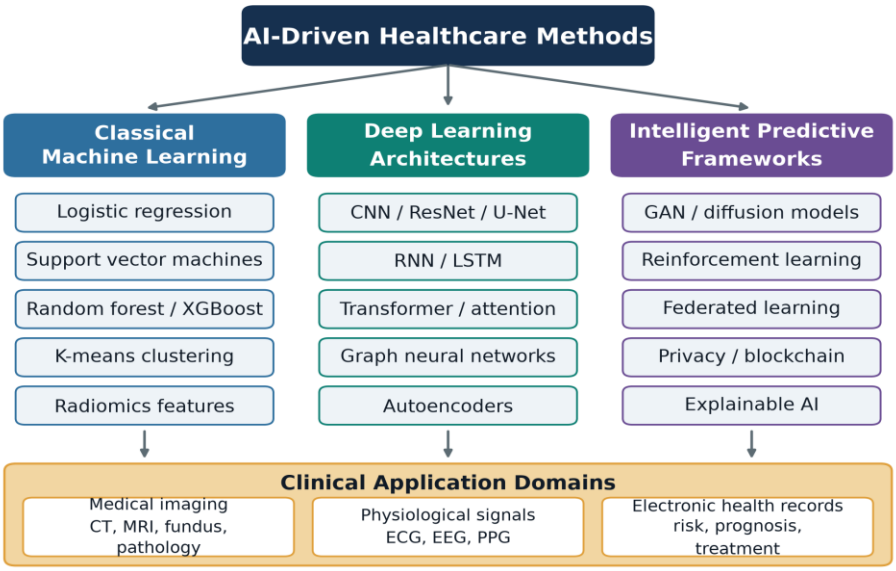

Figure 2. Taxonomy of methods and applications covered in this survey.

Figure 2 summarizes the organization. Section 3 reviews classical machine learning and the clinical prediction pipeline. Section 4 formalizes the deep architectures that dominate current practice. Sections 5 and 6 survey applications in medical imaging and in physiological signal and record-based prediction respectively. Section 7 covers federated, privacy-preserving and security-oriented methods. Section 8 addresses explainability and trustworthiness, Section 9 discusses open challenges and future directions, and Section 10 concludes.

Throughout the survey, reported performance figures are quoted as published in the cited studies. They are obtained on different cohorts, with different labeling protocols and evaluation splits, and are therefore indicative rather than directly comparable; the caveats in [20] apply to any cross-study reading of the tables.

3. CLASSICAL MACHINE LEARNING IN CLINICAL PREDICTION

3.1 The clinical prediction pipeline

Classical machine learning treats clinical prediction as a supervised learning problem over engineered features. A typical pipeline, shown in Figure 3, acquires raw data, cleans and normalizes it, extracts domain-informed features, trains a model with cross-validation and evaluates it on held-out patients before any deployment consideration. Although deep learning collapses the feature extraction stage into the model itself, the surrounding stages of the pipeline remain unchanged, and most reported clinical AI systems, including recent ones, still pass through this workflow.

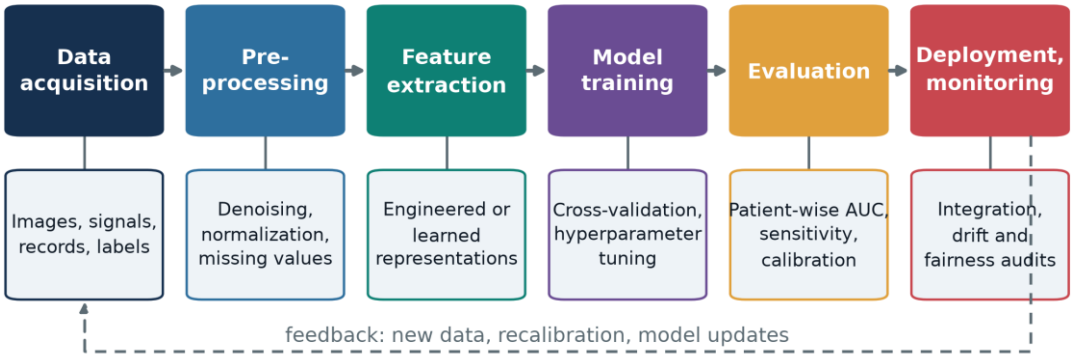

Figure 3. Standard machine learning pipeline for clinical prediction tasks.

Several stages of this pipeline carry disproportionate risk in healthcare. Data cleaning must respect the fact that missingness in clinical data is rarely random: a laboratory test is ordered because a clinician suspects something, so the presence of a value is itself informative, and naive imputation can erase or distort that signal. Feature engineering encodes domain knowledge, for example heart-rate variability statistics from beat intervals or texture descriptors from tumor regions, and its quality typically dominates the choice of classifier on small cohorts. Evaluation must be patient-wise rather than record-wise, since multiple records from one patient scattered across training and test folds leak identity-specific characteristics and inflate every reported metric. Finally, class imbalance is the norm rather than the exception, because disease prevalence is usually low; area under the receiver operating characteristic curve, precision-recall analysis and calibration curves are therefore preferred over raw accuracy, and resampling or cost-sensitive losses are routinely applied during training.

3.2 Core models

Logistic regression remains the reference baseline for clinical risk scoring because its coefficients admit direct clinical interpretation. For a feature vector $x \in \mathbb{R}^d$ with weights w and bias b, the probability of the positive class is

$$P(y = 1 \mid x) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \qquad (1)$$

Support vector machines (SVMs) extend linear classification with maximum-margin separation and kernel functions [23]. Given training pairs (x_i, y_i) with $y_i \in \{-1, +1\}$, the soft-margin SVM solves

$$\min_{w,b,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i \quad \text{s.t.} \quad y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i,\ \xi_i \geq 0 \qquad (2)$$

where $\phi(\cdot)$ maps inputs to a kernel-induced space and C trades margin width against training error. SVMs with radial basis kernels were for many years the strongest performers on small clinical datasets with hundreds rather than millions of samples.

Ensemble methods aggregate many weak learners into a strong one. Random forests build decorrelated decision trees on bootstrap samples and random feature subsets [24], while gradient boosting frameworks such as XGBoost fit trees sequentially to the residuals of the current ensemble with explicit regularization [30]. On tabular clinical data, laboratory panels, vital signs and demographic variables, boosted trees frequently remain competitive with or superior to deep networks while training in seconds and offering feature importance measures that clinicians can audit.

Unsupervised methods complete the classical toolbox. Clustering supports phenotype discovery and image quantification; a hybrid K-means pipeline, for example, has been used for volumetric quantification of lung tumors from computed tomography, converting cluster assignments into tumor volume estimates that track disease burden [31].

