



An Intelligent Multimodal and Multilingual Clinical Decision Support System Using LLM and RAG with Explainable Medical Insights: A Survey

Gangadhar R

Dept. of Computer Science and
Engineering
RV Institute of Technology and
Management
Bengaluru, Karnataka, India
gangadharr414@gmail.com

Gururaj S H

Dept. of Computer Science and
Engineering
RV Institute of Technology and
Management
Bengaluru, Karnataka, India
hiremathgururaj073@gmail.com

M G Nikhil

Dept. of Computer Science and
Engineering
RV Institute of Technology and
Management
Bengaluru, Karnataka, India
mgnikhil9456@gmail.com

Mallikarjunayya S

Dept. of Computer Science and
Engineering
RV Institute of Technology and
Management
Bengaluru, Karnataka, India
salimallu56@gmail.com

Mrs. Bhavya N Javagal

Assistant Professor
Dept. of Computer Science and Engineering
RV Institute of Technology and Management
Bengaluru, Karnataka, India
bhavyanj.rvitm@rvei.edu.in

How to Cite this Article:

R, G., H, G. S., Nikhil, M. G. & S, M. (2026). An Intelligent Multimodal and Multilingual Clinical Decision Support System Using LLM and RAG with Explainable Medical Insights: A Survey. International Journal of Creative and Open Research in Engineering and Management, 2(10), 1-9. <https://doi.org/10.55041/ijcope.v2i9.241>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i9.241>

Abstract— Modern healthcare environments generate heterogeneous clinical data spanning symptoms, laboratory reports, radiological images, vitals, and patient history, yet most existing Clinical Decision Support Systems (CDSS) remain confined to structured datasets and lack multimodal integration, multilingual accessibility, and explainability. This paper surveys recent Large Language Model (LLM) and Retrieval-Augmented Generation (RAG) based clinical reasoning systems and proposes an Intelligent Multimodal and Multilingual Clinical Decision Support System that unifies multilingual symptom normalization, Machine Learning based disease screening (XGBoost), Deep Learning based chest X-ray analysis (DenseNet121) and fracture detection (EfficientNet-B0), RAG-based evidence retrieval from PubMed, MedlinePlus and PubMedQA, ICD-10 standardization through the WHO ICD API, and automated PDF report generation. A systematic review of ten recent works reveals a consistent gap: existing systems address either a single modality, a narrow disease category, or lack standardized interoperable ICD output, and none provide integrated multilingual accessibility for low-resource regional languages such as Kannada. The proposed Fusion Engine architecture directly addresses this gap by combining symptom-derived, laboratory, vital-sign, and imaging-derived evidence with retrieval-grounded medical literature into a single explainable diagnostic hypothesis, further supporting specialist recommendation and GeoHealth analytics for regional disease surveillance.

Index Terms— Clinical Decision Support System (CDSS); Retrieval-Augmented Generation (RAG); Large Language Models (LLM); Multimodal Fusion; Multilingual NLP; Explainable AI; ICD Mapping; Chest X-ray Classification; Fracture Detection; GeoHealth Analytics.

I. INTRODUCTION

Healthcare systems worldwide generate massive volumes of clinical information every day, including patient-reported symptoms, laboratory reports, medical imaging, prescriptions, and electronic health records. Traditional Clinical Decision Support Systems have historically focused on structured, single-modality datasets and rule-based logic, and consequently they struggle to integrate heterogeneous inputs such as free-text symptoms, radiological images, and laboratory values into a single coherent diagnostic narrative. This limitation is compounded by the absence of multilingual support, which restricts accessibility for non-English-speaking patient populations, and by the absence of explainability, which limits clinician trust in AI-generated recommendations.

The recent maturation of Large Language Models (LLMs), Deep Learning, and Retrieval-Augmented Generation (RAG) has opened new opportunities for building intelligent, evidence-grounded healthcare systems capable of multimodal reasoning. This paper first surveys the recent literature on LLM- and RAG-based clinical decision support to identify the state of the art and its limitations, and then presents the design of an Intelligent Multimodal and Multilingual Clinical Decision Support System that addresses the identified gaps.

