Vantage: Journal of Thematic Analysis
(Vantage J. Them. Anal.) ISSN(E): 2582-7391 Vol. 7(1), April, 2026, 60-80

Review Article

Understanding the Advancement of Artificial Intelligence in Healthcare

Sanjana Yadav[1], Megha Khanna[2],

[1] Department of Computer Science, University of Delhi, India.

[2] Department of Computer Science, Sri Guru Gobind Singh College of Commerce, University of Delhi, India.

Corresponding Author: sanjanaphd25@cs.du.ac.in

Received: 18/02/2026 | Accepted: 12/04/2026 | Published: 05/05/2026

ABSTRACT

Artificial Intelligence (AI) has seen tremendous progress over the years. This growth in AI technology has significantly impacted almost every industry. One of the major beneficiaries of this rapid expansion in AI knowledge is the healthcare industry. This review presents a structured and evolutionary compilation of AI applications in the healthcare domain, tracing three major eras of AI: traditional machine learning models, deep learning models and foundational and multimodal models. This review focuses on representative empirical studies that introduced paradigm shifts in modeling, validation, and deployment. We analyze key application domains including diagnostics, patient management, operational efficiency, and biomedical discovery, while critically examining methodological limitations across eras.

Keywords: Artificial Intelligence, Federated Learning, Healthcare, Disease Prediction, Protein Structure

1. INTRODUCTION

AI is the biggest technical revolution since the industrial revolution and onset of information technology. Most of the industries have embraced this new technology from transportation to information technology. Even though the healthcare industry is very heavily regulated, the use of AI in this field has seen a huge leap of about 4000 percent from 2015 to 2023 (Stanford HAI, 2025). This paper summarizes the advancement of the AI hierarchy from classical machine learning model to advanced models. Each level added a better outcome in precision in predicting and detecting diseases (Liu et al., 2013; Polat et al., 2007; Q. Zou et al., 2018), surgical operations' success rates (Fagerstrom et al., 2019; Nemati et al., 2018). As the capturing of better datasets (Cameron et al., 2012; Shen et al., 2025) has improved the efficacy of modern tools used in healthcare, AI models have also seen a great improvement in their performance. Primitive AI models focused on predictions of diseases (Liu et al., 2013; Polat et al., 2007; Q. Zou et al., 2018), and better management of hospital admissions (Futoma et al., 2015) among other uses. As AI models became more sophisticated, the data storage in the healthcare industry also became more structured. This gave rise to the implementation of newer AI models that could be trained on the handwritten clinical notes as well images (Moor et al., 2023; Yang et al., 2022). This has brought a shift in approach to early detection of terminal diseases (Liang et al., 2010; Liu et al., 2013; Polat et al., 2007; Rajpurkar et al., 2017) as well as lifestyle problems (Gulshan et al., 2016; Q. Zou et al., 2018). As the AI architecture became computationally intensive, the paradigm also shifted from simple outcome prediction to knowledge generation (Jin et al., 2020; Jumper et al., 2021; Lin et al., 2023). As the AI models become more sophisticated in generating contextual text, they have further democratized health literacy. These new age models help patients understand the complex medical terminologies and better understand the disease and its symptoms (Alsentzer et al., 2019; Moor et al., 2023; Yang et al., 2022). For doctors, they help centralize access to emerging research.

With rapid progress in AI, the literature seems fragmented, often organized by specific seminars or tasks rather than comprehensive compilation of methodological evolution. This review paper aims to bridge this gap by providing chronological and conceptual application of AI in healthcare.

The format of the review paper follows this pattern: Section 2 explores related studies that are done to understand the advancements of AI. Section 3 describes the selection criteria for literature which is included in this paper. Section 4 talks about traditional machine learning models and provides a review of their use in the healthcare industry. Section 5 discusses early deep learning models and how these are implemented in accordance with the healthcare industry. Section 6 moves to the current stage of technology and its application in the healthcare industry. Section 7 sheds light on the emerging trends and technology in the AI landscape. Section 8 talks about the challenges in adoption of AI technology and section 9 concludes the paper.

2. RELATED WORK

Many studies have been published that have shown the advancements of AI in healthcare. Kitsios et al. (2023) conducted a review study on the recent breakthroughs in AI in healthcare. In this study, Kitsios et al. (2023) did a systematic literature review where they classify the existing literature into four broad dimensions. These aspects include the use of AI in healthcare, the benefits and drawbacks for the healthcare industry, ethical concerns around AI, and social sustainability with AI. The literature that concentrates on the application of AI technology to assist clinics, patients, and medical professionals makes up the dimension of healthcare activity utilizing AI. The positive and negative effects of AI technology on the healthcare sector are the subject of the second dimension of benefits and drawbacks for the sector. Early diagnosis is one of the many benefits that AI offers to people (Liu et al., 2013; Rajpurkar et al., 2017), as well as patient monitoring (Futoma et al., 2015; Nemati et al., 2018) and drug discovery (Jin et al., 2020; Jumper et al., 2021; Lin et al., 2023).

If the data gathered from hospitals and clinics is of poor quality, most AI applications in healthcare could have serious detrimental effects (Goldberg et al., 2010; Hasan et al., 2017; Ling et al., 2013). The third dimension comprises literature about the increasing discussions on ethics of AI in healthcare research (Floridi et al., 2018). Kitsios et al. (2023) in their literature review, categorized social sustainability and AI as the fourth dimension. When using AI technology, one should consider its impact on society and the planet (Kitsios et al., 2023). While Kitsios et al. (2023) discuss the advancement of AI in healthcare, they do not organize AI technology according to its evolutionary progression.

Kalra et al. (2024) in their study, categorized AI advancement based on diagnostic techniques. Their study focuses on the three divisions of diagnostic methods: medical imaging diagnostic, clinical workflow in digital pathology, and predictive diagnostic techniques. These divisions take into consideration image segmentation and classification under medical imaging diagnostic, automated analysis of cytology in clinical workflow in digital pathology and predictive analytics offers a valuable technique for physicians by

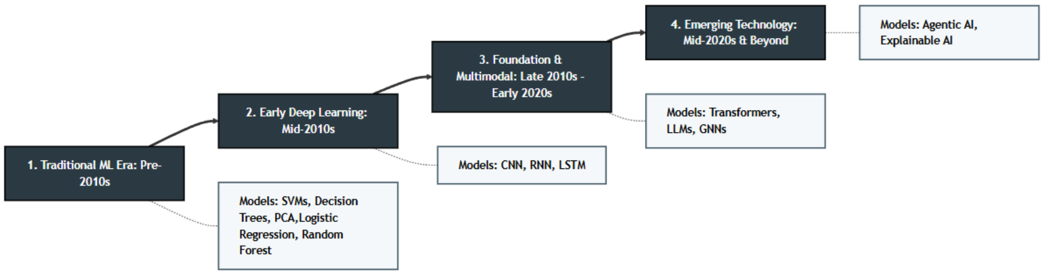

Figure 1: Visual representation of the four categories.

identifying critical complication patterns. A categorical comprehension based on diagnostic techniques make this study important in order to understand how AI has helped develop better diagnostic practices. However, this paper fails to address the current stage of AI development. This study does not track evolution of better AI technologies and their applications in the healthcare industry.

