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



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151171

MedAgentix AI: Multi-Agent Medical Expert System for Diagnostic Support

Sayali Bharat Jadhav[1], Priyanka Devendra Lanjewar[2], Tanisha Chudaman Badgujar[3], Khushi Chandrakant Ahire[4], Rehan Rahuf Shekh[5]

UG Student, Department of Artificial Intelligence and Machine Learning, R. C. Patel Institute of Technology, Shirpur, India[1,3,4,5]

Assistant Professor, Department of Artificial Intelligence and Machine Learning, R. C. Patel Institute of Technology, Shirpur, India[2]

Abstract: Diagnostic errors and delays in emergency triage remain problems in healthcare. Artificial intelligence can help doctors make decisions; however many systems out there don't explain why they make the predictions they do. This paper introduces MedAgentix AI, a multi-agent clinical decision support system. MedAgentix AI is designed to help with diagnosis, disease risk assessment and emergency triage. The system uses machine learning, deep learning and natural language processing. These tools work together through agents in a single workflow. MedAgentix AI takes symptoms from what patients say makes predictions checks for possible risks and looks through medical documents to find important information. It also checks for emergency warning signs to spot cases. Retrieval-Augmented Generation gives doctors the medical information to answer questions and handle cases where the system is not sure. The system creates reports for doctors and for patients. The data, behind MedAgentix AI includes 246,945 cases. These cover 773 conditions and 377 standard symptoms. The systems multi-class diagnostic model works with 721 types of diseases. The findings show that multi-agent intelligence and explainable artificial intelligence can help make clinical decisions more transparent. More testing is needed to make sure MedAgentix AI is reliable and can be used in hospitals.

Keywords: Artificial Intelligence in Healthcare, Clinical Decision Support Systems, Multi-Agent Systems, Explainable Artificial Intelligence, Differential Diagnosis, Emergency Triage, Natural Language Processing, Machine Learning.

I. INTRODUCTION

Clinical diagnosis becomes difficult when patients show symptoms that can belong to many diseases. Clinicians must look at each patient's symptoms past records, other illnesses and all other relevant clinical information to give a correct assessment. These decisions become harder in emergency rooms and outpatient clinics, where quick assessment is essential. Clinical decision support gives a way to organize patient data and help doctors make evidence-based choices. Yet the ability to adapt, to be clear about how a decision was made and to mix medical knowledge into the system remain big challenges for AI-based healthcare systems [1], [5].

Machine learning and natural language processing help with diagnosis by spotting patterns in medical records and pulling useful details from what patients say. Still many predictive models offer explanation and large language models can sound medically correct while giving wrong facts. Multi-agent systems offer another option: they split a job into smaller parts letting specialized agents work together to produce a clear assessment. When combined with artificial intelligence and quick access to biomedical facts these methods can make the process more open and give doctors useful evidence for clinical decision making [2]–[5].

To meet these needs this paper introduces MedAgentix AI, a multi-agent decision support framework that helps doctors, with initial patient assessment listing possible diagnoses checking disease risk and triaging emergencies. The framework uses agents to examine symptoms produce diagnostic guesses, spot possible emergency warning signs and pull in relevant biomedical facts. The system blends prediction, reasoning, safety checks and evidence gathering into one smooth workflow. It is built to aid healthcare professionals not to replace their judgment.

II. LITERATURE REVIEW AND RESEARCH GAP

Clinical decision support systems are very important for making diagnoses and helping with healthcare that is based on evidence. Thorakkattil et al. [1] Talked about the benefits of these systems and the problems that come with using them

 This work is licensed under a Creative Commons Attribution 4.0 International License

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151171

in healthcare like putting data together and making different systems work with each other. Earlier work by Gupta and Mukhopadhyay [2]. By Chakraborty and Gupta [3] looked at multi-agent systems for medical diagnosis. They focused on how agents work share knowledge and handle tasks that are spread out in the medical field.