3.3 Representative clinical studies

Classical models continue to produce clinically useful systems, particularly where cohorts are small and features are well understood. In nephrology, comparative evaluations of machine learning models for kidney disease prognosis illustrate the standard methodology: multiple classifiers are trained on routine clinical and laboratory variables, tuned by cross-validation and compared on accuracy, sensitivity and specificity to select a deployable model [32]. In hematology, comparative implementations across blood cancer causes and disease categories similarly benchmark classical classifiers to establish which features carry diagnostic signal [33]. In non-invasive screening, photoplethysmography (PPG) signals acquired from a fingertip sensor have been used with signal-derived features to screen for diabetes and cerebral infarction, demonstrating that low-cost optical measurements carry systemic disease information [34]. Radiomics extends this feature-engineering philosophy to imaging: the spectral entropic radiomics feature extraction (SERFE) approach derives adaptive entropy-based descriptors from magnetic resonance images for glioblastoma classification [35]. Table 1 consolidates these and other representative classical studies.

Table 1. Representative classical machine learning studies in healthcare.

Study	Application	Method family	Data type	Reported outcome
[32]	Kidney disease prognosis	Comparative ML classifiers	Clinical and laboratory variables	Cross-validated model selection for prognosis
[31]	Lung tumor quantification	Hybrid K-means clustering	Thoracic computed tomography	Volumetric tumor estimates in a case study
[34]	Diabetes and cerebral infarction screening	Feature-based classification	Photoplethysmograph signals	Early non-invasive screening feasibility
[33]	Blood cancer analysis	Comparative classical classifiers	Hematological records	Benchmark of implementations across disease categories
[35]	Glioblastoma classification	Entropy-based radiomics features	Brain magnetic resonance imaging	Adaptive spectral entropic descriptors improve classification
[23-24, 30]	General clinical prediction	SVM, random forest, boosting	Tabular clinical data	Foundational algorithms underpinning the above studies

The durability of these methods is not an accident. They degrade gracefully on small samples, expose interpretable structure and impose modest computational cost, three properties that matter in hospital settings. On structured tabular problems with

dozens to hundreds of features, the practical ranking observed across the studies in Table 1 is remarkably stable: gradient-boosted trees and random forests lead, kernel SVMs follow closely, and logistic regression provides the interpretable floor that any more complex model must beat convincingly to justify its opacity. Their limitation is equally structural: performance is bounded by the quality of the engineered features, and no amount of classifier tuning recovers information that the features discard. It was precisely this bound that motivated the shift to representation learning surveyed next, where the feature extractor itself is learned from data.

4. DEEP LEARNING ARCHITECTURES FOR HEALTHCARE

4.1 From features to representations

Deep learning composes many parameterized nonlinear layers so that the model learns its own features from raw input [7]. Training minimizes an empirical loss by stochastic gradient descent; for a K-class problem with softmax outputs $\hat{y}$ and one-hot labels y, the cross-entropy objective over N samples is

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log\hat{y}_{ik} \qquad (3)$$

Figure 4 sketches the four architecture families that dominate healthcare applications: convolutional networks for images, recurrent networks for waveforms and longitudinal records, transformers for long-range dependencies across any tokenized modality, and graph neural networks for relational and connectivity data.

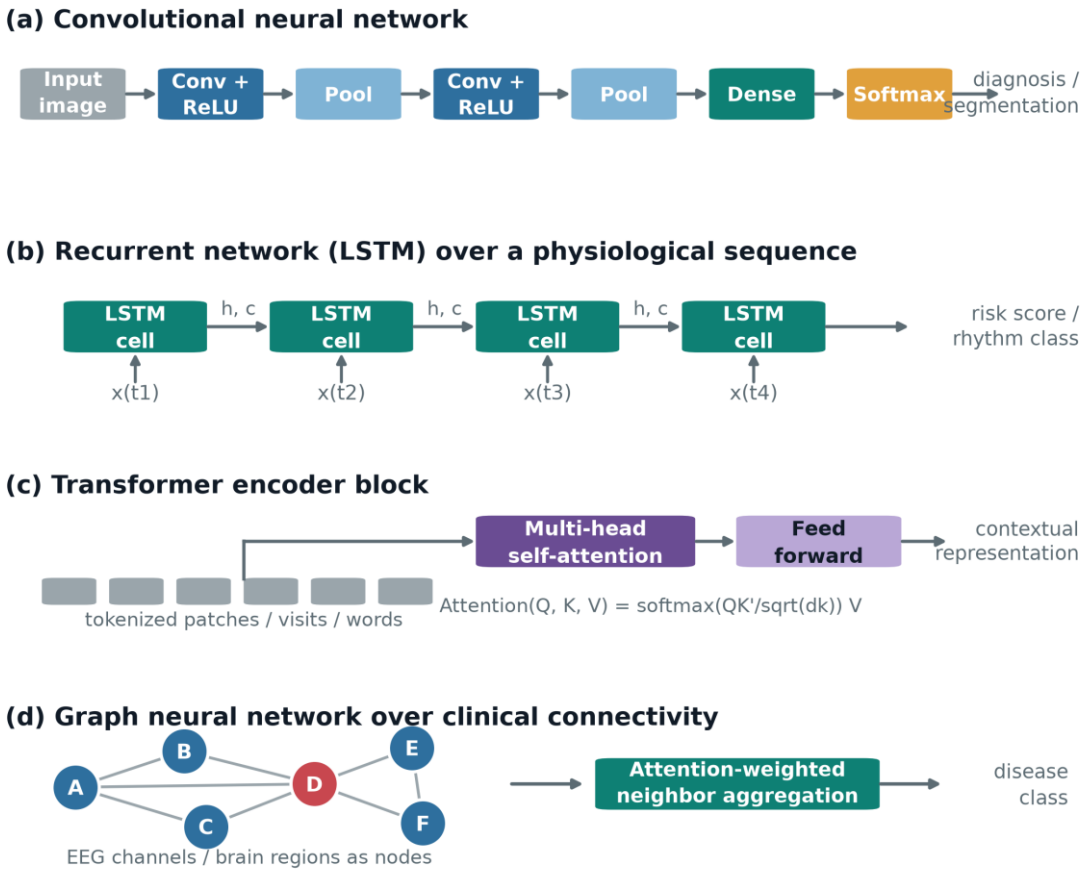

Figure 4. Principal deep architecture families used in healthcare.