Recent studies involving agentic AI have shown that agentic AIs are very good at external knowledge gathering and answering medical questions (Vatsal et al., 2026). This will significantly relieve the medical practitioners' workload related to management and give them ample time to interact and educate the patients even more (J. Zou & Topol, 2025). Neupane et al. (2025) in their review paper show that agentic AI can be trained to be HIPAA compliant. These agentic AI models are sensitive while handling protected health information. This shows that regulatory compliance through dynamic context-aware policy can be enforced (Neupane et al., 2025)

Similar studies have been in circulation that either emphasize the potential of use of AI in the healthcare industry or role of AI in this industry (Al Kuwaiti et al., 2023; Davenport & Kalakota, 2019). Most of these studies help us in understanding the extensive use of AI and its challenges in the healthcare industry. This paper focuses not only on the use of AI but also tries to tie its use in the healthcare industry along with AI's evolution. By tracing AI's evolution along with its application in healthcare, this review paper connects the methodological paradigm shifts to clinical breakthroughs. This review paper not only addresses the paradigm shift in AI but it also discusses their limitations.

3. LITERATURE SELECTION

The review adopts a selective and representative approach to literature selection. Applying this approach, the review aims to trace key methodological and conceptual transitions in AI for healthcare. Sources for literature were selected after conducting a thorough search on Google Scholar, Nature Journal, PubMed, IEEE Explore, JAMA Network and arXiv. The queries used were "Artificial Intelligence in healthcare," "Machine learning disease prediction models," "Deep Neural Networks in Disease Prediction," "Disease classification using Neural Networks," "Early disease prediction machine learning models," "Federated learning in healthcare," "Multimodal AI in healthcare," "Computer vision in healthcare," "Use of transformers in healthcare," "Drug discovery using transformers," "Application of foundation model in healthcare," "Use of AI in healthcare," "Explainable AI CVD," "Big data in Healthcare" and "Agentic AI in healthcare." The selected literature is organized into four categories - traditional machine learning, early deep learning, foundation and multimodal models and emerging trends and technology. These categories reflect AI's evolution. Figure 1 shows the timeline and evolution of these categories.

3.1 Inclusion and exclusion criteria

Studies included are based on their empirical knowledge, methodological significance, new modelling paradigm shifts, validation of quality and relevance to real time application in healthcare. The empirical studies included in the paper that are related to application of AI in healthcare range from 2007 to early 2026.

Inclusion criteria

- Empirical studies addressing the studies in human healthcare only.
- Empirical studies applying AI technology to clinical, medical or biomedical domains.
- Empirical studies demonstrating AI's application in disease prediction, diagnostics, biology of proteins, and drug discovery along with hospital management.
- Empirical studies addressing the challenges and uses of explainable AI.
- Empirical studies that demonstrate use of Agentic AI in healthcare.
- Empirical studies that establish the architectural background of AI technologies.

Exclusion criteria

- Studies in AI applications that are not directly related to healthcare.
- Studies related to healthcare that did not involve AI applications.
- Studies that were not written in the English language.
- Similar studies carried out by the same authors.

4. TRADITIONAL MACHINE LEARNING ERA

Traditional machine learning models are primarily simple, rule-based algorithms. These models stem from very early regression techniques derived from statistics that have their foundation in the 1800s (Wolberg, 2010) and their use in training of machine learning models as early as the 1960s (Samuel, 1959). With the advancement of AI, healthcare also adapted to the integration of progressive informational technology infrastructure in every field and started gathering more organized clinical data (Shen et al., 2025).

4.1 Models

The models in this era used were regression models, decision trees, random forest, Support Vector Machines (SVMs), Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The models like SVMs are supervised machine learning models applied to tasks involving regression and classification. These models were good at handling structured tabular datasets and predicting outcomes that included statistical linear classification or prediction of probability. These models operated on structured or semi-structured data gathered through lab results, physical examination, or engineered features extracted from clinical data. The datasets were compiled from Electronic Health Records (EHR) (Futoma et al., 2015), specific biomarkers related to cardiac health (Liu et al., 2013), brain activity (Bron et al., 2015) and prediabetic conditions (Q. Zou et al., 2018). These algorithms solved simple vector equations that did not require high performance Graphic Processing Units' (GPUs) computational prowess. Apart from the supervised learning models, there were simple unsupervised machine learning models that implemented k- nearest neighbor algorithm to classify patient clusters (e.g., heart disease risk) based on feature similarity (Polat et al., 2007).

4.2 Related Research

Empirical studies during this era focused mainly on predicting diseases. Prominent application areas for these predictions included detection of sepsis (Nemati et al., 2018), breast cancer (Liang et al., 2010), heart diseases (Ling et al., 2013;

Table 1: Representation of Empirical research in implementing traditional machine learning models in healthcare.

Study	Research Area	Model Used	Dataset Used	Key Result	Performance Metrics
(Futoma et al., 2015)	30-day hospital readmission prediction	Logistic Regression, Random Forest	New Zealand National Minimum Dataset consisting of nearly 3.3 million hospital admissions	Established rigorous benchmarking practices for readmission prediction and highlighted limits of structured EHR features	The mean Area Under Receiver Operating Characteristics Curve (AUC) of the best method is 0.69, ranging from 0.57 to 0.95
(Nemati et al., 2018)	Early sepsis prediction in ICU	Machine Learning Based Classifier Model	ICU admissions at two Emory University hospitals and the Medical Information Mart for Intensive Care-III ICU database, which is openly accessible	Demonstrated accurate prediction of the onset of sepsis in an ICU patient 4–12 hours prior to clinical recognition	The model achieved AUC in the range of 0.83-0.85
(Bron et al., 2015)	Alzheimer's disease diagnosis from MRI	Support Vector Machines (SVM)	Alzheimer's disease (AD) patient classification using MRI data from the Alzheimer's disease neuroimaging initiative	Illustrated effectiveness of feature-engineered ML for medical imaging	The AUC increased from 90.3% without selecting any feature to 92.0% when selecting 1.5-3% feature selection.
(Liu et al., 2013)	ECG arrhythmia classification	Support Vector Machines (SVM)	MIT-BIH Arrhythmia Database	Showed integration of signal processing and ML for automated bio signal diagnosis	According to the classification results, 96.4% accuracy was attained.
(Q. Zou et al., 2018)	Diabetes mellitus prediction	Decision Trees, Random Forest	Hospital physical examination data in Luzhou, China	Demonstrated scalability of tree ensembles for chronic disease risk prediction	Prediction with random forest has the highest accuracy of 0.81

Continued....

