IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15446

Artificial Intelligence in Healthcare: A Comprehensive Survey on Disease Prediction

Buddareddy Sumeeth[1], Akula Sai Venkat[2], E Madhumitha[3], Aditya Santosh Naik[4],

Dr. Muhibur Rahman T.R[5]

6th Sem B.E.(CS&E), Ballari Institute of Technology and Management (BITM), Ballari, Karnataka – 583104, India[1-4]

Associate Professor, Department of Computer Science and Engineering,

Ballari Institute of Technology and Management (BITM), Ballari, Karnataka – 583104, India[5]

Abstract: In recent years, machine learning has found growing application in clinical settings, particularly for automating disease-level risk assessment tasks that were historically reliant on physician judgment alone. This paper surveys the current state of AI-driven healthcare systems, drawing on peer-reviewed publications indexed in IEEE Xplore, Springer, ScienceDirect, and PubMed. We reviewed work spanning core algorithmic approaches—including Random Forest, SVM, KNN, and various neural architectures—alongside applied systems for symptom triage, drug safety screening, and patient-facing health assistants. To bring structure to this diverse body of literature, we introduce a four-tier classification scheme organized around increasing system sophistication: from basic symptom-to-diagnosis mapping, through individualized care recommendations, into clinical support tooling, and finally toward fully integrated AI health platforms. Performance dimensions examined include classification accuracy, precision-recall balance, response latency, and scalability characteristics. A recurring pattern across reviewed studies is the absence of any single system that simultaneously covers real-time symptom intake, medication interaction checking, personalized guidance, and conversational AI interaction within one coherent architecture. We discuss the practical and theoretical implications of this gap and sketch directions for future work.

Keywords: Artificial Intelligence; Healthcare Systems; Disease Prediction; Machine Learning; Random Forest; Support Vector Machine; K-Nearest Neighbors; Neural Networks; Symptom Analysis; Drug Interaction; Medication Safety; Clinical Decision Support; Personalized Medicine; Healthcare Chatbot; Predictive Analytics.

I. INTRODUCTION

Medical diagnosis has long depended on direct clinician-patient interaction—a model that, while effective in well-resourced settings, breaks down where specialist access is scarce. Patients in rural districts, or those seeking care at off-peak hours, frequently encounter delays that meaningfully worsen outcomes. Over the past decade, machine learning has emerged as a partial answer to this gap, offering the ability to process structured patient data and return diagnostic probability estimates at a fraction of the time required for a full clinical workup.

Early applications were narrow. A model trained on ECG data could flag arrhythmia. One trained on retinal scans could detect diabetic retinopathy. The more ambitious goal of general-purpose disease prediction—where a user describes symptoms in natural language and receives a clinically-grounded response—required integrating multiple sub-problems: text understanding, probabilistic classification, drug safety lookup, and risk stratification. Progress toward that goal has been uneven, and no single reviewed system has achieved it fully.

This survey was undertaken to map that progress honestly. We pull from work published across IEEE, Springer, ScienceDirect, and PubMed, covering studies from 2015 through early 2025. Our focus is on systems where empirical performance metrics were reported—accuracy, precision, recall, or demonstrable real-world utility. Purely conceptual frameworks were reviewed but kept separate. Four main contributions structure our analysis: (1) a tiered classification of AI healthcare system types by functional depth; (2) a curated literature review of 15 representative studies; (3) a cross-paper comparative table examining methodology, performance, and limitations; and (4) a gap analysis identifying what remains undone and what the field needs next.

II. THEORETICAL BACKGROUND

Before comparing specific systems, it helps to establish the mathematical scaffolding common to most AI healthcare approaches. The following subsections lay out the key modeling primitives used across reviewed work.

DOI: 10.17148/IJARCCE.2026.15446

A. Machine Learning Model

At the most general level, supervised healthcare prediction reduces to learning a function f that maps patient input features X to a diagnostic output Y, given parameters θ:

$$Y = f(X, \theta) \quad (1)$$

$$\hat{Y} = \arg\max P(Y \mid X) \quad (2)$$

Here X encodes observable patient features—symptoms, lab values, demographic information—while θ is learned from training data by minimizing some prediction loss. Algorithms like Random Forest and SVM instantiate f differently but share this underlying formulation. The central engineering challenge is selecting features and training procedures that generalize beyond the specific cohort used for development.