II. BACKGROUND AND MOTIVATION

Modern healthcare environments face increasing challenges such as delayed diagnosis, fragmented patient information, rising patient load, and limited specialist availability, particularly in the difficulty of jointly interpreting symptoms, laboratory reports, vitals, and imaging. Existing healthcare software systems commonly operate as isolated modules and fail to provide unified, explainable, AI-assisted clinical reasoning across modalities.

Retrieval-Augmented Generation mitigates a well-known limitation of LLMs in clinical settings, namely hallucination, by



grounding generated text in retrieved evidence from trusted medical sources such as PubMed and MedlinePlus. Combined with Deep Learning based imaging models and ML-based structured-data screening, RAG creates the opportunity to build a genuinely multimodal clinical reasoning pipeline rather than a text-only chatbot.

III. OBJECTIVES

- Develop an intelligent multimodal Clinical Decision Support System.
- Perform disease screening using symptoms and structured clinical data.
- Detect thoracic disease using chest X-ray Deep Learning models.
- Detect musculoskeletal fractures using Deep Learning techniques.
- Support multilingual symptom understanding (English, Kannada, and mixed input).
- Implement Retrieval-Augmented Generation for evidence-backed reasoning.
- Generate structured, AI-assisted PDF medical reports.
- Implement ICD-based disease standardization using the WHO ICD API.
- Provide GeoHealth analytics and regional disease-trend monitoring.
- Support explainable AI-assisted clinical reasoning throughout the pipeline.

IV. OVERVIEW OF THE PROPOSED SYSTEM

The proposed system follows a modular pipeline architecture. Patient symptoms, vitals, laboratory reports, and medical images are accepted as multimodal inputs. Multilingual symptom processing converts patient input into standardized medical terminology, after which an XGBoost-based Machine Learning engine performs disease screening and prioritization. In parallel, a DenseNet121-based model performs chest X-ray abnormality detection and an EfficientNet-B0-based model performs fracture detection on musculoskeletal radiographs. A RAG layer retrieves trusted medical evidence from PubMed, MedlinePlus, and PubMedQA, which is combined with ICD-10 mapping obtained through the WHO ICD API. A Fusion Engine finally integrates all of these signals into a structured diagnostic hypothesis, from which specialist recommendations, GeoHealth analytics, and AI-assisted PDF clinical reports are generated.

V. LITERATURE SURVEY

This section presents a systematic review of ten recent works related to LLM- and RAG-based clinical decision support, multimodal medical reasoning, and imaging-based diagnostic AI.

A. Knowledge-Graph and Retrieval-Augmented Clinical Reasoning

Zhao et al. [1] propose MedRAG, which combines Retrieval-Augmented Generation with a knowledge graph to improve diagnostic accuracy, though the resulting knowledge-graph design is architecturally complex. Zhu et al. [2] present REALM, which applies RAG to multimodal Electronic Health Record (EHR) data and reduces hallucination, but at the cost of high system complexity. Shukla and Bhati [3] review RAG methods for clinical decision support and identify open research gaps without proposing an implemented system, while Ke et al. [4] evaluate an LLM-RAG pipeline in a real clinical setting, providing valuable empirical evaluation but within a limited domain scope.

B. Multimodal and Multi-Agent LLM Systems

A broader survey by Lui [5] reviews multimodal LLM applications across healthcare but does not propose a concrete implemented system. Agarwal et al. [6] design an LLM-based multimodal chatbot that jointly processes text and images for diagnosis, offering genuine multimodal support but remaining susceptible to hallucination risk. Öğdü et al. [7] propose an adaptive multi-agent LLM architecture for clinical decision support, achieving advanced reasoning capability at the expense of substantially increased system complexity.

C. Clinician Workflow Support

Shetgaonkar et al. [8] apply RAG to mitigate clinician information overload, demonstrating reliable retrieval performance; however, the resulting system remains largely text-based and does not incorporate imaging or structured vitals data.

D. Chronic Disease Management and Medical Imaging Reviews

Serugunda and Jianquan [9] review LLM applications in chronic disease management workflows and associated evaluation metrics, offering a useful overview of clinical workflows without an implemented system. Wang et al. [10] similarly review LLM applications in medical imaging, providing valuable imaging-specific insight but stopping short of a unified Clinical Decision Support System.

VI. CATEGORIZATION OF RELATED WORKS

Table I categorizes the ten surveyed works into four groups based on their principal technical contribution and highlights the limitation that motivates the proposed system.