Study	Research Area	Model Used	Dataset Used	Key Result	Performance Metrics
(Polat et al., 2007)	Automatic detection of heart disease	KNN based classifier with Fuzzy Resource Allocation	Cleveland Clinic Foundation Database	First to successfully integrate Fuzzy Logic into the immune-inspired AIRS model	Classification accuracy of 87%
(Liang et al., 2010)	Identify protein profile changes between cancerous and normal breast tissues	Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)	Breast cancer patients from Penang Hospital, Malaysia	Regardless of the grade, stage, and receptor status of the tissues, as well as the patient's race, the LDA showed that the classification of cancerous and normal tissues using a combination of biomarkers can be carried out with high accuracy levels. Six of the ten PCs that were produced by PCA of the entire data set were able to reliably distinguish between normal and cancerous tissue.	88.9% is the accuracy of the cross-validated samples classified, compared to 91.7% of the initial grouped samples.
(Chou et al., 2024)	Forecast patient death in the initial stages of intensive care unit admission	XGBoost and Random Forest	Electronic medical Records from 12377 patients admitted to tertiary medical center in Taiwan from 2009 to 2018	Machine learning models are more accurate than traditional scoring systems	AUC of XGBoost and Random Forest = 0.88 and 0.87 and it was better than APACHE 0.74 and SOFA 0.75

Polat et al., 2007) and Alzheimer's induced dementia (Bron et al., 2015). The research also focused on early identification of diabetes mellitus from captured biomarkers (Q. Zou et al., 2018). Furthermore, the research area also focused on the study of EHR and predicting the readmission rates in the hospital (Futoma et al., 2015). Studies typically aimed to demonstrate that machine learning models could outperform traditional rule based clinical scoring systems like Sequential Organ Failure Assessment (SOFA) or Acute Physiology and Chronic Health Evaluation (APACHE) scores) (Chou et al., 2024). The summarization of aforementioned research works is tabulated in table 1.

4.3 Limitations

Despite the breakthroughs in the healthcare industry brought by the classical ML Models, the adoption of these techniques was very limited. The predictions made by the models usually didn't fit the real-world data. The models were very well- trained on the data set and predicted the desired outcomes, but the scope of predictions was very limited. Since the training dataset was either semi-structured or limited, these models usually faced the problem of overfitting. These are sensitive to class imbalance, often requiring oversampling or under-sampling. Since these models relied heavily on properly documented structured tabular datasets, they were not suited to handle unstructured free-text clinical notes or raw images. The models from this era are only capable of getting trained in a single mode of input, which was mostly structured datasets. This limited capability restricted the clinical utility of these models and restricted their widespread use in the industry. These constraints ultimately motivated the transition to better machine learning models that involved deep learning approaches which can understand more features and learning richer representation directly from the raw clinical notes or images.

5. EARLY DEEP LEARNING MODEL ERA

As the traditional machine learning models were facing challenges with comprehending unstructured data and images, this model paved the way for newer and better tools in the form of early deep learning models.

5.1 Models

In the early deep learning era, the breakthrough proposed models were Convolutional Neural Network (CNN) (LeCun et al., 1989), Recurrent Neural Network (RNN) (Elman, 1990) and RNN based Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) models. These models are designed to mimic human neurons by implementing multiple hidden layers of the basic building block of any neural network, a perceptron (Fukushima, 1980). Like the human neuron, perceptron also consists of an input layer, responsible for handling input, a hidden layer which consists of a linear function followed by a non-linear activation function (LeCun et al., 1989). CNNs are the models used for image-based prediction. These models extract features on their own from the unstructured raw pixel data (LeCun et al., 1989).

For time-series and sequential data, RNNs are used. An RNN model mostly works like a CNN model as it also performs convolution on matrices. But these models also have an added element of memory. This new element of memory gives these models the ability to retain the memory previously learned parameters. Equation 1 represents the idea of the memory element in the mathematical form. The data input, xt, is used adjusted with input weight, Wx, combined with the previous memory ht-1 also adjusted with hidden weights Wh to get ht, the updated memory of a perceptron at time step, t after passing all of it through a non-linear function (σ). This memory is then used to calculate the output, yt after multiplying a matrix of output weights Wy, this is represented in the equation 2. The RNN models solved the following mathematical equations:

$$h_t = \sigma\left(W_x x_t + W_h h_{t-1} + b\right) \tag{1}$$

$$y_t = W_y h_t \tag{2}$$

The RNNs introduced the concept of memory in a neural network; however, these models have a tendency of forgetting. To overcome this problem, a new type of RNN models was introduced. These were LSTM (Hochreiter & Schmidhuber, 1997). LSTMs solved the problem of vanishing gradient in RNNs by simply switching the continuous matrix multiplication with addition. This helps in preventing exponential delay in case of backpropagation and thus solving the issue of vanishing gradient descent.

5.2 Related Research

These models were used to identify patterns in Chest X-Ray images and predict pneumonia (Rajpurkar et al., 2017). CNN models were trained to allow pathologists to exclude 65–75% of slides while retaining 100% sensitivity, speeding up the process of cancer detection (Campanella et al., 2019). Deep CNN models were deployed to detect diabetic retinopathy screening from fundus photos (Gulshan et al., 2016). These empirical studies show that CNN-based models achieved expert or near expert level accuracy across radiology and pathology along with ophthalmology (Campanella et al., 2019; Gulshan et al., 2016; Rajpurkar et al., 2017). RNN and RNN based LSTM models surpassed the accuracy levels of classical models in predicting patient readmission and early sepsis detection (Fagerström et al., 2019). RNN models enabled precise predictive modelling from irregular and noisy EHR sequences (Choi et al., 2017a). With the emergence of attention mechanisms, the opacity of RNN in clinical natural language processing started bridging the gap between complex medical terminologies and patient's understanding of it (Alsentzer et al., 2019). Table 2 provides a comparative summary of representative empirical studies from this era.