New research has started to use language models with multi-agent systems. Chen et al. [4] Shared an agent-based method for looking at lab results patient measurements and the clinical situation. They made sure the system was clear and followed rules for AI. Öğdü et al. [5] Created a multi-agent system for clinical decision support that includes biomedical research and web information. They pointed out the need for quality information lower costs and testing in real situations.

Wu et al. [6] Created a system that uses an orchestrator with multi-agents to help with diagnosing headaches in care.. Ndlovu [7] studied how to improve decision support using multi-agent systems that generate information. They talked about the issues like the quality of information how useful it is for doctors the cost and the need to protect privacy and be responsible.

Research Gap: Current studies show that multi-agent diagnosis and biomedical information retrieval are getting better.. There are still problems with combining diagnosis predictions explaining how the system works, checking how confident it is handling emergencies and getting reliable medical information. These problems are why the MedAgentix AI framework was created. It brings together agents in one system, for clinical decision support.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Overall System Architecture

MedAgentix AI uses specialized agents coordinated through LangGraph to support clinical assessment, disease prediction, risk evaluation and emergency triage. The agents process medical information, perform specific diagnostic tasks and contribute to a final assessment. A document intelligence component extracts information from medical files, while the triage component identifies potential emergency warning signs. A supervisor agent combines the outputs, evaluates predictions and considers their confidence levels. The system generates a clinical report for healthcare professionals and a simplified health guide for patients.

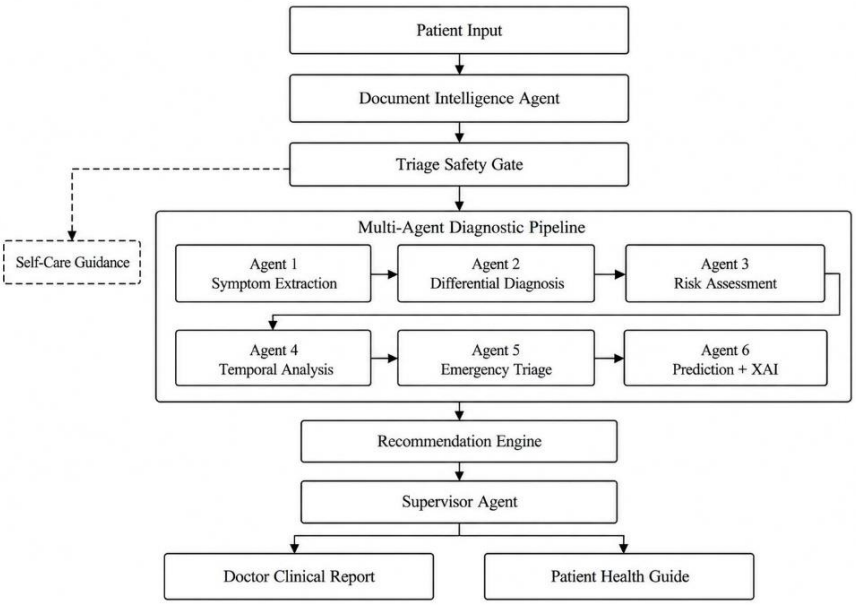


Fig. 1. Overall architecture of the MedAgentix AI multi-agent clinical decision support system.

B. Clinical Data Processing

The system converts reported symptoms, medical history and physical measurements into a standardized format suitable for machine learning. Medical terms are normalized so that different expressions of the same symptom or condition can be matched consistently. This standardized information serves as input for diagnostic and risk-assessment agents.

C. Machine Learning Models and Diagnostic Algorithms

MedAgentix AI employs different machine learning techniques according to the requirements of each agent. Bio_ClinicalBERT processes clinical language, identifies medical terms and extracts relevant information from symptom

DOI: 10.17148/IJARCCE.2026.151171

descriptions. A multi-class XGBoost classifier uses boosted decision trees to generate likely disease predictions. A separate risk-assessment component evaluates information associated with potential long-term conditions. For emergency assessment, the system combines predefined warning rules with multinomial logistic regression to identify potentially critical cases. The supervisor agent integrates these outputs into a unified assessment.