4.2 Convolutional neural networks

Convolutional neural networks (CNNs) exploit the local structure of images by sliding learned kernels across the input. For an image I and a kernel K of size $m \times n$, the discrete convolution producing feature map F is

$$F(i,j) = (I * K)(i,j) = \sum_{u=1}^{m} \sum_{v=1}^{n} I\,(i+u, j+v)\,K(u,v) \qquad (4)$$

Stacked convolutions, pooling and nonlinearities yield features that progress from edges to organ-level structures. Weight sharing across spatial positions gives CNNs two properties that suit medical images: translation equivariance, so that a lesion is detected wherever it appears in the field of view, and a parameter count that depends on kernel size rather than image size, which keeps models trainable on the modest datasets typical of clinical research. In practice, medical imaging models are rarely trained from random initialization; transfer learning from large natural-image corpora, followed by fine-tuning on the clinical task, is the default recipe used by most of the landmark studies in Section 5, and aggressive data augmentation with rotations, flips, intensity shifts and elastic deformations substitutes for data the clinic cannot provide. Residual connections allow networks of great depth to be trained by learning modifications of the identity mapping [36], and encoder-decoder designs with skip connections, exemplified by U-Net, transfer fine spatial detail to the output for pixel-level segmentation [37]. The nnU-Net framework later showed that a self-configuring U-Net, with automated preprocessing and training heuristics rather than architectural novelty, wins or matches the state of the art across dozens of segmentation challenges [38], a result that reset expectations about where segmentation progress comes from.

4.3 Recurrent networks and temporal modeling

Physiological signals and patient histories are sequences, and recurrent neural networks (RNNs) process them by carrying a hidden state across time steps. The long short-term memory (LSTM) cell mitigates vanishing gradients with gated updates [25]. With input x_t, hidden state h_t and cell state c_t, the core update can be written compactly as

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c[h_{t-1}, x_t] + b_c), \quad h_t = o_t \odot \tanh(c_t) \qquad (5)$$

where f_t, i_t and o_t are sigmoid-activated forget, input and output gates. Bidirectional and stacked variants underpin much of the electrocardiogram, electroencephalogram and electronic health record literature surveyed in Section 6.

4.4 Attention and transformers

Attention mechanisms let a model weigh all positions of a sequence when encoding each position, removing the recurrence bottleneck. Scaled dot-product attention over query, key and value matrices Q, K, V with key dimension d_k is

$$\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (6)$$

The transformer built entirely from such attention blocks [39] now defines the state of the art in language modeling and, through vision variants that treat image patches as tokens, increasingly in imaging as well. Multi-head attention runs several such projections in parallel so that different heads can specialize, for example one head tracking rhythm morphology in an electrocardiogram while another tracks beat-to-beat intervals. Because attention is permutation-invariant up to positional encodings, transformers also tolerate the irregular sampling that characterizes clinical time series better than fixed-step recurrent models. In clinical time-series modeling, attention arrived earlier in interpretable form: RETAIN uses reverse-time attention so that the contribution of each past visit and variable to a prediction can be read directly from the attention weights [10]. Large language models trained on clinical corpora encode usable medical knowledge [18], and instruction-tuned successors have since reached expert-level medical question answering, with physician raters often preferring the model's long-form answers [40]. Transformer backbones are also the substrate of proposed generalist medical foundation models that would handle images, signals and text in one system [19].

4.5 Graph neural networks

Many healthcare structures are graphs: brain functional connectivity, molecular structures, hospital transfer networks and organ surface meshes. Graph attention networks (GATs) aggregate information from a node's neighbors with learned

attention coefficients [11]. For node features h_i, weight matrix W and attention vector a, the coefficient between nodes i and j is

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i \parallel Wh_j\right]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(a^T\left[Wh_i \parallel Wh_k\right]\right)\right)} \qquad (7)$$

A recent clinical instantiation is an adaptive multi-band dynamic functional-connectivity GAT that classifies Parkinson's disease from resting-state electroencephalography, reaching 97.31 percent accuracy by attending over frequency-band-specific connectivity graphs [12]. Graph regularization also strengthens image segmentation, as in conditional diffusion segmentation of pancreatic tumors where a graph term enforces anatomical coherence [14].

4.6 Generative, reinforcement and hybrid methods

Generative adversarial networks pit a generator against a discriminator to synthesize realistic data [13], and have been used in healthcare for augmentation, cross-modality synthesis and anomaly detection. Diffusion models, which learn to reverse a gradual noising process, currently lead in sample quality and have entered medical imaging as conditional segmentation and synthesis engines [14]. Generative modeling has itself reached foundation-model scale: a self-improving generative foundation model trained across organs and modalities produces synthetic medical images that measurably improve downstream diagnostic and prediction tasks [41]. Reinforcement learning frames treatment as sequential decision making; the AI Clinician learned sepsis fluid and vasopressor policies from intensive care records whose recommended actions were associated with the lowest observed mortality [15]. Beyond clinical data, deep learning has also transformed the biological substrate of medicine: AlphaFold predicts protein structures at near-experimental accuracy, accelerating drug target characterization [42].

5. APPLICATIONS IN MEDICAL IMAGING

5.1 Screening and diagnosis

Medical imaging is the domain where AI first reached specialist-level performance, and it remains the most thoroughly validated application area [27]. In diabetic retinopathy screening, a CNN trained on 128,175 graded fundus photographs achieved an area under the receiver operating characteristic curve (AUC) of 0.991 on the EyePACS-1 test set, with sensitivity and specificity selectable by operating point [4]; the surrounding literature on fundus analysis, including automated localization of anatomical landmarks such as the fovea, is reviewed in [43] and [29]. In dermatology, a CNN trained on 129,450 clinical images classified biopsy-proven malignancies on par with 21 board-certified dermatologists [3]. In digital pathology, the best algorithm in the CAMELYON16 challenge detected lymph node metastases with an AUC of 0.994, exceeding a panel of pathologists operating under time constraints [44]. In chest radiography, the CheXNeXt system matched radiologists on 10 of 14 thoracic pathologies in a retrospective comparison [45]. In screening mammography, an AI system evaluated on cohorts from the United Kingdom and the United States reduced false positives by 5.7 percent and false negatives by 9.4 percent on the United States data while generalizing across sites [6]. In lung cancer screening, an end-to-end three-dimensional CNN operating on low-dose computed tomography volumes reached an AUC of 0.944 and outperformed six radiologists when prior imaging was unavailable [46]. The screening paradigm is now extending to opportunistic use of routine imaging: a 2025 deep learning system detected gastric cancer from noncontrast computed tomography acquired for unrelated indications, sustaining an AUC of 0.927 on an 18,160-case external cohort from 16 centers and surfacing earlier-stage cancers in real-world deployment [47].

Reading these studies together, three design choices recur. Almost all of them fine-tune architectures pretrained on natural images rather than training from scratch, confirming that transferred low-level features survive the domain shift into medical imaging. All of them invest heavily in label quality, using multi-grader adjudication or biopsy-proven ground truth, because a network trained on noisy reads can only learn the noise. And the strongest evaluations compare against panels of clinicians on the same held-out cases rather than against historical report accuracy, which is the comparison regulators and clinicians actually credit.

A pooled reading of this literature is provided by a systematic review and meta-analysis of 82 studies, which estimated deep learning sensitivity at 87.0 percent and specificity at 92.5 percent, statistically comparable to health-care professionals, while

cautioning that only a small minority of studies used external validation or prospective designs [20]. Figure 5 summarizes reported AUC values from representative landmark studies; the caveat of non-comparable cohorts from Section 2 applies.