B. Disease Prediction Models

Supervised classification is the dominant paradigm for disease prediction. Models commonly used include Random Forest (an ensemble of independently-grown decision trees), Support Vector Machine (which finds a maximum-margin separating hyperplane), K-Nearest Neighbors (which assigns labels by proximity in feature space), and various feedforward neural architectures. The predicted class probability is often expressed through a softmax formulation:

$$P(Y = c \mid X) = (1/Z) \exp(wc \cdot X) \quad (3)$$

where wc is the learned weight vector for class c and Z is a normalizing constant. In practice, tree-based methods tend to outperform this simple form on tabular clinical data.

C. Natural Language Processing (NLP)

Symptom-based systems face an additional challenge: converting free-text patient input into machine-readable feature vectors. NLP pipelines typically proceed through tokenization, stopword removal, and vectorization—either via classical TF-IDF weighting or, in newer systems, dense embeddings from pretrained language models. The resulting feature vector takes the form:

$$X = (x1, x2, x3, \ldots, xn) \quad (4)$$

Each xi corresponds to a symptom token or semantic feature. The quality of the NLP layer strongly influences downstream prediction—noise introduced at text parsing often cannot be corrected at the model level.

D. Performance Metrics

Standard evaluation follows the confusion matrix decomposition into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

$$Precision = TP / (TP + FP), \quad Recall = TP / (TP + FN) \quad (6)$$

In clinical settings, Recall is often prioritized—missing a true positive (a sick patient flagged as healthy) carries far higher cost than a false alarm. System designers must weigh this trade-off against the operational burden of excessive false positives, which can erode user trust.

E. Risk and Decision Models

Some systems go beyond point-estimate prediction and model disease risk as a posterior probability conditioned on observed symptoms:

$$Risk = P(Disease \mid Symptoms) \quad (7)$$

A threshold is then applied to assign cases to action categories—low-risk (self-care and watchful waiting), moderate-risk (scheduled monitoring), or high-risk (immediate consultation). The choice of thresholds is a design parameter that should ideally be calibrated on population-representative data rather than set arbitrarily.

F. System Performance and Scalability

Deployed healthcare systems face latency constraints not present in offline evaluation. Total response time can be decomposed as:

DOI: 10.17148/IJARCCE.2026.15446

$$Tresponse = Tprocessing + Tprediction \quad (8)$$

Where Tprocessing covers NLP and data normalization and Tprediction is the model inference time. In low-bandwidth clinical settings, even well-optimized models can fail usability thresholds if preprocessing is poorly engineered. Scalability under concurrent load is an additional concern that few academic papers address directly.

III. FOUR-TIER TAXONOMY

Reviewing the literature without an organizing framework makes comparison difficult. We propose classifying AI healthcare systems into four tiers, ordered by functional depth. The taxonomy was derived inductively from the reviewed papers rather than imposed from a prior theoretical framework.

Tier 1: Symptom-Level Prediction Systems

These are the most common type in the literature: a user submits a set of symptoms, and a classifier returns a ranked list of probable conditions. Common algorithmic choices at this tier include Random Forest, SVM, and KNN, all trained on labeled symptom-disease datasets. Tier 1 systems are computationally inexpensive and often achieve respectable accuracy on well-defined condition sets. Their limitations are equally clear—they treat each query in isolation, carry no memory of prior interactions, and cannot tailor output to individual patient context. For many practical deployments, however, this level of functionality is sufficient as an initial triage layer.

Tier 2: Personalized Healthcare Systems

Tier 2 systems extend the prediction pipeline with patient-specific context: age, gender, chronic history, medication load, and in some cases genetic markers. The core prediction models are similar to Tier 1, but the feature space is richer and the outputs are correspondingly more individualized. The key benefit is that two patients presenting the same symptoms may receive different risk assessments based on their histories—more clinically faithful than a uniform lookup. The practical costs are real, though: richer inputs require either user self-reporting (often incomplete) or integration with existing health records (often unavailable).