D. Confidence Assessment and Explainable AI

The supervisor agent evaluates the confidence levels and agreement of diagnostic agents. Depending on confidence, the system may use an individual model prediction, combine multiple predictions or request additional medical information through a language model. These confidence-based rules guide the workflow but do not guarantee diagnostic correctness. MedAgentix AI uses SHAP and LIME to identify patient features that influence model predictions. These explanations provide additional context and improve the interpretability of the diagnostic outputs.

E. Biomedical Knowledge Retrieval

The system incorporates Retrieval-Augmented Generation (RAG) to provide relevant biomedical information during clinical assessment. PubMedBERT embeddings are used to retrieve relevant text from a stored medical knowledge set. The retrieved information provides context to the language model when generating explanations and health advice. This component supports the predictive models rather than replacing them.

IV. EXPERIMENTAL SETUP AND EVALUATION

A. Dataset and Evaluation Cohort

The dataset contains approximately 246,945 clinical case instances, covering 773 indexed medical conditions and 377 standardized symptom features. The multi-class differential diagnosis model predicts across 721 disease categories. The dataset is derived from open-access benchmark resources and disease-symptom datasets. The evaluation focuses on benchmark-based model performance. Independent validation using a separately curated clinical cohort or clinician-reviewed cases has not yet been performed.

B. Machine Learning Models and Evaluation Approach

The evaluated classification models include Random Forest, LightGBM, XGBoost and a Random Forest–ExtraTrees ensemble. The multi-class XGBoost model is evaluated across 721 disease categories. Results from the 40-class classification benchmark are presented separately from those of the 721-class differential diagnosis task because they involve different classification settings and evaluation measures.

C. Evaluation Metrics

The evaluation considers Top-1, Top-3 and Top-5 accuracy to measure the ranking quality of diagnostic predictions. Precision, recall and F1-score assess classification performance, while macro and weighted F1-scores provide information about performance across disease categories with different frequencies. Expected Calibration Error (ECE) is considered for evaluating prediction confidence. Clinical safety assessment includes sensitivity and under-triage rate for emergency detection. Computational efficiency is evaluated using component execution times and complete workflow latency.

V. CLINICAL WORKFLOW AND MULTI-AGENT COORDINATION

MedAgentix AI follows a multi-agent workflow in which different components collaborate to support preliminary clinical diagnosis, emergency detection, risk assessment, and medical information analysis. Instead of depending on a single model for the entire diagnostic process, the system distributes tasks among specialized agents. Each agent focuses on a particular aspect of clinical decision support, while the orchestration mechanism coordinates their activities and combines the relevant outputs.

A. Patient Information and Symptom Analysis

The workflow begins with the collection of patient information, including reported symptoms and available clinical details. The symptom analysis component processes the provided information and identifies relevant clinical features. Bio_ClinicalBERT supports the processing of clinical text, helping convert unstructured medical descriptions into meaningful information for subsequent analysis. The extracted symptoms are standardized according to the system's predefined symptom features.

B. Emergency Detection and Risk Assessment

After processing the patient information, the system evaluates potential emergency indicators and risk-related features. The emergency detection component identifies warning signs that may indicate a critical condition. A rule-based warning

 This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IJARCCE.2026.151171

mechanism and classification-based assessment support the identification of potentially high-risk cases. This stage helps prioritize critical indicators before presenting diagnostic recommendations.

C. Disease Prediction and Differential Diagnosis

The processed clinical features are passed to the diagnostic prediction component. Machine-learning models, including Random Forest, LightGBM, and XGBoost, are used for disease classification and prediction. For differential diagnosis, the system generates a ranked list of possible conditions rather than relying exclusively on a single predicted disease. The XGBoost model supports Top-1, Top-3, and Top-5 prediction outputs, allowing multiple possible diagnoses to be considered.