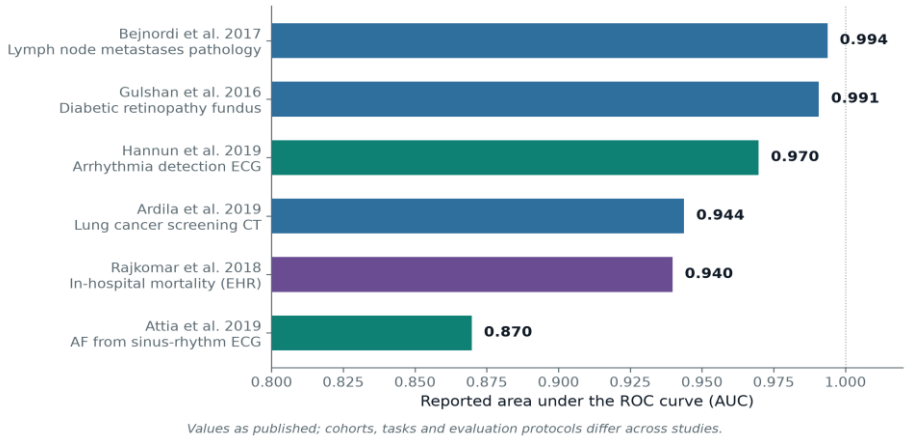

Figure 5. Reported area under curve in landmark clinical studies.

5.2 Segmentation and quantification

Where Section 5.1 concerns categorical decisions, segmentation assigns a label to every pixel or voxel and underpins radiotherapy planning, surgical navigation and quantitative follow-up. The U-Net encoder-decoder [37] and its self-configuring successor nnU-Net [38] define the baseline; nnU-Net alone achieved state-of-the-art results across 53 public segmentation tasks without manual tuning. Segmentation quality is typically measured by the Dice similarity coefficient between predicted region P and ground truth G,

$$DSC = \frac{2|P \cap G|}{|P| + |G|} \qquad (8)$$

which is also used directly as a differentiable training loss in most modern pipelines. Dice-based losses counteract the extreme foreground-background imbalance of medical segmentation, where a tumor may occupy a fraction of one percent of the voxels, a regime in which plain cross-entropy converges to predicting background everywhere. Annotation cost is the second structural constraint: expert contours are expensive, so the field increasingly leans on weak supervision from bounding boxes or image-level labels, semi-supervised consistency training over unlabeled scans, and self-supervised pretraining on unannotated archives before fine-tuning on the small labeled subset. Difficult low-contrast organs continue to drive methodological innovation: conditional diffusion models with graph regularization have recently been applied to pancreatic tumor segmentation on abdominal computed tomography, using the graph term to keep generated masks anatomically coherent [14]. Diffusion models also address the annotation bottleneck from the data side, with text-guided diffusion augmentation shown to generate controllable synthetic training images that consistently improve segmentation across ultrasound, computed tomography and magnetic resonance benchmarks [48]. Classical clustering remains relevant for quantification when annotated training masks are scarce, as in K-means-based volumetric lung tumor measurement [31], and radiomics descriptors such as SERFE convert segmented regions into diagnostic features [35]. Table 2 compares representative imaging studies.

Table 2. Deep learning studies in medical imaging.

Study	Modality	Task	Architecture	Reported outcome
[4]	Fundus photography	Diabetic retinopathy detection	Inception CNN	AUC 0.991 on EyePACS-1
[3]	Dermoscopic and clinical photos	Skin cancer classification	Inception v3 CNN	On par with 21 dermatologists

Study	Modality	Task	Architecture	Reported outcome
[44]	Digital pathology	Lymph node metastasis detection	Deep CNN ensemble	Best AUC 0.994, above pathologist panel
[45]	Chest radiography	14-pathology diagnosis	DenseNet-121 CNN	Matched radiologists on 10 of 14 findings
[6]	Mammography	Breast cancer screening	Deep CNN ensemble	False positives reduced 5.7 percent (United States)
[46]	Low-dose computed tomography	Lung cancer screening	Three-dimensional CNN	AUC 0.944 end to end
[47]	Noncontrast computed tomography	Opportunistic gastric cancer screening	Deep CNN	AUC 0.927 on 18,160-case external cohort
[37-38]	Multimodal	Biomedical segmentation	U-Net, self-configuring	State of the art on 53 tasks
[14]	Abdominal computed tomography	Pancreatic tumor segmentation	Conditional diffusion with graph regularization	Anatomically coherent masks on a difficult organ

The trajectory visible in Table 2 runs from supervised CNN classification toward architecture-light automated pipelines and generative segmentation, with the most recent work spending its novelty on data efficiency, anatomical priors and robustness rather than raw benchmark gains.

6. PHYSIOLOGICAL SIGNALS AND PREDICTIVE HEALTH ANALYTICS

6.1 Cardiovascular and neurological signals

Continuous physiological waveforms are natural inputs for temporal deep learning. A 34-layer CNN trained on 91,232 single-lead ambulatory electrocardiogram (ECG) recordings detected 12 rhythm classes with an average AUC of 0.97 and exceeded averaged cardiologist F1 performance [5]. Deep learning also extracts information invisible to human readers: an AI-enabled ECG identified patients with paroxysmal atrial fibrillation from sinus-rhythm recordings alone with an AUC of 0.87, effectively converting a routine 10-second test into a screen for an intermittent arrhythmia [49]. On the optical side, PPG signals support cuffless, low-cost screening for systemic conditions such as diabetes and cerebral infarction [34], a direction relevant to wearable deployment. Cross-modal learning now connects the two signal families directly: a 2025 framework maps wearable-acquired PPG to electrocardiography-level representations, improving cardiovascular disease prediction while synthesizing interpretable ECG waveforms from the optical signal [50]. In neurology, electroencephalogram (EEG) connectivity modeling with dynamic multi-band graph attention achieved 97.31 percent accuracy for Parkinson's disease classification [12], illustrating the migration of graph methods from network science into clinical electrophysiology.

Signal-based models face preprocessing choices that materially change outcomes. Filtering must preserve the diagnostic band of the signal, since aggressive baseline removal can destroy the morphology that carries the label; segmentation into fixed windows interacts with rhythm irregularity; and inter-patient evaluation, in which no subject contributes to both training and test sets, is the only protocol that predicts field performance. Wearable acquisition adds motion artifacts, sensor detachment and irregular sampling, which is why models validated on clean clinical recordings often degrade sharply on ambulatory data, and why the ambulatory training corpus in [5] was itself a major contribution.