TABLE I
CATEGORIZATION OF SURVEYED WORKS

Category	Works	Technique	Limitation
KG / Retrieval Reasoning	[1]-[4]	KG-RAG, multimodal EHR RAG	Complex / narrow domain
Multimodal / Multi-Agent LLM	[5]-[7]	Multimodal fusion, multi-agent	Hallucination / complexity
Clinician Workflow Support	[8]	RAG-assisted retrieval	Text-only, no imaging
Chronic Care / Imaging Reviews	[9] [10]	Literature review	No unified CDSS



VII. RESEARCH GAP AND MOTIVATION

The systematic review reveals a consistent gap: no surveyed work integrates multilingual symptom understanding, structured-data disease screening, dual imaging modalities (chest X-ray and musculoskeletal fracture detection), RAG-grounded evidence retrieval, ICD-10 standardization, and regional disease surveillance within a single explainable pipeline. Works closest to this goal, such as MedRAG [1] and REALM [2], remain limited to knowledge-graph or EHR-only retrieval without imaging fusion, while multimodal chatbot systems [6] incorporate imaging but not multilingual accessibility or ICD interoperability. The proposed system directly addresses this gap through a unified Fusion Engine architecture described in Section VIII.

VIII. PROPOSED METHODOLOGY

The proposed system combines Machine Learning, Deep Learning, RAG, multilingual Natural Language Processing, imaging AI, and rule-based reasoning into a unified healthcare workflow, illustrated as a thirteen-stage pipeline.

Patient symptoms, laboratory reports, vitals, and medical imaging are accepted as inputs. Multilingual symptom processing and normalization convert patient input into standardized medical terminology, which is then passed to an XGBoost-based disease screening engine for disease prioritization. The imaging module performs chest X-ray abnormality detection using DenseNet121 and fracture detection using EfficientNet-B0. Concurrently, the RAG layer retrieves trusted medical evidence from PubMed, MedlinePlus, and PubMedQA to improve explainability and reduce hallucination. A Fusion Engine integrates symptoms, imaging findings, laboratory analysis, vitals, RAG evidence, and ICD mapping into a structured clinical workflow, from which specialist recommendations, GeoHealth analytics, and AI-assisted PDF medical reports are generated.

IX. SYSTEM MODULES

Patient Input Module: Collects symptoms, vitals, lab reports, and medical images as the entry point of the system.

Multilingual Processing Module: Supports English, Kannada, and mixed-language input, converting regional-language symptoms into medical terminology.

Symptom Normalization Module: Converts raw symptom text into standardized clinical terms via synonym mapping (e.g., “high temperature” → “fever”).

Disease Prediction Module: Uses XGBoost and rule-based reasoning to generate disease probability scores and ranking.

Vitals and Lab Analysis Module: Analyzes blood pressure, oxygen saturation, CBC, and sugar levels to generate risk alerts.

Chest X-ray AI Module: Uses DenseNet121 for multi-label detection of pneumonia, pleural effusion, cardiomegaly, and lung opacity.

Fracture Detection Module: Uses EfficientNet-B0 on musculoskeletal radiographs to generate fracture and trauma alerts.

RAG Evidence Retrieval Module: Retrieves trusted evidence from PubMed and MedlinePlus to reduce hallucination and improve reasoning.

ICD Mapping Module: Maps predicted diseases to standardized ICD codes using the WHO ICD API for interoperability.

Fusion Engine: Combines symptoms, imaging, vitals, lab analysis, RAG evidence, and ICD mapping into a confidence-weighted diagnostic hypothesis.

Clinical Summary Generation Module: Generates structured, human-readable clinical summaries with cited evidence sources.

Specialist Recommendation Module: Suggests appropriate specialists (pulmonologist, cardiologist, orthopedic, etc.) based on screening results.

GeoHealth Analytics Module: Supports regional outbreak analysis, disease clustering, and public-health trend monitoring.

X. HARDWARE AND SOFTWARE REQUIREMENTS

The proposed system requires only moderate computing resources, as summarized in Table II.

