2026 Volume: 02 Issue: 11 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-13

AI-BASED MULTIMODAL CLINICAL DECISION SUPPORT SYSTEM

***Nikitha S., Chiranjeet S., Nischith B. M., Prajwal Manjunath Pattar**

Department of Computer Science Engineering Jyothy Institute of Technology Bengaluru, India.

Article Received: 24 September 2026
Article Revised: 14 October 2026
Published on: 03 November 2026

***Corresponding Author: Nikitha S.**
Department of Computer Science Engineering Jyothy Institute of Technology Bengaluru, India.

DOI: https://doi-doi.org/101555/ijrpa.1189

ABSTRACT

Artificial Intelligence (AI) has significantly improved Clinical Decision Support Systems (CDSS) by enabling intelligent analysis of healthcare data for disease diagnosis and clinical decision-making. Earlier CDSS mainly relied on rule-based systems and structured healthcare records, which limited their adaptability and diagnostic capability. Recent advancements in deep learning, Natural Language Processing (NLP), and multimodal learning have enabled the integration of medical images, patient symptoms, clinical text, and structured healthcare data for more accurate and context-aware diagnosis. Multimodal AI systems combine complementary information from different healthcare modalities to improve prediction accuracy and clinical reasoning. This survey reviews recent developments in multimodal clinical diagnosis systems, including structured healthcare data systems, CNN-based medical imaging models, NLP-based symptom analysis systems, multimodal fusion approaches, foundation models, and generative AI techniques in healthcare.

KEYWORDS: Clinical Decision Support System (CDSS), Multimodal AI, Medical Imaging, Deep Learning, Natural Language Processing (NLP), CNN, MultiCaRe Dataset, Explainable AI, Vision-Language Models, Healthcare AI.

I. INTRODUCTION

Clinical Decision Support Systems (CDSS) have evolved from traditional rule-based systems into intelligent AI-driven healthcare frameworks capable of supporting disease diagnosis and clinical decision-making [1], [2]. Early CDSS mainly depended on predefined medical rules and expert knowledge for generating clinical recommendations [1], [2]. Although these systems provided interpretability and transparency, they lacked

scalability and adaptability when handling continuously growing healthcare data and complex medical conditions [2], [3]. Several studies further emphasized that conventional rule-based systems were insufficient for handling heterogeneous healthcare information and dynamic clinical environments [3], [4].

The digitization of healthcare systems through Electronic Health Records (EHRs), laboratory reports, physiological measurements, and demographic records generated large volumes of structured healthcare data that could be used for predictive clinical analysis [3], [4]. Several studies demonstrated that machine learning algorithms could analyze structured clinical data for disease prediction, patient monitoring, and treatment recommendation [4], [5]. These systems improved the efficiency of clinical analysis compared to traditional rule-based approaches but were limited because they relied mainly on numerical and structured patient information [3], [4], [5]. Research in healthcare analytics also highlighted the importance of integrating patient history, laboratory findings, and demographic information for improving predictive clinical performance [4], [5], [6].

Along with structured healthcare systems, deep learning techniques significantly improved medical image analysis using imaging modalities such as MRI, CT scan, X-ray, ultrasound, and retinal imaging [6], [7]. CNN-based systems were widely applied for tumor detection, cardiovascular disease analysis, neurological disorder diagnosis, and ophthalmological screening [6], [7]. These systems achieved high accuracy in image classification and abnormality detection by automatically learning spatial and hierarchical image features [7], [8]. Several studies additionally explored multimodal image fusion and advanced feature extraction methods for improving medical image interpretation and disease localization [7], [8], [20].

However, several studies highlighted that image-based systems alone were insufficient for real-world clinical diagnosis because they lacked contextual clinical understanding such as patient symptoms, disease history, laboratory findings, and physician observations [8], [9]. To address this limitation, NLP-based clinical systems were developed to process unstructured healthcare text including discharge summaries, radiology reports, physician notes, and symptom descriptions [9], [10]. NLP and transformer-based architectures enabled semantic understanding of clinical narratives and improved healthcare text analysis [13], [15]. Recent studies further demonstrated that NLP systems could support

automated symptom understanding, report summarization, conversational healthcare assistance, and clinical documentation analysis [9], [13], [15].