6.2 Electronic health records and risk prediction

Electronic health records (EHRs) combine codes, laboratory values, medications and text over irregular timelines, and deep learning over such records is reviewed comprehensively in [28]. Unsupervised patient representations came first: Deep Patient showed that stacked denoising autoencoders over raw records produce features that improve prediction across dozens of future diseases [51]. Interpretable attention followed with RETAIN [10]. Scalability was demonstrated on 216,221 hospitalizations, where deep models over the complete raw record predicted in-hospital mortality with AUC up to 0.94 and

unplanned readmission with AUC 0.76 [52]. For deterioration prediction, a recurrent model over 703,782 veterans' records predicted 55.8 percent of inpatient acute kidney injury episodes up to 48 hours in advance [53], and in intensive care, reinforcement learning derived sepsis treatment policies associated with reduced observed mortality [15]. Modeling records raises problems that images and waveforms do not. Events arrive irregularly, so the time between observations is itself a feature; vocabularies of diagnosis and procedure codes are large, sparse and revised over time, which motivates learned code embeddings analogous to word embeddings; and the record reflects the process of care as much as the state of the patient, so a model can learn to predict mortality from the fact that comfort-care order sets were opened rather than from physiology. Guarding against such label leakage requires censoring the input window well before the prediction target and auditing the features that drive each prediction. Progress in this area is tightly coupled to open data, most prominently the MIMIC-III critical care database of over forty thousand intensive care stays [9], which has served as the common benchmark on which mortality, length-of-stay and intervention models are compared. Comparative classical baselines remain standard practice in disease-specific prognosis, as in kidney disease modeling [32]. Table 3 consolidates representative signal and record studies.

Table 3. Predictive studies on physiological signals and health records.

Study	Data	Task	Method	Reported outcome
[5]	Single-lead ambulatory ECG	12-class arrhythmia detection	34-layer CNN	Average AUC 0.97, above cardiologist F1
[49]	12-lead sinus-rhythm ECG	Atrial fibrillation identification	CNN	AUC 0.87 from normal-rhythm traces
[34]	Photoplethysmograph	Diabetes and cerebral infarction screening	Feature-based ML	Non-invasive early screening feasibility
[50]	Wearable PPG	Cardiovascular disease prediction	Cross-modal PPG-to-ECG deep model	Improved prediction with synthesized ECG
[12]	Resting-state EEG	Parkinson's disease classification	Multi-band dynamic GAT	97.31 percent accuracy
[51]	Raw EHR	Multi-disease risk representation	Stacked denoising autoencoders	Improved prediction across disease categories
[52]	Complete raw EHR	Mortality, readmission, length of stay	RNN and attention ensemble	AUC 0.94 mortality, 0.76 readmission
[53]	Longitudinal EHR	Acute kidney injury early warning	RNN	55.8 percent of episodes predicted 48 hours early
[15]	Intensive care records	Sepsis treatment policy	Reinforcement learning	Policies associated with lowest observed mortality

Two methodological lessons recur across Table 3. First, label leakage and record-wise rather than patient-wise data splitting inflate reported accuracy, so patient-independent evaluation is essential. Second, calibration and alert burden, such as the false alerts accompanying each true acute kidney injury warning in [53], determine clinical usability as much as discrimination does.

7. FEDERATED LEARNING, PRIVACY AND SECURITY

7.1 Federated and communication-efficient learning

Health data cannot in general be pooled centrally, for legal, ethical and competitive reasons. Federated learning (FL) trains a shared model while data remains at each institution: clients optimize locally and a server aggregates their updates. In the canonical FedAvg algorithm [16], with K clients holding n_k samples each and local weights w_{t+1}^k, the server computes

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k \qquad (9)$$

Surveys of digital health FL identify statistical heterogeneity across hospitals, system heterogeneity across devices and communication cost as the central obstacles [17]. Statistical heterogeneity is intrinsic to medicine: hospitals differ in scanners, protocols, case mix and coding culture, so client datasets are not independent and identically distributed, and naive averaging can converge to a model that serves no site well. Mitigations include client-specific fine-tuning layers, proximal regularization that keeps local updates near the global model, and clustering clients with similar distributions before aggregation. Communication cost is acute in wireless and bandwidth-constrained clinical networks, motivating communication-efficient federated learning (CEFL) designs: CEFL for computed tomography image classification compresses update traffic for bandwidth-constrained wireless healthcare networks [26], and a hybrid CEFL scheme combines gradient sparsification with quantization so that only a small fraction of significant gradient components, encoded at low precision, is exchanged per round for medical imaging tasks [54]. These engineering advances matter because they decide whether smaller clinics with limited connectivity can participate in collaborative model training at all. Beyond the algorithms, governance is emerging as its own research object: a 2025 scoping review of federated learning initiatives in healthcare maps the governance structures, stakeholder roles and policy gaps that determine whether multi-institutional training can be deployed responsibly [55]. Figure 6 shows the resulting architecture.

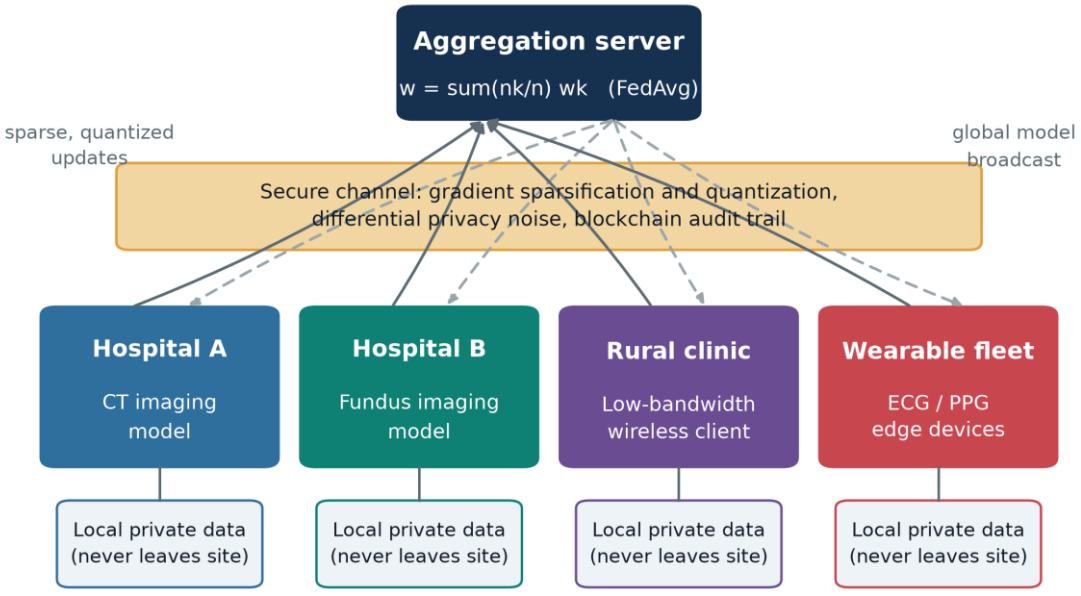

Figure 6. Privacy-preserving federated learning architecture for healthcare networks.