5.3 Limitations

With the early deep learning models the predictions and outcomes were far better than traditional machine learning models however, they came with their own challenges. Unlike traditional models, where features were engineered, these models extract the features from the data on their own. However, this ability makes them more data hungry, requiring a vast amount of labelled data which is often unavailable. This ability of feature extraction acts like a double-edged sword which can prove very beneficial in predictions, but it also makes the feature selection even more opaque (Hassija et al., 2024). This makes their adoption even less reliable due to a lack of interpretability. There has been significant work done to increase its interpretability by implementing Explainable AI (XAI). Another drawback of these models that came with their better accuracy prediction was an increase in the computational cost. These models now require high performance Graphic Processing Units (GPUs) which are an additional cost over the cost of high-end Central Processing Unit (CPU) machines. The datasets are another limiting factor for these models. If the models are trained on very high-quality images but the real-world images are poorly lit or blurry, the output accuracy of these models dip significantly. These models can be trained on images and texts, but these models can process only one type of input mode for solving the given problem. These models, if trained on text, then cannot take images as the input and vice versa. Their output was also limited to a single type. RNN and LSTM brought the capacity of generating texts,

Table 2: Representation of empirical research in implementing early deep learning models in healthcare

Study	Research Area	Model Used	Dataset Used	Key Result	Performance Metrics
(Rajpurkar et al., 2017)	Pneumonia detection from chest X-rays	Deep CNN (DenseNet-121)	Hospital-scale chest x-ray database containing 112,120 frontal-view X-ray images	Established chest X-ray benchmarks and accelerated adoption of CNNs in radiology research	F1 score of Model is 0.44 higher than the radiologist of 0.39
(Campanella et al., 2019)	Cancer detection in histopathology	Weakly supervised CNN with multiple-instance learning	A dataset of 44,732 whole slide images from 15,187 patients	Demonstrated scalability of deep learning for pathology using slide-level labels	AUC above 0.98
(Gulshan et al., 2016)	Diabetic retinopathy screening	Deep CNN	Retrospective development data set of 128,175 retinal images	Provided the first large-scale evidence that deep learning could achieve screening-level diagnostic performance in ophthalmology	AUC of 0.99, sensitivity 90.3%, specificity 98.1%
(Choi et al., 2017a)	Clinical outcome prediction	Two-level RNN	Electronic health records from Sutter Health	Introduced attention-based interpretability tailored to clinical decision support	AUC of 0.87
(Alsentzer et al., 2019)	Clinical Natural Language Processing (NLP)	BERT	Clinical text from approximately 2 million notes in MIMIC-III v1.4	Established domain-specific pretraining as critical for clinical language understanding	Accuracy of 85.4% on MedNLI
(Fagerström et al., 2019)	Sepsis Prediction	LSTM	MIMIC database	Models detect patients up to 20 hours earlier than Cox proportional hazards model	AUC of 0.83
(Tinauer et al., 2022)	Classification accuracy in Alzheimer's disease	CNN classifiers with relevance-guided heat maps	MRI Scans of multiple patients and controls	Achieves higher classification accuracies than conventional CNNs	Accuracy 86.19%, AUC 0.92, Specificity 92.66%

images or speech, the output of the model can be only one of these. Although these models outperformed the traditional models in prediction, these studies remained retrospective and the performance gains did not translate into improved clinical outcomes. These challenges motivated the industry toward more powerful models that could generalize across tasks.

6. FOUNDATION AND MULTIMODAL MODELS ERA

The publication of 'Attention is all you Need' paper by Google's researchers, introduced the concept of Attention and Transformers to the world (Vaswani et al., 2023). Transformers signaled a dramatic shift in the AI industry's terrain. AI has moved into a new model space known as foundation models with the advent of transformers. Adaptable to a variety of downstream tasks, these models are trained on large amounts of data (Bommasani et al., 2022). The most common use of these models is Large Language Models (LLMs). LLMs are transformers that are language models trained on a vast amount of text, designed for NLP tasks, especially text generation. With the advancements in transformer training, another set of models called Multimodal Models started generalizing across the tasks. These models are trained on multiple modes of input and can also generate output in multiple forms like text, speech, images or videos.

6.1 Models

The models in this era are mainly transformers. These models use the concept of self-attention and unlike RNNs, can remember the context for longer periods. These models enabled self-supervising or weakly supervised pretraining on huge healthcare data. Unlike models from the early deep learning era, these models learn general-purpose representations that can be adopted across domains and multiple modes of data. This eliminates the need for multiple different early deep learning models required for doing the same task.

An attention mechanism known as self-attention, or intra-attention, links various points in a single sequence to calculate a representation of the sequence (Vaswani et al., 2023). The input is made up of a collection of queries grouped together as matrix Q, as well as keys and values grouped together as K and V. Equation 3 illustrates how the output matrix is calculated (Vaswani et al., 2023):

$$\text{Attention(Q,K,V)} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

Here Q stands for Query, K stands for Key and V stands for Value; d_k is the dimensionality of the key and query vectors.

A definitive milestone in the capabilities of these transformer-based models came in the form of AlphaFold2 model (Jumper et al., 2021). This model solved the 50-year-old problem of protein structure prediction by showcasing what these attention architecture-based models can do when trained at scale (Jumper et al., 2021). This brought a huge paradigm shift in AI in healthcare. Now, the AI models are not only good for predicting outcomes based on the feature selection but can also help in new knowledge generation (Jumper et al., 2021; Lin et al., 2023). As we have seen that models from the traditional and early deep learning phase were limited to a unimodal input or output, with the onset of transformers and other better algorithms, this limitation did not affect models from this generation. The models from this generation could be trained on multiple sources like, text, images or speech and are also capable of generating output in multiple modes. These multimodal models train on huge datasets of images and texts, with some separating images and text during training, while others examine them concurrently (Radford et al., 2021). These models when trained on clinical text datasets like MIMIC-IV and X-rays images can be used to explain the medical reports (Moor et al., 2023; Sellergren et al., 2026).

Models in this era further developed the concept of deep learning by proposing new models. These state-of-the-art models used core concepts of computer science and mathematics and elevated to neural network level. Graph Neural Networks (GNNs) used the concept of graphs and used it to create a new deep learning model. This model is much cheaper to train than the transformers and can predict with extremely high accuracy (Jin et al., 2020). These new age deep learning models along with the concept of attention mechanism, are being used at the frontiers of healthcare (Choi et al., 2017b).

This era also saw the rise of concern over data privacy. Due to strict data privacy laws, a dataset crisis started to emerge. For the data hungry models like transformers, this would result in huge loss in accuracy and training. A novel training paradigm called Federated Learning (FL) emerged to mitigate this dataset crisis (Rieke et al., 2020). Models are deployed on local clients to train the models and use their parameters to train a global model at the server. FL makes it possible to obtain insights cooperatively, such as through a consensus model, without transferring patient data outside of the firewalls of the organizations where they are housed. Rather, only model characteristics (such as parameters and gradients) are transferred, and the machine learning process takes place locally at each participating institution (Rieke et al., 2020). This helps in protecting the patients' sensitive data while training the models.