As clinical diagnosis generally depends on multiple sources of information, recent studies increasingly focused on multimodal AI systems capable of integrating medical images, structured healthcare data, and clinical text into unified diagnostic frameworks [10], [11]. These systems aimed to mimic clinical reasoning by combining complementary information from different modalities for improved diagnostic performance. The reviewed studies showed that multimodal systems achieved better prediction accuracy and robustness compared to unimodal approaches in applications such as cancer diagnosis, cardiovascular analysis, and neurological disease detection [10], [12].

Recent advancements in foundation models, vision-language systems, and multimodal Large Language Models (LLMs) further accelerated the development of intelligent healthcare systems [13], [14]. These models enabled cross-modal reasoning, automated report generation, image captioning, and conversational clinical assistance by jointly processing image and textual information. Generative AI and Retrieval-Augmented Generation (RAG)-based healthcare systems also emerged as promising approaches for intelligent patient interaction and clinical support [15], [16].

Jyothy Institute of Technology

Despite significant progress, several challenges remain in the deployment of multimodal clinical diagnosis systems. Major issues identified in the literature include healthcare data heterogeneity, lack of standardized multimodal datasets, annotation inconsistency, computational complexity, explainability limitations, and privacy concerns [17], [18]. In addition, reliable explainability mechanisms are necessary to improve physician trust and ensure safe clinical deployment of AI systems [19].

This survey examines recent developments in multimodal clinical diagnosis systems that integrate structured healthcare data, medical imaging, and clinical text for intelligent disease prediction and clinical decision-making [10], [11]. The survey reviews advancements in deep learning-based medical imaging, NLP-driven healthcare text analysis, multimodal fusion strategies, foundation models, and generative AI techniques used in modern healthcare systems [12], [13], [15]. It also analyzes the major challenges associated with multimodal healthcare AI, including data heterogeneity, explainability limitations, computational complexity, and clinical deployment issues [17], [18], [19].

II. LITERATURE SURVEY

Early research in Clinical Decision Support Systems mainly focused on rule-based architectures and structured healthcare data obtained from Electronic Health Records (EHRs), laboratory reports, physiological measurements, and demographic records [1], [3]. These systems used machine learning algorithms to identify disease patterns and support clinical prediction tasks. Studies showed that structured healthcare data significantly improved predictive analytics and patient monitoring compared to traditional rule-based systems [4], [5]. Structured healthcare records provided important clinical information such as disease history, laboratory measurements, medication details, and patient demographics, which supported disease risk prediction and treatment planning. However, these systems could not process visual diagnostic information present in medical imaging modalities.

The development of deep learning significantly improved automated medical image analysis using CNN-based architectures [6]. Several studies applied CNN models to imaging modalities such as MRI, CT scan, X-ray, retinal imaging, and histopathology images for disease classification and abnormality detection [7]. CNN-based systems achieved strong performance in applications such as tumor detection, cardiovascular imaging, neurological analysis, and ophthalmological diagnosis. Research papers also introduced multimodal image fusion techniques combining MRI and CT images for improved visualization and structural analysis [20]. Although CNN models improved imaging diagnosis, the surveyed studies highlighted that these systems remained limited because they processed only image data without incorporating patient symptoms or clinical context [8].

To overcome the limitations of image-only systems, several studies focused on applying NLP techniques for processing unstructured healthcare text including physician notes, discharge summaries, pathology reports, and symptom descriptions [8], [9]. NLP systems transformed clinical narratives into semantic representations that could support disease prediction and clinical analysis. Transformer-based architectures and LLMs further improved contextual understanding in healthcare NLP applications [15]. The reviewed studies demonstrated that NLP systems could support symptom understanding, report summarization, patient monitoring, and clinical documentation. However, text-based systems alone could not interpret visual abnormalities present in

medical imaging data.