7.2 Privacy, integrity and security mechanisms

Federation alone does not guarantee privacy, since model updates can leak training data. Differential privacy provides a formal guarantee: a randomized mechanism M is (ϵ, δ)-differentially private if for all datasets D and D' differing in one record and all output sets S,

$$\Pr[M(D) \in S] \leq e^{\epsilon} \Pr[M(D') \in S] + \delta \qquad (10)$$

Differentially private stochastic gradient descent achieves this for deep networks by clipping and noising per-example gradients [56], and combined secure aggregation, encryption and privacy accounting frameworks for medical imaging are reviewed in [57]. Beyond confidentiality, integrity and availability of health-adjacent records must be protected. Blockchain-

assisted designs address tamper evidence: a deep learning and blockchain security framework protects cloud-hosted data by coupling anomaly detection with immutable audit trails [58], and two-tier integrity verification protects individualized education and therapy records in distributed systems serving special-needs populations [59], a record class that is medically sensitive yet frequently outside hospital security perimeters. Finally, deployed models are themselves attack surfaces: imperceptible adversarial perturbations can flip medical image classifications, with documented incentives for fraud in billing and approval pipelines [22]. Table 4 compares the surveyed protection mechanisms.

Table 4. Privacy, security and efficiency methods for healthcare AI.

Study	Setting	Technique	Primary contribution
[16]	Decentralized clients	Federated averaging	Foundational aggregation rule for FL
[17]	Digital health consortia	FL survey and roadmap	Obstacles and requirements for clinical FL
[26]	Wireless healthcare networks	Communication-efficient FL	Bandwidth-constrained CT classification
[54]	Federated medical imaging	Gradient sparsification plus quantization	Reduced update traffic per training round
[56]	Centralized training	Differentially private SGD	Formal (ϵ, δ) privacy for deep models
[57]	Medical imaging pipelines	Secure and private ML review	End-to-end privacy-preserving workflow
[58]	Cloud data protection	Deep learning with blockchain	Anomaly detection with immutable audit
[59]	Distributed therapy records	Two-tier integrity verification	Tamper-evident sensitive record storage
[22]	Deployed medical models	Adversarial risk analysis	Documented attack surface and incentives

8. EXPLAINABILITY AND TRUSTWORTHINESS

Clinical adoption depends on whether predictions can be examined, contested and audited. Post-hoc attribution methods dominate current practice: SHAP assigns each feature an additive contribution grounded in cooperative game theory [60], and attention-based architectures such as RETAIN expose visit-level and variable-level influence directly [10]. A contrary position holds that for high-stakes decisions inherently interpretable models should be preferred to explained black boxes, since post-hoc explanations can be unfaithful to the underlying computation [61]. Explanations must also be evaluated on the clinicians who receive them: a 2025 user study found that while model predictions reduced clinician estimation error, adding explanations helped some clinicians and worsened others while inflating their confidence, so explanation design cannot be validated on accuracy metrics alone [62]. The stakes are concrete: a widely deployed commercial risk score was found to systematically underestimate the illness of Black patients because it used healthcare cost as a proxy label for health need, and correcting the target definition nearly tripled the fraction of Black patients flagged for extra care [21]. Explainability, fairness auditing and careful label design are therefore not optional refinements but preconditions for safe deployment, a conclusion echoed across clinical AI reviews [1-2, 8].

Trustworthiness also requires knowing when a model does not know. Softmax confidence is a poor uncertainty measure, since deep networks are systematically overconfident on inputs far from their training distribution, which is exactly the situation a deployed clinical model encounters first. Calibration techniques such as temperature scaling, ensembling and Monte Carlo dropout produce probability estimates that can be compared against observed event rates, and selective prediction, in which the model abstains and refers difficult cases to a clinician, converts residual uncertainty into a safe workflow rather than a silent failure. Reporting standards are beginning to demand these analyses alongside discrimination metrics, and prospective deployments increasingly pair every model output with an uncertainty estimate and an audit log.

9. OPEN CHALLENGES AND FUTURE DIRECTIONS

The surveyed literature supports a consistent diagnosis: the accuracy problem is largely solved for narrow, well-labeled tasks, while the surrounding system problems, validation, governance, trust and integration, are not. This section organizes those problems and the research directions attached to them.

Generalization and validation. Most published systems are validated retrospectively on data resembling their training distribution; external, prospective and randomized evaluations remain rare [20]. Distribution shift across scanners, populations and coding practices routinely erodes reported performance, and patient-wise evaluation discipline is still not universal.

Data access and heterogeneity. Progress remains coupled to a small number of open resources such as MIMIC-III [9]. Federated learning eases access barriers but introduces statistical and system heterogeneity [17], and communication-efficient schemes [26, 54] are still maturing toward plug-and-play deployment.

Trust, fairness and security. Bias inherited from proxy labels [21], adversarial fragility [22] and the faithfulness limits of post-hoc explanation [61] together define a trustworthiness agenda that is as demanding as the accuracy agenda that preceded it.

Multimodal foundation models. Large language models already encode clinical knowledge at examination-passing and, more recently, expert levels [18, 40], and generalist medical AI proposes single models spanning images, signals, records and text [19]. Generalist diagnostic language models fine-tuned to emulate physician reasoning have begun to improve physician diagnostic performance in workflow studies [63]. Grounding such models, quantifying their uncertainty and regulating their updates are open problems.

Regulation and human factors. Software that adapts after approval sits uneasily in device regulation designed for static artifacts, and change-control plans for continuously learning systems are still being worked out by regulators. Equally decisive are human factors: alert fatigue from poorly calibrated early-warning systems, automation bias in which clinicians over-trust model output, and skill drift when screening is delegated to algorithms. Studies that measure clinician-plus-model performance, rather than model performance alone, remain scarce, yet the combined system is what patients actually experience.

Deployment economics. Inference cost, integration with hospital information systems, clinician workflow fit and post-deployment monitoring determine real-world impact; lightweight models, edge inference and automated pipelines such as nnU-Net [38] point toward sustainable deployment. Table 5 summarizes the challenge landscape.

Table 5. Open challenges and representative research directions.