Tier 3: Clinical Decision Support Systems (CDSS)

Rather than serving patients directly, Tier 3 systems are designed as tools for practicing clinicians. They ingest structured clinical data—lab results, imaging metadata, procedure histories—and surface risk flags, guideline-concordant treatment suggestions, or differential diagnosis rankings. CDSS implementations have the longest history in the literature and the largest body of validation evidence. Adoption in practice remains uneven: systems that require tight EHR integration are costly to deploy, and some clinicians report alert fatigue when the system flags too frequently.

Tier 4: Intelligent Health Assistant Systems (Proposed)

No reviewed paper operated at this level in its entirety, which is itself a finding worth noting. A Tier 4 system would unify all previous capabilities within a single conversational interface: real-time symptom parsing via NLP, multi-condition probabilistic prediction, medication safety checking including pairwise drug interaction screening, personalized follow-up recommendations, and risk-stratified escalation guidance. The interaction layer would be chatbot-style, allowing follow-up queries and context accumulation across a session. Extensions envisioned for this tier include wearable data integration for continuous passive monitoring, multilingual interface support, and model updating as new interaction data accumulates. Whether such a system is achievable with current tooling and under realistic privacy constraints is the central open question the field has not yet answered.

IV. LITERATURE REVIEW

The 15 papers reviewed here were drawn from IEEE Xplore, Springer, ScienceDirect, PubMed, and arXiv. Selection criteria required that each paper report at least one quantitative performance metric (accuracy, AUC, precision/recall, or clinically-measured outcome) or, in the case of survey papers, provide substantial comparative evidence rather than pure description. Purely speculative or non-empirical works were excluded. Table I presents the full review summary.

TABLE I: LITERATURE REVIEW SUMMARY

Sl.	Author(s)	Year & Title	Method / Technique	Key Findings	Venue & Index
1	Nafiseh Nia et al.	2023 – AI Techniques in Disease Diagnosis	ML, DL models, data preprocessing	Strong prediction accuracy reported; data labeling and privacy remain open issues	Discover AI, 2023
2	J. Awwalu et al.	2015 – AI in Personalized Medicine	SVM, ANN, Fuzzy Logic	Tailored treatment plans showed measurable improvement in patient outcomes	IJCTE, 2015
3	Vadapalli et al.	2022 – AI for Personalized Medicine	ML with genomic data	Disease risk modeled from large genomic datasets with promising accuracy	Briefings in Bioinformatics
4	Wang & Torriani	2020 – AI in Body Composition	CNN, ANN, U-Net	Diagnostic accuracy in imaging tasks surpassed conventional methods	Seminars in Musculoskeletal Radiology
5	Faiyazuddin	2025 – AI in Healthcare	Literature + case studies	AI-assisted workflows cut diagnosis errors; data governance concerns persist	PMC
6	Nopour et al.	2025 – ML in Health Prediction	ML models	Preventive care systems benefited from ML-based risk stratification	ScienceDirect
7	Khalifa	2024 – AI Clinical Prediction	Predictive AI models	Clinician decision support improved; structured data inputs required	ScienceDirect
8	Islam et al.	2024 – Chronic Disease Prediction	RF, SVM, KNN	Multi-algorithm approach effective across several chronic conditions	Springer
9	Alhumaidi	2025 – ML for Healthcare Data	Big data + ML	Real-world hospital data handled efficiently; privacy safeguards needed	PMC
10	Frontiers AI Team	2024 – Symptom-Based Diagnosis	RF, Decision Tree, NLP	Symptom checker using NLP + RF outperformed single-model baselines	Frontiers AI
11	Wiedermann et al.	2023 – AI Symptom Checkers	Survey + case studies	AI symptom tools shown to cut avoidable outpatient visits meaningfully	J. Personalized Medicine