As clinical diagnosis generally depends on multiple sources of information, recent studies increasingly focused on multimodal AI systems integrating medical images, structured healthcare data, and clinical text into unified diagnostic frameworks [10], [11]. These systems attempted to replicate clinical reasoning by combining complementary information from multiple healthcare modalities. Several papers explored multimodal fusion strategies such as early fusion, feature-level fusion, and decision-level fusion [12]. Among these methods, feature-level fusion was commonly preferred because it enabled direct interaction between image and textual features before prediction generation [10]. The surveyed literature consistently showed that multimodal systems achieved better diagnostic performance compared to unimodal systems in cancer diagnosis, cardiovascular disease analysis, neurological disorder prediction, and ophthalmological applications [10], [11].

Recent research also introduced foundation models and vision-language architectures capable of jointly processing medical images and clinical text [13], [14]. These systems supported advanced healthcare tasks such as automated report generation, image captioning, visual question answering, and multimodal reasoning. Several studies discussed the growing role of generative AI and multimodal LLMs in healthcare applications [15]. These systems enabled conversational diagnosis support, patient interaction, and AI-assisted healthcare guidance. Retrieval-Augmented Generation (RAG)-based healthcare architectures further improved clinical assistance by integrating external medical knowledge retrieval with multimodal reasoning [16]. Although these systems demonstrated promising results, the literature highlighted challenges related to computational requirements, explainability, reliability, and regulatory validation.

Based on the reviewed literature, several multimodal approaches were widely applied across organ-specific clinical diagnosis applications.

A. *Heart (Cardiovascular Disease Analysis)*

Several studies applied multimodal AI systems for cardiovascular disease diagnosis using ECG signals, cardiac imaging, and structured healthcare records [4], [10]. CNN-based approaches improved automated analysis of echocardiograms, CT scans, and cardiac MRI images for identifying structural abnormalities and arterial blockages.

Research papers also highlighted the importance of integrating patient history, symptoms, and laboratory findings with imaging information for improving cardiovascular risk prediction [3], [5]. Multimodal approaches demonstrated improved diagnostic accuracy and reduced prediction uncertainty compared to single-modality systems.

B. *Brain (Neurological Disease Analysis)*

Neurological disease diagnosis represented another important application area for multimodal AI systems. Several studies combined MRI and CT imaging with clinical information for brain tumor detection and neurological disorder analysis [7], [20].

Advanced multimodal image fusion techniques improved visualization of brain abnormalities by integrating complementary imaging modalities. The reviewed studies showed that multimodal systems improved tumor localization, classification accuracy, and structural analysis compared to unimodal imaging systems [10], [11].

C. *Eye (Ophthalmological Disease Analysis)*

AI-based ophthalmological systems achieved significant success in retinal disease diagnosis using fundus images and retinal scans [6], [7]. CNN-based systems were applied for diabetic retinopathy detection, glaucoma identification, and retinal abnormality analysis.

The reviewed studies further improved ophthalmological diagnosis by integrating demographic information, patient symptoms, and clinical records with retinal imaging [5], [10].

These multimodal systems achieved higher diagnostic performance and supported early disease detection.

III. COMPARATIVE ANALYSIS OF REVIEWED PAPERS

Table 1 presents a comparative summary of the reviewed studies based on methodology, contribution, and limitations.

Ref	Methodology	Contribution	Limitation
[1]	Rule-based CDSS	Clinical decision automation	Limited scalability
[3]	Structured healthcare analytics	EHR-based disease prediction	No imaging support
[6]	CNN-based imaging	Automated disease classification	Unimodal limitation
[9]	NLP clinical systems	Symptom understanding	No image analysis
[10]	Multimodal AI	Multimodal disease diagnosis	Data heterogeneity
[12]	Feature-level fusion	Improved prediction accuracy	Computational complexity
[13]	Foundation models	Medical image-text understanding	High resource requirements
[15]	Generative AI and LLMs	Conversational healthcare support	Reliability concerns
[16]	RAG healthcare systems	Retrieval-assisted clinical support	Deployment complexity
[21]	MultiCaRe dataset	Large-scale multimodal dataset	Annotation inconsistency

IV. DATASET ANALYSIS: MULTICARE

The MultiCaRe dataset represented one of the most important multimodal healthcare datasets discussed in the literature [21]. The dataset contains approximately 93,000 clinical cases and more than 130,000 medical images collected from PubMed Central case reports. It was specifically designed to support multimodal healthcare AI applications by integrating medical imaging data with corresponding clinical text and metadata.