6.2 Related Research

Models in this era worked in almost all sections of the healthcare industry. One of the major breakthroughs that happened is in the field of drug discovery and protein structure (Jin et al., 2020; Jumper et al., 2021; Lin et al., 2023). This fast forwarded the development of drugs. Another major research happened in the field of training transformers on the medical terminology and building Large Language Models (LLMs) that can help bridge the gap between patients and medical terminologies (Moor et al., 2023; Yang et al., 2022). This helped the medical practitioners in making patients understand the complexities and symptoms of various conditions.

Researchers gave transformers medical vision along with their ability to generate clinical text from unstructured EHR along with the medical reports and images (Moor et al., 2023).

The GNNs used the established medical hierarchies to predict very highly accurate outcomes even from the limited datasets (Choi et al., 2017b). Application of GNNs also improved the drug discovery and protein structure analysis which has accelerated the development of new therapeutics (Jin et al., 2020). Table 3 summarizes representative foundation and multimodal AI studies and their contributions to healthcare and biomedical research.

6.3 Limitations

These models ushered in a new age of innovative methods and techniques to solve some age-old issues and problems in healthcare. However, these also bring a set of issues of their own. These parameter-heavy transformers still have the reliability problem because of hallucinations (Reddy et al., 2024). These can create spurious correlation and generate data that does not at all exist in the real world (Reddy et al., 2024). Even though these models can articulate highly medical advice but because of lack of any standard evaluation mechanism, these are always considered with caution. The ever-increasing complexity of these models also made them very opaque which again raises their adoption concern (Tonekaboni et al., 2019).

Table 3: Representative empirical studies from the foundation and multimodal AI era.

Study	Research Area	Model Used	Dataset Used	Key Result	Performance Metrics
(Jumper et al., 2021)	Drug Discovery and Protein Structure Analysis	AlphaFold2 (attention-based foundation model)	Protein Data Bank	Near-experimental accuracy in protein structure prediction, which marked the transition from task-specific deep learning to biomedical foundation models	Very high median accuracy of 0.96 Å
(Lin et al., 2023)	Protein function and structure modeling	Protein Language Model Transformer	UniRef protein		
sequence database	Enabled reusable protein representations for downstream biomedical applications	TM-score of 0.83, measures accuracy of projection in comparison with actual structure			
(Jin et al., 2020)	Drug discovery and molecule generation	Graph based molecular foundation model	Large molecular graph datasets	Generative design of novel drug-like molecules	Reconstruction accuracy of 79.9%
(Yang et al., 2022)	Clinical NLP and decision support	Large clinical language model	Wikipedia, PubMed publications, and clinical notes from the University of Florida (UF) Health.	Improved performance across multiple clinical NLP tasks	Clinical Extraction F1 Score - 0.90 Medical relation extraction F1 Score - 0.96 MedNLI Accuracy - 0.90
(Moor et al., 2023)	Medical Vision and Clinical text	Vision–language transformer OpenFlamingo-9B	Interleaved		
image-text dataset	Can perform multimodal				
in-context learning specialized for the medical domain	Clinical evaluation, rated by three medical doctors on a scale of 0 to 10, 5.61				

Contd....Representative empirical studies from the foundation and multimodal AI era.

Study	Research Area	Model Used	Dataset Used	Key Result	Performance Metrics
(Horikawa, 2025)	Decoding brain activity to uncover mental content	Language Model pretrained for masked language modelling	fMRI volumes measured during viewing and recalling each video	Implementation of Large Language Models to decode the brain activity	Video identification accuracy of 50%
(Choi et al., 2017b)	Biomedicine	Graph Attention Model	Electronic health records (EHR)	Brings the concept of Attention and GNN and uses the available data for prediction.	AUC of HF onset prediction of 0.84
(Li et al., 2019)	Privacy preserving training model	Federated Deep Learning Network	Brain tumour segmentation on the BraTS dataset	Viability of using differential privacy approaches in a federated learning environment to safeguard patient data	66.3% accuracy of FedAvg Algorithm

Transformers need very high-performance GPUs for training and for processing (Wang et al., 2022), however this has led to various environment impacting issues. The computational cost of transformers is directly related to the power consumed by them, which raises concerns about sustainability.

7. EMERGING TRENDS AND TECHNOLOGY

As the hardware capabilities have caught up with the existing and new research in AI, its landscape is changing rapidly. With new and more powerful GPUs, the foundation models are now giving way to multimodal large language models (MLLMs) that are very good at reasoning, generation, predicting patterns and performing several tasks (Karunanayake, 2025). These new models, unlike traditional AI models, can perform some predefined tasks and operations without any manual intervention. These models are called Agentic AI.

To increase the usability of AI applications, these applications should be more interpretable. Recent developments have been focused on creating an XAI. The XAI idea aims to explain individual choices, facilitate comprehension of overall strengths and shortcomings, and provide a knowledge of future system behaviour (Gunning & Aha, 2019).

7.1 Models

In dynamic situations, the agentic AI models exhibit flexibility, sophisticated decision-making skills, and self-sufficiency. These models employ various reasoning techniques like Chain of Thought (CoT) (Wei et al., 2023), Reason and Act (ReAct) (Yao et al., 2022) or Tree of Thought (ToT) (Yao et al., 2023) to achieve these tasks. With MLLMs at the heart of their reasoning, these agentic AI models can operate independently and alter clinical procedures like clinical summarization and medical report production with little assistance from humans (Acharya et al., 2025).

Gunning & Aha (2019) set the idea of XAI and it was followed with a lot of research. The XAI models proved that one doesn't need to understand the entire complex neural network of a model to trust it. Instead, by probing the model with slight changes in the input and verifying how the output prediction changes, an explanation can be reached upon for a particular dataset (Ribeiro et al., 2016). The use of mathematical standards to explain the inner workings of a neural network further increased the reliability and interpretability of deep learning networks in the healthcare industry (Lundberg et al., 2017; Ribeiro et al., 2016). Local interpretable model agnostic explanation (LIME) can be used to highlight the most influential factors in treatment recommendation or individual diagnoses. Shapley additive explanations (SHAP) can provide insights into the contribution of different features to a diagnosis or a treatment plan.

7.2 Related Research

A recent study has shown that agentic AI can be implemented for organ health monitoring (Suura, 2025) that can help in early detection of organ rejects. This shows that agentic AI can be implemented to create an AI doctor that can help the patients answer not only their medical queries but also explain the processes of medical formalities. The ongoing research shows that medical LLMs have become more reliable by using search engine tools to retrieve medical knowledge.