TABLE II
HARDWARE AND SOFTWARE STACK

Component	Specification / Tool
Processor	Intel i5 / Ryzen 5 or above
RAM	Minimum 8 GB
Storage	256 GB SSD or above
GPU	NVIDIA GPU (optional, for DL training)
Frontend	React.js, Vite, Tailwind CSS
Backend	Python, FastAPI
ML / DL	Scikit-learn, XGBoost, PyTorch
Database	SQLite
NLP / Models	HuggingFace Transformers
External APIs	WHO ICD API, OpenFDA API, PubMed API
Tooling	VS Code, Git & GitHub

XI. EXPECTED OUTCOMES AND APPLICATIONS

The proposed system is expected to produce reliable multimodal disease screening, chest X-ray and fracture detection, multilingual symptom understanding, RAG-grounded evidence-based reasoning, ICD-standardized reporting, and GeoHealth analytics for regional disease surveillance. Direct applications include triage support in outpatient departments, decision support for general physicians in low-resource or multilingual regions, automated radiology-assist tools for musculoskeletal and thoracic imaging, and public-health dashboards for regional disease-trend monitoring.

XII. ADVANTAGES

- Unified multimodal reasoning across symptoms, vitals, labs, and imaging, unlike single-modality prior works [1] [6].



- Reduced hallucination through RAG grounding in PubMed and MedlinePlus, consistent with findings in [2] and [4].
- Multilingual accessibility (English/Kannada) not addressed in any surveyed work.
- ICD-10 standardized, interoperable clinical output suitable for downstream health-record systems.
- Explainable evidence citation improves clinician trust relative to black-box chatbot approaches [6].

XIII. CHALLENGES AND LIMITATIONS

Fusing heterogeneous modalities (text, images, structured vitals) with well-calibrated confidence weighting is non-trivial, and the Fusion Engine's reliability depends on the quality of each upstream model. Multilingual symptom normalization for mixed-language, code-switched input (e.g., English-Kannada) is linguistically harder than monolingual normalization. RAG evidence quality depends on the coverage and currency of the underlying PubMed/MedlinePlus corpus, and imaging models such as DenseNet121 and EfficientNet-B0 require sufficiently large, well-labeled regional datasets to generalize reliably. As with any AI-assisted CDSS, the system is intended to support, not replace, clinician judgment.

XIV. FUTURE SCOPE

Future work includes extending multilingual support to additional Indian regional languages, incorporating additional imaging modalities such as CT and MRI, integrating federated learning to preserve patient-data privacy across institutions, and validating the Fusion Engine's diagnostic-hypothesis quality against clinician-annotated ground truth in a prospective clinical study. Extending the ICD mapping module to SNOMED-CT would further improve interoperability with international health-record standards.

XV. CONCLUSION

This paper has surveyed ten recent works on LLM- and RAG-based clinical decision support and identified a consistent gap: no existing system integrates multilingual symptom understanding, multimodal imaging, RAG-grounded evidence retrieval, ICD standardization, and regional disease surveillance within a single explainable pipeline. The proposed Intelligent Multimodal and Multilingual Clinical Decision Support System directly addresses this gap through a Fusion Engine that combines ML-based disease screening, DenseNet121/EfficientNet-B0 imaging analysis, RAG-based evidence retrieval, and WHO ICD mapping into a unified, explainable, multilingual clinical-reasoning workflow, with GeoHealth analytics providing an additional public-health dimension absent from prior work.