Several studies highlighted that MultiCaRe provides paired medical images, clinical narratives, image captions, disease labels, and demographic metadata across multiple imaging modalities such as MRI, CT scan, X-ray, ultrasound, retinal imaging, and pathology imaging [21]. Unlike traditional unimodal medical datasets, MultiCaRe enables simultaneous analysis of visual and textual healthcare information, making it highly suitable for multimodal disease diagnosis systems.

The dataset includes a wide variety of disease categories and clinical conditions collected from real-world medical case reports. Its hierarchical disease taxonomy supports both coarse-grained and fine-grained disease classification tasks. This structure enables AI models to perform generalized disease prediction as well as detailed disease-specific analysis.

Another important advantage of MultiCaRe is the availability of image-text relationships within each clinical case. Medical images are associated with detailed clinical descriptions, physician observations, diagnosis summaries, and patient-related information. This multimodal pairing supports tasks such as image-text representation learning, automated report generation, visual question answering, symptom-based disease prediction, and multimodal clinical reasoning [21].

The reviewed literature also emphasized that MultiCaRe supports the development of explainable and context-aware healthcare AI systems because it combines imaging findings with clinical narratives and structured metadata. The dataset can therefore be effectively used for multimodal fusion approaches integrating CNN-based image analysis and NLP-based symptom understanding.

For the proposed project, MultiCaRe is highly relevant because it directly supports the

development of a multimodal Clinical Decision Support System capable of processing both medical images and patient symptom descriptions. The dataset provides sufficient multimodal information for training disease prediction models, generating confidence-based outputs, retrieving similar clinical cases, and producing AI-assisted clinical reports.

Despite its advantages, several limitations of MultiCaRe were identified in the literature [17], [21]. These include annotation inconsistency, heterogeneous image quality, imbalance among disease categories, and overrepresentation of rare clinical cases. Variations in textual descriptions and imaging modalities may also affect model generalization and prediction stability. In addition, preprocessing and careful disease selection are necessary to reduce dataset imbalance and improve training performance.

Nevertheless, MultiCaRe remains a valuable benchmark dataset for multimodal healthcare AI research due to its large-scale image-text structure, diverse disease coverage, and support for multimodal clinical reasoning applications.

V. DISCUSSION

The reviewed literature clearly demonstrates the evolution of clinical diagnosis systems from rule-based healthcare frameworks to advanced multimodal AI architectures. Early studies mainly focused on structured healthcare data obtained from EHRs, laboratory reports, and demographic records for disease prediction and clinical analytics [1], [3]. Although these systems improved predictive analysis, they lacked the capability to process visual diagnostic information.

The introduction of CNN-based medical imaging systems significantly improved automated disease diagnosis using MRI, CT scan, X-ray, and retinal imaging data [6], [7]. These systems achieved high accuracy in tumor detection, cardiovascular analysis, and neurological diagnosis. However, image-only systems could not incorporate patient symptoms or contextual clinical information required for real-world diagnosis [8].

To address this limitation, researchers increasingly applied NLP techniques for processing unstructured healthcare text including symptom descriptions, physician notes, and discharge summaries [9]. Transformer-based architectures and LLMs improved semantic understanding of clinical narratives and enabled automated symptom analysis

and healthcare text interpretation [15]. Nevertheless, NLP systems alone lacked the capability to analyze visual medical abnormalities.

The major transition identified in the literature was the shift toward multimodal AI systems integrating medical images, structured healthcare data, and clinical text into unified frameworks [10], [11]. These systems improved diagnostic performance by combining complementary information from multiple healthcare modalities. The reviewed studies showed that multimodal fusion approaches achieved better prediction accuracy and robustness compared to unimodal systems in applications such as cancer diagnosis, cardiovascular disease prediction, and neurological disorder analysis [10], [12].

Another important trend observed in the literature was the emergence of foundation models, vision-language architectures, and multimodal LLMs for healthcare applications [13], [14]. These systems enabled cross-modal reasoning, automated report generation, image captioning, and conversational healthcare support. Generative AI and RAG-based architectures further improved clinical assistance through multimodal reasoning and knowledge retrieval [15], [16].