Much research has been done to understand the working of machine learning models but only after (Lundberg et al., 2017; Ribeiro et al., 2016) research work, helpful and workable insights were drawn into the black box nature of the machine learning model. These papers brought in a new trend of machine learning techniques that focused on interpretability and working

Table 4: Representative empirical studies in emerging AI technology.

Study	Research Area	Model Used	Dataset Used	Key Result	Performance Metrics
(Suura, 2025)	Organ health monitoring	Time-series model	Clinical and para-clinical data from local research hospital CRO of Cagliari and with the hospital of Pavia	80% alert happened in due time	27 agents triggered alerts in due time out of 35 agents deployed.
(Muneer et al., 2025)	CVD prediction	SHAP and LIME XAI along with XGBoost model	Clinical and lifestyle attributes	The method outperforms previous strategy by attaining the highest precision	The SHAP model showed that 'systolic blood pressure' is the feature of high importance instead of previously thought 'age'
(Vani et al., 2025)	Personalized health monitoring	CNN with attention mechanism and SHAP interpretability	EHR	High performing predictive model and explainability at both population and individual patient level	The accuracy of the model dropped from 97.86% to 90.30 after dropping all the SHAP enhancements
(Yesmin et al., 2026)	Maternal health assessment	XGBoost model with XAI	UCI Maternal Health Risk Dataset	Persistent tendency among clinicians for mixed explanations as opposed to single-method approaches	71.4% of healthcare professionals accepted the hybrid explanation while 54.8% expressed trust for clinical use
(Alshammri et al., 2026)	Detection of bladder cancer	Hierarchical Representation Learning Model SHAP	Bladder cancer classification dataset	Model effectively mitigates image noise and improves overall visual quality and XAI method to interpret the model's decision-making process	The accuracy of the model with feature engineering using SHAP is 98.75%

of complex models. Due to their white box nature, clinics and medical experts have started using XAI for diagnosing and creating treatment plans for cardiovascular disease (CVD) (Muneer et al., 2025), personalized health monitoring (Vani et al., 2025), maternal health assessment (Yesmin et al., 2026), and enhanced bladder cancer detection models (Alshammri et al., 2026). These studies have shown that with better visibility and transparency AI technology can be integrated with confidence into the healthcare industry. Table 4 summarizes all the above representative research studies in agentic AI and XAI along with their contribution to healthcare.

7.3 Limitations

The availability of varied, high-quality data to train AI models is the main obstacle to healthcare AI. Seamlessly incorporating agentic AI into existing clinical workflows is a formidable challenge. Many healthcare organizations still use operational systems and outdated EHRs that aren't built to handle new AI tools. One of the main obstacles to the adoption of AI in healthcare is still the lack of acceptance and trust between patients and clinicians. To solve this, healthcare workers should get organized training on AI-assisted decision-making, and AI agentic models should be rigorously validated in the real world. Furthermore, the system is opaque in its operations due to the sensitive data and unstructured text, which are frequently governed by privacy regulations. This necessitates using both these agentic AIs and XAI (Karunanayake, 2025). Developing strategies to explain these ”black box” models in high stakes decisions, but attempting to do so instead of developing interpretable models in the first place is likely to encourage unethical behavior and could have a significant negative impact on society (Rudin, 2019). Unlike the explainable AI models that are available today, modern AI models and data are too complex with high dimensionality, that they can't be simply explained based on the inputs and outputs (Ghassemi et al., 2021).

8. CHALLENGES IN ADOPTION OF AI IN HEALTHCARE

As we have seen that as AI models advanced, their application in the healthcare industry also increased. Empirical studies have been on a continuous rise, either through the implementation of new models on existing datasets or by applying traditional models on the new datasets. Even though information technology infrastructure has been well integrated in the healthcare industry, the availability of quality data remains one major hurdle in training of AI models. However, despite this rapid growth, there are still some challenges that AI models need to solve to be adopted by medical practitioners and healthcare experts. Another major hurdle in the adoption of AI models by the healthcare industry is its opaque nature (Tonekaboni et al., 2019). The opaqueness of these AI systems is basically the machine's inability to explain its decisions and actions to users. As the AI models are getting increasingly complex, their opacity has also increased (Hassija et al., 2024). For someone not from the background of computer science, understanding the outcomes of these models seems very less reliable. Also, with the new age AI models, the problem of hallucinations is also a roadblock in their adoption (Reddy et al., 2024). The second major challenge in adoption of AI models is their deployment. Even though now there are various cloud platforms that can host AI models, there is a lack of data pipelining which makes training these models harder and their use even harder. Another major issue in their adoption comes from the concern of power consumption. As these models become more computation heavy, they need a lot of power and other resources to keep them running.

AI models are now also facing issues related to data privacy and data governance (Rieke et al., 2020). With stricter data privacy laws, it is becoming harder for researchers to get new datasets. This is also a concern and a challenge in adoption of AI in healthcare.

9. CONCLUSION

The advent of AI has brought about a transformative change in the healthcare domain. While traditional machine learning models primarily focused on predictive analysis of diseases, they were not adept in handling real-time unstructured data. This limitation was overcome using deep learning models, which are capable of efficient feature learning from raw data and images. However, these models are data demanding and require extensive labelled datasets to provide accurate results. The advent of multimodal models and transformers has been a breakthrough in the healthcare domain as predictive analytics would now not just be dependent on just one mode i.e. textual data or image data but would integrate text, images or may even biological data concurrently for efficient prediction. Transformers use their self-attention mechanism to process medical records to summarize information, analyze bio signals to flag anomalies, examine genetic data for discovering new drugs and enhance predictive disease analysis. Furthermore, the current trends of explainable models in the healthcare domain make the results more trustworthy and actionable. Predictive analysis can be validated using XAI by flagging key contributors to the patient's health state. In future, strong and effective healthcare ecosystems that integrate AI into disease diagnostics, personalized treatments using genetic information, remote diagnostics and complex robotic surgeries would ensure proactive care and precise treatments.

Conflict of Interests

The author has no conflict of interest to declare.

Funding

Not applicable.