Sl.	Author(s)	Year & Title	Method / Technique	Key Findings	Venue & Index
12	Harada et al.	2024 – AI Diagnostic Accuracy	Observational study	Accuracy gains tied directly to training data volume and quality	JMIR Research
13	V. Jackins et al.	2021 – AI Smart Prediction (RF & Naive Bayes)	Random Forest, Naive Bayes	RF consistently outperformed Naive Bayes across diabetes, cardiac, cancer datasets	Springer
14	A. A. Soladoye et al.	2025 – Alzheimer's Prediction Optimization	RF, Feature Selection, Optimization	Achieved ~95% accuracy; feature pruning reduced compute without losing precision	Elsevier (ScienceDirect)
15	Shaheer Ahmad Khan et al.	2025 – Explainable Disease Surveillance	RF, XAI, EHR analysis	3–12 month early prediction demonstrated on real EHR data with XAI transparency	arXiv

Note: AI = Artificial Intelligence. ML = Machine Learning. DL = Deep Learning. NLP = Natural Language Processing. RF = Random Forest. SVM = Support Vector Machine. KNN = K-Nearest Neighbors. ANN = Artificial Neural Network. XAI = Explainable Artificial Intelligence. EHR = Electronic Health Records.

V. COMPARATIVE ANALYSIS

Several patterns surface when the reviewed papers are placed side by side. We organize observations around four recurring themes rather than proceeding study by study.

Baseline algorithms hold up well. Random Forest, SVM, and KNN appear across the majority of reviewed papers, and they consistently report accuracy figures in the 80–95% range on structured symptom or clinical datasets. Ensemble methods—especially Random Forest—show the strongest robustness against class imbalance, a persistent problem in disease datasets where positive cases are relatively rare. There is no evidence in the reviewed literature that newer deep architectures systematically outperform well-tuned tree methods on tabular health data, though imaging tasks are a different matter.

Optimization and feature engineering deliver meaningful gains. Studies such as Soladoye et al. [14] demonstrate that careful feature selection can raise accuracy into the 95% band while simultaneously reducing model size. This matters for deployment: smaller models are faster to retrain as new data arrives and cheaper to serve at scale. The implication for system designers is that investing in data preprocessing and feature pruning often yields larger returns than switching to a more complex architecture.

Real-world deployment surfaces problems that benchmark evaluation misses. Papers relying on Electronic Health Records—such as Khan et al. [15]—tend to report more modest accuracy figures than those using clean research datasets, and they also identify new failure modes: inconsistent field completion, non-standardized diagnostic codes, and population shift between training and deployment sites. These findings suggest that benchmark-era accuracy claims should be interpreted cautiously when generalizing to clinical deployment.

Integration depth remains the field's main limitation. Even the most sophisticated reviewed systems handle one or two capabilities well—prediction, or interaction, or personalization—but none strings all four together. Systems built around chatbot interfaces lack medication safety modules. Systems with strong drug interaction checking lack natural language symptom intake. The result is a fragmented landscape where healthcare workers or patients must switch between tools to cover their needs. Table II below documents these patterns systematically.

DOI: 10.17148/IJARCCE.2026.15446

TABLE II: COMPARATIVE ANALYSIS OF REVIEWED PAPERS

Sl.	Paper	Protocol / Technique	Performance	Advantages	Limitations	AI/ML?
1	Nia et al. [1]	ML & DL; data preprocessing	High	Early diagnosis enabled across multiple conditions	Demands large labeled data; annotation overhead significant	Yes
2	Awwalu et al. [2]	SVM, ANN, Fuzzy Logic	Moderate–High	Personalization of treatment plans	Real-world clinical deployment not validated	Yes
3	Vadapalli et al. [3]	ML with genomic data	High	Disease risk estimation from large biological datasets	Biological data complexity limits generalization	Yes
4	Wang & Torriani [4]	CNN, ANN, U-Net	High (Imaging)	Superior imaging diagnostics over manual methods	GPU-intensive; high infrastructure cost	Yes
5	Faiyazuddin [5]	Literature + case studies	Conceptual	Broad overview of AI impact on diagnosis workflows	No empirical model; privacy challenges unresolved	No
6	Nopour et al. [6]	ML prediction models	High	Preventive care enhanced via ML stratification	No live data integration mechanism described	Yes
7	Khalifa [7]	AI predictive models	High	Boosts clinician confidence in diagnosis	Only structured clinical inputs supported	Yes
8	Islam et al. [8]	RF, SVM, KNN	High	Works across multiple chronic disease types	Performance varies by dataset size and source	Yes
9	Alhumaidi [9]	Big data, ML	High	Handles messy, real-world hospital data	Data handling raises privacy and consent issues	Yes
10	Frontiers AI [10]	RF, Decision Tree, NLP	High	NLP interface improves symptom capture accuracy	Safety checking absent from the pipeline	Yes
11	Wiedermann et al. [11]	Survey + case studies	Conceptual	Reduction in unnecessary consultations documented	No integrated prediction or treatment module	No
12	Harada et al. [12]	Observational study	High	Accuracy scales with training data	Cross-demographic generalization untested	Yes
13	Jackins et al. [13]	Random Forest, Naive Bayes	High	Handles multiple diseases; RF robust to noise	Limited to classification; no chatbot layer	Yes