D. Biomedical Knowledge Retrieval

To provide additional medical context, the system incorporates a Retrieval-Augmented Generation (RAG) approach. Biomedical text is represented using PubMedBERT-based embeddings, which support semantic retrieval of relevant medical information. Retrieved information provides supplementary context for understanding the diagnostic results and their associated medical concepts.

E. Explainability and Report Generation

The explainability component uses SHAP and LIME to interpret model predictions and identify influential clinical features. These explanations help users understand the factors contributing to a prediction. The system combines relevant patient information, diagnostic outputs, risk-related findings, and explanatory information to support the generation of clinical reports.

F. Multi-Agent Coordination

The coordination mechanism manages communication between the specialized components and organizes their outputs into a unified workflow. This allows symptom analysis, emergency detection, disease prediction, biomedical retrieval, and explainability to contribute to the overall decision-support process. The final output is intended to assist healthcare professionals in reviewing possible diagnoses and associated clinical information. MedAgentix AI functions as a preliminary decision-support system and does not replace professional medical judgment.

VI. CLINICAL SAFETY AND PREDICTION RELIABILITY

Clinical safety is an important consideration in MedAgentix AI, particularly when identifying potentially critical patient conditions. The system evaluates emergency warning signs to identify cases that may require urgent attention. Sensitivity and under-triage rate are considered to assess emergency detection performance and the system's ability to avoid assigning low urgency to serious cases.

Prediction reliability is assessed through confidence calibration, which examines whether the confidence assigned to model predictions corresponds to their observed accuracy. Expected Calibration Error (ECE) is considered for evaluating the difference between predicted confidence and actual model performance.

The supervisor agent also examines the confidence and agreement of diagnostic agents before combining their outputs or requesting additional medical information. However, confidence-based predictions do not guarantee diagnostic correctness. Benchmark-based evaluation alone cannot establish clinical safety. Independent validation using properly governed clinical datasets and qualified medical reviewers is necessary before practical clinical deployment.

VII. RESULTS AND DISCUSSION

A. Classification Model Performance

The classification models were evaluated using the 40-class benchmark. The results include training accuracy, test accuracy and F1-score.

TABLE I. COMPARATIVE PERFORMANCE OF CLASSIFICATION MODELS

Model	Training Accuracy	Test Accuracy	F1-Score
Random Forest	99.65%	98.02%	98.00%
LightGBM	99.75%	97.82%	97.79%
XGBoost	99.75%	97.42%	97.40%
Soft-Voting Ensemble	99.75%	97.42%	97.38%

 This work is licensed under a Creative Commons Attribution 4.0 International License

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151171

Random Forest achieved the highest test accuracy of 98.02% and an F1-score of 98.00%. The training accuracies were slightly higher than the test accuracies, indicating a generalization gap.

B. Differential Diagnosis Performance

The 721-class differential diagnosis task was evaluated using Top-1, Top-3 and Top-5 accuracy, along with weighted and macro F1-scores.

TABLE II. PERFORMANCE COMPARISON ON THE 721-CLASS BENCHMARK

Metric	XGBoost	Random Forest–ExtraTrees
Top-1 Accuracy	82.98%	70.05%
Top-3 Accuracy	94.01%	80.03%
Top-5 Accuracy	96.53%	82.79%
Weighted F1-Score	82.65%	71.10%
Macro F1-Score	66.12%	54.30%

XGBoost achieved 82.98% Top-1 accuracy, increasing to 94.01% Top-3 accuracy and 96.53% Top-5 accuracy. Its weighted and macro F1-scores were 82.65% and 66.12%, respectively. XGBoost achieved higher reported values than the Random Forest–ExtraTrees model across all listed evaluation metrics.