REFERENCES

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals - A comprehensive survey. *IEEE Access*, *13*, 18912–18936. https://doi.org/10.1109/ACCESS.2025.3532853

Al Kuwaiti, A., Nazer, K., Al-reedy, A., Al-shehri, S., Al-muhanna, A., Subbarayalu, A. V., Al Muhanna, D., & Al-muhanna, F. A. (2023). A review of the role of artificial intelligence in healthcare. *Journal of Personalized Medicine*, *13*(6), Article 951. https://doi.org/10.3390/jpm13060951

Alsentzer, E., Murphy, J., Boag, W., Weng, W.-h., Jindi, D., Naumann, T., & Mcdermott, M. (2019). Publicly available clinical BERT embeddings. In A. Rumshisky, K. Roberts, S. Bethard, & T. Naumann (Eds.), *Proceedings of the 2nd Clinical Natural Language Processing Workshop* (pp. 72–78). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-1909

Alshammri, G. H., Obayya, M., Negm, N., Alaqil, M. A., Alshahrani, M. M., Alsini, R., Alsuhaibani, R., & Alhashmi, A. A. (2026). Towards enhanced bladder cancer detection using explainable artificial intelligence with a hybrid feature engineering framework on biomedical images. *Engineering Applications of Artificial Intelligence*, *167*, Article 113729. https://doi.org/10.1016/j.engappai.2026.113729

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Von Arx, S., ... & Liang, P. (2021). *On the opportunities and risks of foundation models.* arXiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Bron, E. E., Smits, M., Niessen, W. J., & Klein, S. (2015). Feature selection based on the svm weight vector for classification of dementia. *IEEE Journal of Biomedical and Health Informatics*, *19*(5), 1617–1626. https://doi.org/10.1109/JBHI.2015.2432832

Cameron, D., Bhagwan, V., & Sheth, A. P. (2012). Towards comprehensive longitudinal healthcare data capture. *2012 IEEE International Conference on Bioinformatics and Biomedicine Workshops*, 240–247. https://doi.org/10.1109/BIBMW.2012.6470310

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., Brogi, E., Reuter, V. E., Klimstra, D. S., & Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine*, *25*(8), 1301–1309. https://doi.org/10.1038/s41591-019-0508-1

Choi, E., Bahadori, M. T., Kulas, J. A., Schuetz, A., Stewart, W. F., & Sun, J. (2017a). RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in Neural Information Processing Systems*, *29,* 3512–3520. https://doi.org/10.48550/arXiv.1608.05745

Choi, E., Bahadori, M. T., Song, L., Stewart, W. F., & Sun, J. (2017b). GRAM: Graph-based attention model for healthcare representation learning. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F129685*, 787–795. https://doi.org/10.1145/3097983.3098126

Chou, R. H., Hsu, B. W. Y., Yu, C. L., Chen, T. Y., Ou, S. M., Lee, K. H., Tseng, V. S., Huang, P. H., & Tarng, D. C. (2024). Machine-learning models are superior to severity scoring systems for the prediction of the mortality of critically ill patients in a tertiary medical center. *Journal of the Chinese Medical Association: JCMA*, *87*(4), Article 369. https://doi.org/10.1097/JCMA.0000000000001066

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, *14*(2), 179–211. https://doi.org/10.1207/s15516709cog1402_1

Fagerström, J., Bång, M., Wilhelms, D., & Chew, M. S. (2019). LiSep LSTM: A machine learning algorithm for early detection of septic shock. *Scientific Reports*, *9*(1), Article 15132. https://doi.org/10.1038/s41598-019-51219-4

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People-an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

Futoma, J., Morris, J., & Lucas, J. (2015). A comparison of models for predicting early hospital readmissions. *Journal of Biomedical Informatics*, *56*(9), 229–238. https://doi.org/10.1016/j.jbi.2015.05.016

Ghassemi, M., Oakden-rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, *3*(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

Goldberg, S. I., Shubina, M., Niemierko, A., & Turchin, A. (2010). A weighty problem: Identification, Characteristics and risk factors for errors in EMR data. *AMIA Annual Symposium Proceedings*, *2010*, 251–5. https://pubmed.ncbi.nlm.nih.gov/21346979

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, *316*(22), 2402–2410. https://doi.org/10.1001/jama.2016.17216

Gunning, D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence program. *AI Magazine*, *40*(2), 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Hasan, M., Fukuda, A., Maruf, R. I., Yokota, F., & Ahmed, A. (2017). Errors in remote healthcare system: Where, how and by whom? *TENCON 2017 - 2017 IEEE Region 10 Conference*, 170–175. https://doi.org/10.1109/TENCON.2017.8227856

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting black-box models: A review on explainable artificial intelligence. *Cognitive Computation*, *16*(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horikawa, T. (2025). Mind captioning: Evolving descriptive text of mental content from human brain activity. *Science Advances*, *11*(45), eadw1464. https://doi.org/10.1126/sciadv.adw1464

Jin, W., Barzilay, R., & Jaakkola, T. (2020). Hierarchical generation of molecular graphs using structural motifs. *Proceedings of the 37th International Conference on Machine Learning, Proceedings of Machine Learning Research*, *119*, 4839–4848. https://doi.org/10.48550/arXiv.2002.03230

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kalra, N., Verma, P., & Verma, S. (2024). Advancements in AI based healthcare techniques with FOCUS on diagnostic techniques. *Computers in Biology and Medicine*, *179*, Article 108917. https://doi.org/10.1016/j.compbiomed.2024.108917

Karunanayake, N. (2025). Next-generation agentic AI for transforming healthcare. *Informatics and Health*, *2*(2), 73–83. https://doi.org/10.1016/j.infoh.2025.03.001

Kitsios, F., Kamariotou, M., Syngelakis, A. I., & Talias, M. A. (2023). Recent advances of artificial intelligence in healthcare: A systematic literature review. *Applied Sciences, 13*(13), Article 7479. https://doi.org/10.3390/app13137479

Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, *1*(4), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M. J., & Feng, A. (2019). Privacy-Preserving federated brain tumour segmentation. *Lecture Notes in Computer Science*, *11861*, 133–141. https://doi.org/10.1007/978-3-030-32692-0_16

Liang, S., Singh, M., Dharmaraj, S., & Gam, L. H. (2010). The PCA and LDA analysis on the differential expression of proteins in breast cancer. *Disease Markers*, *29*(5), 231–242. https://doi.org/10.3233/DMA-2010-0753

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., ... & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, *379*(6637), 1123-1130. https://doi.org/10.1126/science.ade2574

Ling, Y., An, Y., Liu, M., & Hu, X. (2013). An error detecting and tagging framework for reducing data entry errors in electronic medical records (EMR) system. *2013 IEEE International Conference on Bioinformatics and Biomedicine*, 249–254. https://doi.org/10.1109/BIBM.2013.6732498

Liu, S.-h., Cheng, D.-c., & Lin, C.-m. (2013). Arrhythmia identification with two-lead electrocardiograms using artificial neural networks and support vector machines for a portable ECG monitor system. *Sensors*, *13*(1), 813–828. https://doi.org/10.3390/s130100813