REFERENCES

- [1] X. Zhao, S. Liu, and S.-Y. Yang, "MedRAG: Enhancing Retrieval-Augmented Generation with Knowledge Graph-Elicited Reasoning for Healthcare Copilot," arXiv preprint arXiv:2502.04413, 2025.
- [2] Y. Zhu, C. Ren, et al., "REALM: RAG-Driven Enhancement of Multimodal Electronic Health Records Analysis via Large Language Models," arXiv preprint, 2024.
- [3] N. Shukla and D. Bhati, "Retrieval-Augmented Generation in Healthcare: A Survey," 2025.
- [4] Y. Ke, L. Jin, K. Elangovan, et al., "Development and Testing of Retrieval-Augmented Generation in Large Language Models: A Case Study," 2024.
- [5] Lui, "Multimodal Large Language Models in Health Care: A Review," 2024.
- [6] I. Agarwal, V. Sakthivel, and P. Prakash, "LLM-Based Multimodal Chatbot for Clinical Diagnosis," 2024.
- [7] Ç. U. Ögdü, K. Arslanoğlu, and M. Karaköse, "Adaptive Multi-Agent LLM Clinical Decision Support," 2024.
- [8] A. Shetgaonkar, D. Pradhan, and L. Arora, "Mitigating Clinician Information Overload using Generative AI and RAG," 2024.
- [9] H. M. Serugunda and O. Jianquan, "Large Language Models for Chronic Disease Management," 2025.
- [10] P. Wang, W. Lu, and C. Lu, "Large Language Models for Medical Images: A Review," 2023.
- [11] Q. Liu, R. Yang, Q. Gao, T. Liang, X. Wang, S. Li, B. Lei, and K. Gao, "A Review of Applying Large Language Models in Healthcare," IEEE Access, vol. 13, pp. 6878–6892, 2025.
- [12] L. Liu, X. Yang, J. Lei, Y. Shen, J. Wang, P. Wei, Z. Chu, Z. Qin, and K. Ren, "A Survey on Medical Large Language Models: Technology, Application, Trustworthiness, and Future Directions," arXiv preprint arXiv:2406.03712, 2024.
- [13] X. Yang, A. Chen, N. PourNejatian, H. C. Shin, K. E. Smith, C. Parisien, C. Compas, C. Martin, A. B. Costa, M. G. Flores, Y. Zhang, T. Magoc, C. A. Harle, G. Lipori, D. A. Mitchell, W. R. Hogan, E. A. Shenkman, J. Bian, and Y. Wu, "A Large Language Model for Electronic Health Records (GatorTron)," npj Digital Medicine, vol. 5, 2022, arXiv:2203.03540.
- [14] A. Hernández Cano, A. Romanou, A. Bonnet, Z. Chen, K. Matoba, F. Salvi, M. Pagliardini, S. Fan, A. Köpf, A. Sallinen, A. Mohtashami, A. Sakhaeirad, V. Swamy, I. Krawczuk, D. Bayazit, A. Marmet, S. Montariol, M.-A. Hartley, M. Jaggi, and A. Bosselut, "MEDITRON-70B: Scaling Medical Pretraining for Large Language Models," arXiv preprint arXiv:2311.16079, 2023.
- [15] A. Pal, L. K. Umapathi, and M. Sankarasubbu, "Med-HALT: Medical Domain Hallucination Test for Large Language Models," arXiv preprint arXiv:2307.15343, 2023.
- [16] E. Bolton, A. Venigalla, M. Yasunaga, D. Hall, B. Xiong, T. Lee, R. Daneshjou, J. Frankle, P. Liang, M. Carbin, and C. D. Manning, "BioMedLM: A 2.7B Parameter Language Model Trained on Biomedical Text," arXiv preprint arXiv:2403.18421, 2024.
- [17] R. Luo, L. Sun, Y. Xia, T. Qin, S. Zhang, H. Poon, and T.-Y. Liu, "BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining," Briefings in Bioinformatics, vol. 23, 2022, arXiv:2210.10341.
- [18] K. Zhang, R. Zhou, E. Adhikarla, Z. Yan, Y. Liu, J. Yu, Z. Liu, X. Chen, B. D. Davison, H. Ren, J. Huang, C. Chen, Y. Zhou, S. Fu, W. Liu, T. Liu, X. Li, Y. Chen, L. He, J. Zou, Q. Li, H. Liu, and L. Sun, "BiomedGPT: A Generalist Vision-Language Foundation Model for Diverse Biomedical Tasks," arXiv preprint arXiv:2305.17100, 2023.



- [19] Q. Chen, Y. Hu, X. Peng, Q. Xie, Q. Jin, A. Gilson, M. B. Singer, X. Ai, P.-T. Lai, Z. Wang, V. K. Keloth, K. Raja, J. Huang, H. He, F. Lin, J. Du, R. Zhang, W. J. Zheng, R. A. Adelman, Z. Lu, and H. Xu, “Benchmarking Large Language Models for Biomedical Natural Language Processing Applications and Recommendations,” *Nature Communications*, 2025, arXiv:2305.16326.
- [20] A. Pal, L. K. Umapathi, and M. Sankarasubbu, “Med-HALT: Medical Domain Hallucination Test for Large Language Models,” in *Proc. 27th Conf. Computational Natural Language Learning (CoNLL)*, Singapore, 2023, pp. 314–334.