DOI: 10.17148/IJARCCE.2026.15446

Sl.	Paper	Protocol / Technique	Performance	Advantages	Limitations	AI/ML?
14	Soladoye et al. [14]	RF, Feature Selection, Optimization	High (~95%)	Optimization reduces model bloat without accuracy loss	Alzheimer-specific; not generalized to other diseases	Yes
15	Khan et al. [15]	RF, XAI, EHR analysis	High	Months-ahead prediction with explainability for clinicians	Requires structured EHR access; deployment-heavy	Yes

Note: AI/ML? column indicates whether machine learning or deep learning techniques are integrated into the system's core prediction, decision-making, or optimization pipeline.

VI. RESEARCH GAP

The survey reveals consistent patterns of omission across the reviewed body of work. Seven gaps are identified below, ordered roughly from the most practically urgent to the more systemic.

Gap 1 — No Fully Integrated Healthcare Platform: Every reviewed system addresses a subset of the required functionality. Systems with strong prediction accuracy lack conversational interfaces; systems with natural language intake lack medication safety checking. The composite system—prediction, drug interaction screening, personalized recommendations, and real-time interaction in one platform—has not been built. This is the most immediately actionable gap.

Gap 2 — Absence of Real-Time Adaptive Prediction: Reviewed systems take a snapshot of symptoms at query time and return a static estimate. None implements rolling risk assessment that updates as the user provides additional information, as vitals change, or as treatment progresses. Dynamic inference of this kind is technically feasible with streaming architectures but has not been applied in the healthcare AI literature reviewed here.

Gap 3 — Drug Interaction Safety Neglected: Despite medication error being a leading cause of preventable hospital harm, none of the reviewed prediction systems incorporates a drug interaction module. The omission is difficult to justify from a patient safety standpoint and represents a clear engineering gap rather than a fundamental research challenge.

Gap 4 — Shallow Personalization: Systems that call themselves "personalized" typically condition on age and gender—basic demographic slices. Behavioral data, chronic condition history, medication load, and lifestyle factors that substantially change disease risk are rarely incorporated. Until feature pipelines expand to include longitudinal patient context, personalization claims in this literature should be read conservatively.

Gap 5 — Privacy Engineering Underdeveloped: Several reviewed papers acknowledge that their systems require access to sensitive health records but treat privacy as a future concern rather than a design constraint. Federated learning, differential privacy, and secure multi-party computation are available techniques that could address this; they appear in almost none of the reviewed systems.

Gap 6 — Black-Box Models in High-Stakes Settings: The majority of reviewed systems use models—Random Forest, deep networks—that do not produce interpretable outputs by default. In clinical contexts, a prediction without an explanation is difficult for a physician to trust or act on. Explainable AI methods exist and are occasionally applied (see Khan et al. [15]), but remain the exception rather than the rule.

Gap 7 — Accessibility Constraints: Most systems in the reviewed literature implicitly assume users with reliable internet, functional digital literacy, and access to devices. Rural and low-income populations—who arguably have the most to gain from AI-assisted triage—are rarely considered in system design. Multilingual support and low-bandwidth operation modes are almost entirely absent.