Lundberg, S. M., Allen, P. G., & Lee, S.-i. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

Moor, M., Huang, Q., Wu, S., Yasunaga, M., Zakka, C., Dalmia, Y., Reis, E. P., Rajpurkar, P., & Leskovec, J. (2023). Med-Flamingo: A multimodal medical few-shot learner. *Proceedings of Machine Learning Research*, *225*, 353–367. https://doi.org/10.48550/arXiv.2307.15189

Muneer, S., Abbas, S., Shah, A. A., Alharbi, M., Aldossary, H., Fatima, A., Ghazal, T. M., & Adnan, K. M. (2025). Responsible CVD screening with a blockchain assisted chatbot powered by explainable AI. *Scientific Reports*, *15*(1), Article 11558. https://doi.org/10.1038/s41598-025-96715-y

Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. *Critical Care Medicine*, *46*(4), 547–553. https://doi.org/10.1097/CCM.0000000000002936

Neupane, S., Mittal, S., & Rahimi, S. (2025). *Towards a HIPAA compliant agentic ai system in healthcare.* arXiv:2504.17669. https://doi.org/10.48550/arXiv.2504.17669

Polat, K., Şahan, S., & Güneş, S. (2007). Automatic detection of heart disease using an artificial immune recognition system (AIRS) with fuzzy resource allocation mechanism and k-nn (nearest neighbour) based weighting preprocessing. *Expert Systems with Applications*, *32*(2), 625–631. https://doi.org/10.1016/j.eswa.2006.01.027

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *Proceedings of Machine Learning Research*, *139*, 8748–8763. https://doi.org/10.48550/arXiv.2103.00020

Rajpurkar, P., Irvin, J.A., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D.Y., Bagul, A., Langlotz, C., Shpanskaya, K.S., Lungren, M.P., & Ng, A. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *ArXiv, abs/1711.05225*. https://doi.org/10.48550/arXiv.1711.05225

Reddy, G. P., Kumar, Y. P., & Prakash, K. P. (2024, April). Hallucinations in large language models (LLMs). In *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)* (pp. 1-6). IEEE. https://doi.org/10.1109/eStream61684.2024.10542617

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why Should I. Trust You?" Explaining The Predictions Of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *Npj Digital Medicine*, *3*(1), Article 119. https://doi.org/10.1038/s41746-020-00323-1

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and*

Development, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Sellergren, A., Gao, C., Mahvar, F., Kohlberger, T., Jamil, F., Traverse, M., … & Golden, D. (2026). *Medgemma 1.5 technical report.* arXiv:2604.05081. https://doi.org/10.48550/arXiv.2604.05081

Shen, Y., Yu, J., Zhou, J., & Hu, G. (2025). Twenty-five years of evolution and hurdles in electronic health records and interoperability in medical research: comprehensive review. *Journal of Medical Internet Research*, *27*, Article e59024. https://doi.org/10.2196/59024

Suura, S. R. (2025). Agentic AI systems in organ health management: early detection of rejection in transplant patients. *Journal of Neonatal Surgery*, *14*(4s), 490–500. https://doi.org/10.52783/jns.v14.1823

Stanford Hai. (2025). *The 2025 AI index report.* Stanford University Human-Centered Artificial Intelligence. Retrieved February 14, 2026, from https://hai.stanford.edu/ai-index/2025-ai-index-report

Tinauer, C., Heber, S., Pirpamer, L., Damulina, A., Schmidt, R., Stollberger, R., Ropele, S., & Langkammer, C. (2022). Interpretable brain disease classification and relevance-guided deep learning. *Scientific Reports*, *12*(1), Article 20254. https://doi.org/10.1038/s41598-022-24541-7

Tonekaboni, S., Joshi, S., Mccradden, M. D., & Goldenberg, A. (2019, October). What clinicians want: contextualizing explainable machine learning for clinical end use. In *Machine Learning for Healthcare Conference* (pp. 359-380). PMLR. https://doi.org/10.48550/arXiv.1905.05134

Vani, M. S., Sudhakar, R. V., Mahendar, A., Ledalla, S., Radha, M., & Sunitha, M. (2025). Personalized health monitoring using explainable AI: bridging trust in predictive healthcare. *Scientific Reports*, *15*(1), Article 31892. https://doi.org/10.1038/s41598-025-15867-z

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1706.03762

Vatsal, S., Dubey, H., & Singh, A. (2026). Agentic AI in healthcare and medicine: A seven-dimensional taxonomy for empirical evaluation of LLM-based agents. *IEEE Access*, *14*, 4840-4863. https://doi.org/10.1109/ACCESS.2026.3651218

Wang, X., Wei, Y., Xiong, Y., Huang, G., Qian, X., Ding, Y., Wang, M., & Li, L. (2022). LightSeq2: Accelerated training for transformer-based models on GPUs. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC, 2022-November.* https://doi.org/10.1109/SC41404.2022.00043

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). Chain-of-Thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837. https://doi.org/10.48550/arXiv.2201.11903

Wolberg, E. J. (2010). The method of least squares. In *Designing quantitative experiments: Prediction analysis* (pp. 47–89). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11589-9_3

Yang, X., Chen, A., Pournejatian, N., Shin, H. C., Smith, K. E., Parisien, C., … & Wu, Y. (2022). *Gatortron: A large clinical language model to unlock patient information from unstructured electronic health records.* arXiv:2203.03540. https://doi.org/10.48550/arXiv.2203.03540

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances In Neural Information Processing Systems*, *36*, 11809–11822. https://doi.org/10.48550/arXiv.2305.10601

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). *ReAct: Synergizing reasoning and acting in language models.* arXiv:2210.03629. https://doi.org/10.48550/arXiv.2210.03629

Yesmin, F., Shirmin, N., & Bristy, S. S. (2026). *Bridging the trust gap: clinician-validated hybrid explainable ai for maternal health risk assessment in Bangladesh.* arXiv:2601.07866. https://doi.org/10.48550/arXiv.2601.07866

Zou, J., & Topol, E. J. (2025). The rise of agentic AI teammates in medicine. *The Lancet*, *405*(10477), Article 457. https://doi.org/10.1016/S0140-6736(25)00202-8

Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in Genetics*, *9*, Article 416440. https://doi.org/10.3389/fgene.2018.00515

COPYRIGHT AND LICENSE

HOW TO CITE?

S. Yadav & M. Khanna. (2026). Understanding the Advancement of Artificial Intelligence in Healthcare. *Vantage: Journal of Thematic Analysis*, 07(01), 60-80. https://vantagejournal.com/