Challenge	Manifestation	Representative directions
External validity	Performance drops under distribution shift	Prospective trials, multi-site evaluation [20]
Data governance	Siloed, privacy-restricted cohorts	Federated and communication-efficient learning [17, 54]
Privacy leakage	Updates and outputs reveal records	Differential privacy, secure aggregation [56-57]
Integrity and security	Tampering and adversarial inputs	Blockchain audit, robust training [58, 22]
Interpretability	Opaque high-stakes decisions	Attention, SHAP, interpretable models [10, 60-61]
Fairness	Proxy labels encode inequity	Label auditing, subgroup reporting [21]
Scale and scope	One model per task	Multimodal foundation models [18-19]

10. CONCLUSION

This survey traced artificial intelligence in healthcare from classical machine learning over engineered features, through the deep architecture families that now dominate imaging and temporal prediction, to the intelligent predictive frameworks, attention, graphs, diffusion, reinforcement and federation, that define current research. Landmark clinical results demonstrate specialist-comparable performance in retinal, dermatological, pathological, radiological and electrocardiographic tasks, while record-scale models forecast deterioration and mortality with useful lead time. The consolidated tables show a field whose recent gains come less from new benchmark records and more from data efficiency, anatomical priors, communication efficiency, privacy guarantees and interpretability. The open problems are correspondingly systemic rather than architectural: external validation, equitable labels, secure and private collaboration, and the grounding of emerging multimodal foundation models. Methods surveyed here, from communication-efficient federated learning to graph-regularized diffusion segmentation, indicate that the community is already reorienting toward these requirements. We expect the next phase of AI-driven healthcare to be judged not by leaderboard accuracy but by demonstrated, audited and sustained clinical benefit.

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, doi: 10.1038/s41591-018-0300-7.

[2] K. R. Radhika, S. Gupta, M. J. Oli, D. Sharma, H. Vinutha, and H. N. Shenoy, "Pancreatic tumor segmentation using conditional diffusion and graph regularization on abdominal computed tomography images," Journal of Intelligent Decision Making and Information Science, vol. 3, no. 1s, pp. 1123–1143, 2026, doi: 10.59543/jidmis.v3.489.

[3] Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115-118, 2017, doi: 10.1038/nature21056.

[4] S. Gupta, I. S. Rajesh, C. Singh, U. N. Ranjitha, K. Siddesha, and P. K. Sekharamantry, "A Hybrid CEFL Approach Using Gradient Sparsification and Quantization for Medical Imaging," International Journal of Computational Intelligence in Engineering (IJCIE), vol. 1, no. 1, pp. 1–15, 2026.

[5] Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nature Medicine, vol. 25, no. 1, pp. 65-69, 2019, doi: 10.1038/s41591-018-0268-3.

[6] S. M. McKinney et al., "International evaluation of an AI system for breast cancer screening," Nature, vol. 577, no. 7788, pp. 89-94, 2020, doi: 10.1038/s41586-019-1799-6.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015, doi: 10.1038/nature14539.

[8] Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24-29, 2019, doi: 10.1038/s41591-018-0316-z.

[9] E. W. Johnson, T. J. Pollard, L. Shen, L. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. A. Celi, and R. G. Mark, "MIMIC-III, a freely accessible critical care database," Scientific Data, vol. 3, 2016, Art. no. 160035, doi: 10.1038/sdata.2016.35.

[10] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, "RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism," in Advances in Neural Information Processing Systems, vol. 29, 2016, pp. 3504-3512.

[11] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," in Proceedings of the International Conference on Learning Representations (ICLR), 2018.

[12] S. Gupta, U. N. Ranjitha, J. G. S. Banu, K. L. Divya, T. R. Kulkarni, and S. Mannikeri, "Adaptive multi-band dynamic functional-connectivity graph attention network for Parkinson's disease classification from resting-state EEG,"

International Journal of Computer Information Systems and Industrial Management Applications, vol. 18, no. 11s, pp. 840-858, 2026.

[13] Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139-144, 2020, doi: 10.1145/3422622.

[14] R. Radhika, S. Gupta, M. J. Oli, D. Sharma, H. Vinutha, and H. N. Shenoy, "Pancreatic tumor segmentation using conditional diffusion and graph regularization on abdominal computed tomography images," Journal of Intelligent Decision Making and Information Science, vol. 3, no. 1s, pp. 1123-1143, 2026, doi: 10.59543/jidmis.v3.489.

[15] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, "The Artificial Intelligence Clinician learns optimal treatment strategies for sepsis in intensive care," Nature Medicine, vol. 24, no. 11, pp. 1716-1720, 2018, doi: 10.1038/s41591-018-0213-5.

[16] McMahan, E. Moore, D. Ramage, S. Hampson, and B. Aguera y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273-1282.

[17] N. Rieke et al., "The future of digital health with federated learning," npj Digital Medicine, vol. 3, 2020, Art. no. 119, doi: 10.1038/s41746-020-00323-1.

[18] Singhal et al., "Large language models encode clinical knowledge," Nature, vol. 620, no. 7972, pp. 172-180, 2023, doi: 10.1038/s41586-023-06291-2.

[19] Moor, O. Banerjee, Z. S. H. Abad, H. M. Krumholz, J. Leskovec, E. J. Topol, and P. Rajpurkar, "Foundation models for generalist medical artificial intelligence," Nature, vol. 616, no. 7956, pp. 259-265, 2023, doi: 10.1038/s41586-023-05881-4.

[20] X. Liu et al., "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis," The Lancet Digital Health, vol. 1, no. 6, pp. e271-e297, 2019, doi: 10.1016/S2589-7500(19)30123-2.

[21] 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, doi: 10.1126/science.aax2342.

[22] S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane, "Adversarial attacks on medical machine learning," Science, vol. 363, no. 6433, pp. 1287-1289, 2019, doi: 10.1126/science.aaw4399.

[23] Cortes and V. Vapnik, "Support-vector networks," Machine Learning, vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.

[24] Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5-32, 2001, doi: 10.1023/A:1010933404324.

[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[26] K. R. Radhika, H. N. Shenoy, T. R. Vinay, H. Pooja, D. Sharma, R. Priyanka, and S. Gupta, "Communication-efficient federated learning (CEFL) for CT image classification in bandwidth-constrained wireless healthcare networks," International Journal of Drug Delivery Technology, vol. 16, no. 13S, pp. 163-172, 2026, doi: 10.25258/IJDDT.16.13S.17.

[27] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sanchez, "A survey on deep learning in medical image analysis," Medical Image Analysis, vol. 42, pp. 60-88, 2017, doi: 10.1016/j.media.2017.07.005.

[28] Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589-1604, 2018, doi: 10.1109/JBHI.2017.2767063.

[29] Kiresur, I. S. Rajesh, and A. Bharathi Malakreddy, "Automatic detection of diabetic retinopathy in fundus image: A survey," in Proceedings of the International Conference on Smart Data Intelligence (ICSMDI 2021), 2021.

[30] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785-794, doi: 10.1145/2939672.2939785.

[31] U. N. Ranjitha and M. A. Gowtham, "Hybrid model using K-means clustering for volumetric quantification of lung tumor: A case study," in Machine Learning and Autonomous Systems: Proceedings of ICMLAS 2021, Singapore: Springer Nature Singapore, 2022, pp. 527-536.

[32] Y. M. Dalal, K. Amuthabala, and U. N. Ranjitha, "Precision medicine in nephrology: An exploratory study of machine learning models for accurate kidney disease prognosis," in Proceedings of the 2024 Second International Conference on Advances in Information Technology (ICAIT), vol. 1, 2024, pp. 1-6.