VII. CONCLUSION

This survey covered 15 peer-reviewed papers on AI-driven healthcare systems, spanning roughly a decade of work from 2015 through early 2025. The reviewed literature demonstrates that machine learning—particularly Random Forest,

DOI: 10.17148/IJARCCE.2026.15446

SVM, and KNN on structured clinical data—can achieve strong classification accuracy for targeted disease prediction tasks. More recent papers extend this into personalized care, clinical decision support, and natural language symptom analysis, each representing a meaningful step toward practical deployment.

At the same time, the review makes visible a gap that individual papers naturally obscure: no system in the reviewed literature combines all the capabilities that a useful healthcare AI assistant would need. The four-tier taxonomy we proposed makes this gap concrete. Tiers 1 through 3 are populated with validated work; Tier 4 remains largely theoretical. The specific missing pieces are identifiable—a drug interaction module, a rolling risk update mechanism, a natural language interface, and a privacy-preserving data layer—and none of them represents an unsolved research problem in isolation. The gap is more architectural than algorithmic.

The research direction we see as most valuable from here is not further incremental accuracy improvement on existing benchmark tasks—that curve has flattened—but rather work on system integration: how to assemble validated components into a deployment-grade platform that maintains accuracy, handles real EHR data, explains its outputs to users, and operates within practical privacy and infrastructure constraints. Wearable device integration and continuous passive monitoring add a further dimension worth pursuing once the core platform is stable. These are engineering and systems challenges as much as machine learning ones, and the field will need researchers willing to engage at that level.

REFERENCES

[1]. N. G. Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," Discover Artificial Intelligence, vol. 3, no. 1, pp. 1–15, 2023.
[2]. J. Awwalu, A. G. Garba, A. Ghazvini, and R. Atuah, "Artificial intelligence in personalized medicine: Application of AI algorithms in solving personalized medicine problems," International Journal of Computer Theory and Engineering, vol. 7, no. 6, pp. 439–444, 2015.
[3]. S. Vadapalli, H. Abdelhalim, S. Zeeshan, and Z. Ahmed, "Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine," Briefings in Bioinformatics, vol. 23, no. 5, 2022.
[4]. B. Wang and M. Torriani, "Artificial intelligence in the evaluation of body composition," Seminars in Musculoskeletal Radiology, vol. 24, no. 1, pp. 30–37, 2020.
[5]. M. Faiyazuddin, "The impact of artificial intelligence on healthcare," PubMed Central (PMC), 2025.
[6]. R. Nopour, "Machine learning models in health prediction," ScienceDirect, 2025.
[7]. M. Khalifa, M. Albadawy, "Artificial intelligence for clinical prediction," ScienceDirect, vol. 134, pp. 102–110, 2024.
[8]. R Islam, Azrin Sultana, Mohammad Rashedul Islam, "Chronic disease prediction using machine learning," Springer, vol. 12, no. 3, pp. 210–225, 2024.
[9]. N. H. Alhumaidi, "Machine learning for real-world healthcare data," PubMed Central (PMC), 2025.
[10]. Frontiers AI Research Team, "Enhancing diagnostic accuracy in symptom-based health checkers: A comprehensive machine learning approach," Frontiers in Artificial Intelligence, vol. 7, 2024.
[11]. C. J. Wiedermann, A. Mahlknecht, G. Piccoliori, and A. Engl, "Redesigning primary care: The emergence of artificial-intelligence-driven symptom diagnostic tools," Journal of Personalized Medicine, vol. 13, no. 3, 2023.
[12]. Y. Harada, T. Sakamoto, S. Sugimoto, and T. Shimizu, "Longitudinal changes in diagnostic accuracy of an AI-based symptom checker," JMIR Formative Research, vol. 8, 2024.
[13]. V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, "AI-based smart prediction of clinical disease using random forest and naive Bayes," The Journal of Supercomputing, vol. 77, pp. 5198–5219, 2021.
[14]. A. A. Soladoye, N. Aderinto, D. Osho, and D. B. Olawade, "Enhancing Alzheimer's disease prediction using optimization techniques," ScienceDirect, 2025.
[15]. S. A. Khan, M. U. Shahid, A. Abdullah, I. Hashmat, and M. Farooq, "Explainable disease surveillance system for early prediction of multiple chronic diseases," arXiv:2501.15969, 2025.