The literature also emphasized the importance of multimodal datasets such as MultiCaRe for developing generalized healthcare AI systems [21]. The dataset provides paired image-text healthcare information that supports multimodal disease classification, report generation, and clinical reasoning. However, issues such as annotation inconsistency, class imbalance, and heterogeneous imaging modalities remain important challenges for multimodal healthcare AI research [17].

Overall, the reviewed studies indicate a clear transition from isolated unimodal healthcare systems toward integrated multimodal clinical diagnosis systems capable of simultaneously analyzing medical images, symptoms, and structured healthcare data for improved clinical decision-making.

VI. RESEARCH GAP

Although significant progress has been achieved in multimodal healthcare AI, the reviewed literature reveals several important research gaps. Many existing systems still focus on individual modalities such as medical imaging, symptom analysis, or structured

healthcare records separately [6], [9]. Fully integrated multimodal systems combining medical images, patient symptoms, structured healthcare data, explainability, and report generation remain limited.

Several studies also highlighted the lack of clinically deployable systems capable of providing confidence scores, explainable predictions, and retrieval of similar past clinical cases [15], [16]. In addition, many multimodal models require large computational resources and complex training pipelines, limiting real-world healthcare deployment.

The literature further identified limitations in multimodal datasets including annotation inconsistency, disease imbalance, and heterogeneous imaging modalities [17], [21]. Another major research gap involves explainability and physician trust, as many AI systems continue to operate as black-box models [19].

These gaps justify the need for a multimodal AI-based CDSS capable of integrating medical image analysis, symptom understanding, explainability mechanisms, confidence estimation, similar case retrieval, and AI-generated clinical reporting within a unified healthcare framework.

VII. PROPOSED MULTIMODAL CDSS FRAMEWORK

Based on the reviewed literature, the proposed system aims to develop a multimodal AI-based Clinical Decision Support System capable of disease prediction using medical images and patient symptom descriptions. The system mainly focuses on integrating CNN-based image analysis and NLP-based symptom understanding using the MultiCaRe dataset.

The proposed system accepts three types of input:

- Medical image input
- Symptom description input
- Combined image and symptom input

The image processing module uses CNN architectures for extracting disease-related image features from modalities such as MRI, CT scan, and X-ray images. Simultaneously, the NLP module processes patient symptom descriptions and clinical text to generate semantic feature representations.

A multimodal fusion layer combines image features and symptom embeddings to generate a unified patient representation for disease prediction. The system further provides:

- Disease prediction output
- Confidence score
- Explainability heatmaps using Grad-CAM
- Similar case retrieval
- AI-generated clinical report

The proposed framework addresses several research gaps identified in the literature by integrating multimodal diagnosis, explainability, and clinical assistance into a unified healthcare AI system.

VIII. CHALLENGES, LIMITATIONS, AND FUTURE DIRECTIONS

The literature identified several challenges affecting multimodal clinical diagnosis systems. One major challenge is healthcare data heterogeneity because medical images, symptom descriptions, laboratory records, and clinical reports differ significantly in format and quality [17]. This complicates multimodal fusion and model generalization.

Another important limitation involves annotation inconsistency and class imbalance in multimodal healthcare datasets such as MultiCaRe [21]. Rare disease categories and inconsistent labeling can affect prediction performance and training stability.

Explainability remains another critical challenge because many multimodal AI systems function as black-box models [19]. Several studies emphasized the need for explainable AI techniques such as Grad-CAM, attention visualization, and interpretable prediction mechanisms to improve physician trust and clinical reliability.

The reviewed papers also discussed challenges related to computational complexity, deployment cost, privacy, and integration with existing healthcare systems [13], [14]. Foundation models and multimodal LLMs require large computational resources and extensive training data, making practical deployment difficult in resource-constrained healthcare environments.

Future research should focus on developing lightweight multimodal architectures, improving explainability, enhancing dataset quality, and enabling secure clinical deployment. Integration of multimodal LLMs, RAG systems, federated learning, and personalized healthcare models may further improve intelligent clinical decision support systems.