[33] T. R. Kulkarni, Bharathi, Gururaj, Aditi, and Jaiswal, "Comparison of implementation in blood cancer causes and diseases," International Journal of Trend in Scientific Research and Development, vol. 9, no. 1, pp. 503-512, 2025.

[34] T. R. Kulkarni and N. Dushyanth, "Early and noninvasive screening of common cardio vascular related diseases such as diabetes and cerebral infarction using photoplethysmograph signals," Results in Optics, vol. 3, 2021, Art. no. 100062, doi: 10.1016/j.rio.2021.100062.

[35] V. L. Sowmya, A. Bharathi Malakreddy, S. Natarajan, N. Prathik, and I. S. Rajesh, "Spectral entropic radiomics feature extraction (SERFE): An adaptive approach for glioblastoma disease classification," Frontiers in Artificial Intelligence, vol. 8, 2025, Art. no. 1583079.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[37] Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Lecture Notes in Computer Science, vol. 9351, Cham: Springer, 2015, pp. 234-241, doi: 10.1007/978-3-319-24574-4_28.

[38] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203-211, 2021, doi: 10.1038/s41592-020-01008-z.

[39] Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 5998-6008.

[40] K. Singhal, T. Tu, J. Gottweis, R. Sayres, E. Wulczyn, M. Amin, et al., "Toward expert-level medical question answering with large language models," Nature Medicine, vol. 31, no. 3, pp. 943-950, 2025, doi: 10.1038/s41591-024-03423-7.

[41] J. Wang, K. Wang, Y. Yu, Y. Lu, W. Xiao, Z. Sun, et al., "Self-improving generative foundation model for synthetic medical image generation and clinical applications," Nature Medicine, vol. 31, no. 2, pp. 609-617, 2025, doi: 10.1038/s41591-024-03359-y.

[42] J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," Nature, vol. 596, no. 7873, pp. 583-589, 2021, doi: 10.1038/s41586-021-03819-2.

[43] S. Rajesh, B. M. Arikerie, and B. M. Reshmi, "A review on automatic identification of fovea in retinal fundus images," International Journal of Medical Engineering and Informatics, vol. 12, no. 2, pp. 169-179, 2020.

[44] Ehteshami Bejnordi et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," JAMA, vol. 318, no. 22, pp. 2199-2210, 2017, doi: 10.1001/jama.2017.14585.

[45] Rajpurkar et al., "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," PLOS Medicine, vol. 15, no. 11, 2018, Art. no. e1002686, doi: 10.1371/journal.pmed.1002686.

[46] [46] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng, D. Tse, M. Etemadi, W. Ye, G. Corrado, D. P. Naidich, and S. Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," Nature Medicine, vol. 25, no. 6, pp. 954-961, 2019, doi: 10.1038/s41591-019-0447-x.

[47] Hu, Y. Xia, Z. Zheng, M. Cao, G. Zheng, S. Chen, et al., "AI-based large-scale screening of gastric cancer from noncontrast CT imaging," Nature Medicine, vol. 31, no. 9, pp. 3011-3019, 2025, doi: 10.1038/s41591-025-03785-6.

[48] Z. Zhang, L. Yao, B. Wang, D. Jha, G. Durak, E. Keles, et al., "DiffBoost: Enhancing medical image segmentation via text-guided diffusion model," IEEE Transactions on Medical Imaging, vol. 44, no. 9, pp. 3670-3682, 2025, doi: 10.1109/TMI.2024.3519307.

[49] Z. I. Attia et al., "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction," The Lancet, vol. 394, no. 10201, pp. 861-867, 2019, doi: 10.1016/S0140-6736(19)31721-0.

[50] Z. Ding, Y. Hu, Z. Li, Y. Mao, H. Li, D. Zhou, et al., "AI modeling photoplethysmography to electrocardiography useful for predicting cardiovascular disease," npj Digital Medicine, vol. 9, no. 1, Art. no. 61, 2025, doi: 10.1038/s41746-025-02228-3.

[51] Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep Patient: An unsupervised representation to predict the future of patients from the electronic health records," Scientific Reports, vol. 6, 2016, Art. no. 26094, doi: 10.1038/srep26094.

[52] Rajkomar et al., "Scalable and accurate deep learning with electronic health records," npj Digital Medicine, vol. 1, 2018, Art. no. 18, doi: 10.1038/s41746-018-0029-1.

[53] N. Tomasev et al., "A clinically applicable approach to continuous prediction of future acute kidney injury," Nature, vol. 572, no. 7767, pp. 116-119, 2019, doi: 10.1038/s41586-019-1390-1.

[54] Gupta, I. S. Rajesh, C. Singh, U. N. Ranjitha, K. Siddesha, and P. K. Sekharamantry, "A hybrid CEFL approach using gradient sparsification and quantization for medical imaging," International Journal of Computational Intelligence in Engineering, vol. 1, no. 1, pp. 1-15, 2026.

[55] Eden, I. Chukwudi, C. Bain, S. Barbieri, L. Callaway, S. de Jersey, et al., "A scoping review of the governance of federated learning in healthcare," npj Digital Medicine, vol. 8, no. 1, Art. no. 427, 2025, doi: 10.1038/s41746-025-01836-3.

[56] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 2016, pp. 308-318, doi: 10.1145/2976749.2978318.

[57] G. A. Kaissis, M. R. Makowski, D. Ruckert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, vol. 2, no. 6, pp. 305-311, 2020, doi: 10.1038/s42256-020-0186-1.

[58] N. Ajay, H. S. Mohan, I. S. Rajesh, and S. Gupta, "A deep learning and blockchain-based security framework for cloud data protection," Journal of Discrete Mathematical Sciences and Cryptography, vol. 29, no. 6, pp. 2553-2587, 2026, doi: 10.47974/JDMSC-2823.

[59] P. Guduru, K. V. Lakshmi, and S. Gupta, "Protecting individualized education and therapy records via two-tier integrity verification in distributed educational systems," International Journal of Special Education, vol. 41, no. 5s, pp. 386-401, 2026. [Online]. Available: https://internationalsped.com/index.php/ijse/article/view/2980

[60] M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 4765-4774.

[61] Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Machine Intelligence, vol. 1, no. 5, pp. 206-215, 2019, doi: 10.1038/s42256-019-0048-x.

[62] Nicolson, E. Bradburn, Y. Gal, A. T. Papageorghiou, and J. A. Noble, "The human factor in explainable artificial intelligence: Clinician variability in trust, reliance, and performance," npj Digital Medicine, vol. 8, no. 1, Art. no. 658, 2025, doi: 10.1038/s41746-025-02023-0.

[63] X. Liu, H. Liu, G. Yang, Z. Jiang, S. Cui, Z. Zhang, et al., "A generalist medical language model for disease diagnosis assistance," Nature Medicine, vol. 31, no. 3, pp. 932-942, 2025, doi: 10.1038/s41591-024-03416-6.