IX. CONCLUSION

This survey reviewed recent developments in multimodal clinical diagnosis systems and summarized the evolution of AI-based Clinical Decision Support Systems from rule-based approaches to multimodal healthcare AI architectures. The reviewed literature showed that structured healthcare data systems, CNN-based medical imaging models, NLP-based symptom analysis systems, and multimodal fusion approaches collectively contributed to improved disease diagnosis and clinical prediction.

The literature also demonstrated that multimodal AI systems outperform unimodal systems by integrating complementary information from medical images, patient symptoms, and structured healthcare data. Recent developments in foundation models, multimodal LLMs, and generative AI further expanded the capabilities of healthcare AI systems through cross-modal reasoning and automated clinical assistance.

Despite these advancements, several challenges remain, including data heterogeneity, explainability limitations, annotation inconsistency, and computational complexity. The proposed multimodal CDSS perspective addresses these challenges by integrating image analysis, symptom understanding, explainability, confidence estimation, and AI-generated reporting within a unified healthcare framework. Future advancements in multimodal healthcare AI are expected to significantly improve intelligent clinical diagnosis and patient-centered healthcare systems.

REFERENCES

1. N. Saeidi and M. Torabi, “Artificial intelligence and clinical decision making,” Journal of Artificial Intelligence and Sustainability in Society (JAISIS), vol. 3, no. 2, pp. 43–55, 2022.
2. Alnattah et al., “Artificial intelligence in clinical decision-making: A scoping review,” Cureus, vol. 17, no. 1, 2025.
3. S. Shetty, V. S. Ananthanarayana, and A. Mahale, “Comprehensive review of multimodal medical data analysis,” Acta Informatica Pragensia, vol. 11, no. 3, pp. 423–

457, 2022.

4. Milenkovic et al., “Healthcare AI for physician-centered decision-making,” Computers, vol. 14, no. 320, 2025.
5. Giordano et al., “Accessing artificial intelligence for clinical decision-making,” Frontiers in Digital Health, vol. 3, 2021.
6. *Q. Cai et al., “A survey on multimodal data-driven smart healthcare systems,” IEEE Access, vol. 7, pp. 133583–133599, 2019.*
7. *P. Arora et al., “Multimodal medical image analysis using deep learning,” Discover Computing, vol. 1, no. 1, 2026.*
8. *F. Krones et al., “Review of multimodal machine learning approaches in healthcare,” Information Fusion, vol. 114, p. 102690, 2025.*
9. *B. Jandoubi and M. Akhloufi, “Multimodal artificial intelligence in medical diagnostics,” Information, vol. 16, no. 591, 2025.*
10. *R. Ravikumar et al., “AI-enhanced clinical decision-making through a collaborative approach,” in Artificial Intelligence Applications in Healthcare, IGI Global, 2026.*
11. *S. Huang et al., “Multimodal foundation models for medical imaging,” medRxiv, 2024.*
12. *M. Imran and Y. Lee, “Multimodal vision-language models in medical imaging,” IEEE Access, 2026.*
13. *L. Buess et al., “From large language models to multimodal AI in medicine,” Biomedical Engineering Letters, 2025.*
14. *M. Kayaalp et al., “ChatGPT, DeepSeek, and generative AI in healthcare,” preprint, 2025.*
15. *Agarwal et al., “Toward inclusive healthcare: An LLM-based multimodal chatbot for preliminary diagnosis,” IEEE Access, 2025.*
16. *Agarwal et al., “Multimodal RAG systems for healthcare assistance,” preprint, 2025.*
17. *Hosseinzadeh et al., “Data quality assessment in healthcare AI systems,” BMC Medical Informatics and Decision Making, 2025.*
18. *Amann et al., “Explainability for artificial intelligence in healthcare,” BMC Medical Informatics and Decision Making, vol. 20, no. 1, 2020.*
19. *S. Thomas et al., “Artificial intelligence in modern clinical practice,” Medicine International, 2026.*
20. *P. Arora et al., “Deep multimodal image fusion techniques for healthcare imaging,” Discover Computing, 2026.*
21. *Nievas Offidani et al., “An open-source clinical case dataset for medical image classification and multimodal AI applications,” Data, vol. 10, no. 123, 2025.*