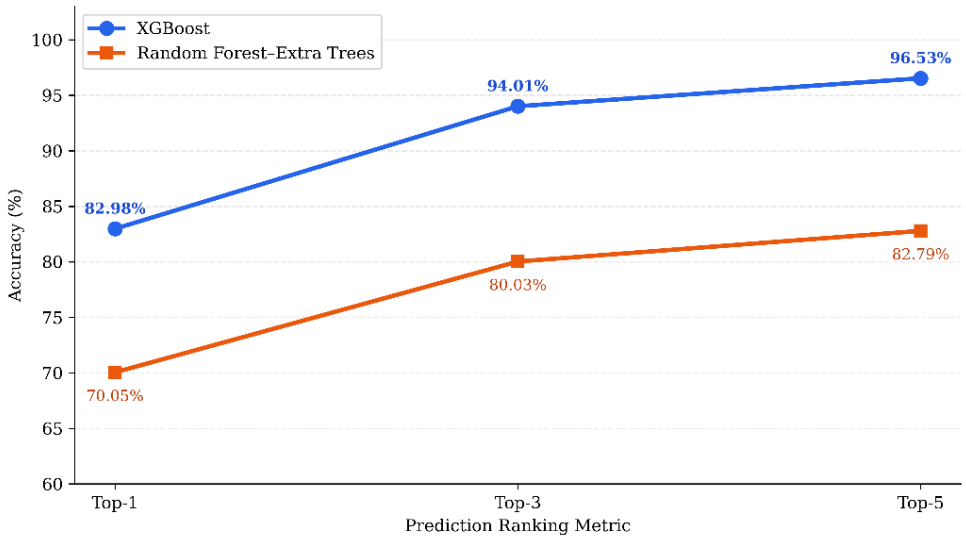


Fig. 1. Top-(k) Accuracy of the Differential Diagnosis Model

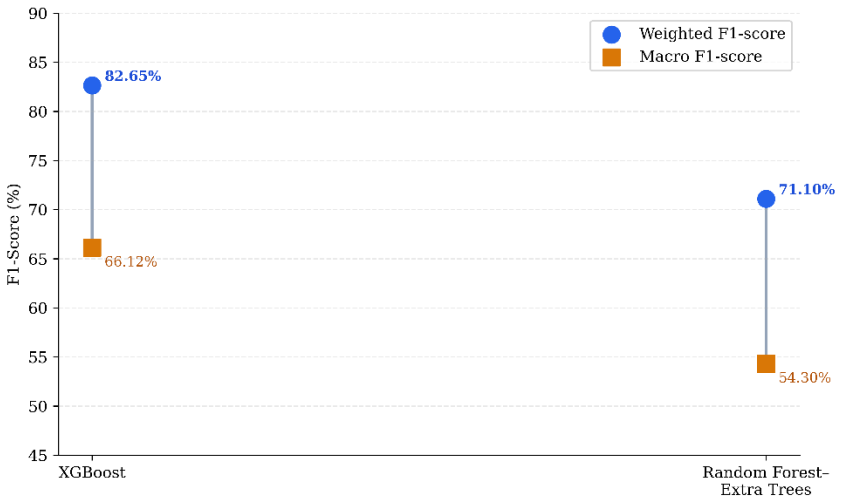


Fig. 2. Comparison of Weighted and Macro F1-Scores for Differential Diagnosis Models.

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

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026

DOI: 10.17148/IJARCCE.2026.151171

C. Multi-Agent Evaluation

The reported multi-agent evaluation includes 96.08% triage accuracy and 90.00% symptom-extraction recall. Critical-vital identification and high-risk case identification each achieved a reported recall of 100%. These measurements provide an initial assessment of the system components but do not independently establish clinical safety.

D. Computational Efficiency

The reported execution times of the major components are presented below.

TABLE III. COMPUTATIONAL EFFICIENCY OF MEDAGENTIX AI

Component	Execution Time
Triage	0.02 ms
Emergency Identification	1.17 ms
Vector Retrieval	42 ms
Diagnosis	150.47 ms
Bio_ClinicalBERT Processing	173.94 ms
Complete Pipeline	2.4 seconds

The complete pipeline was reported to execute in 2.4 seconds. Individual component execution times varied, with Bio_ClinicalBERT processing and diagnosis taking 173.94 ms and 150.47 ms, respectively.

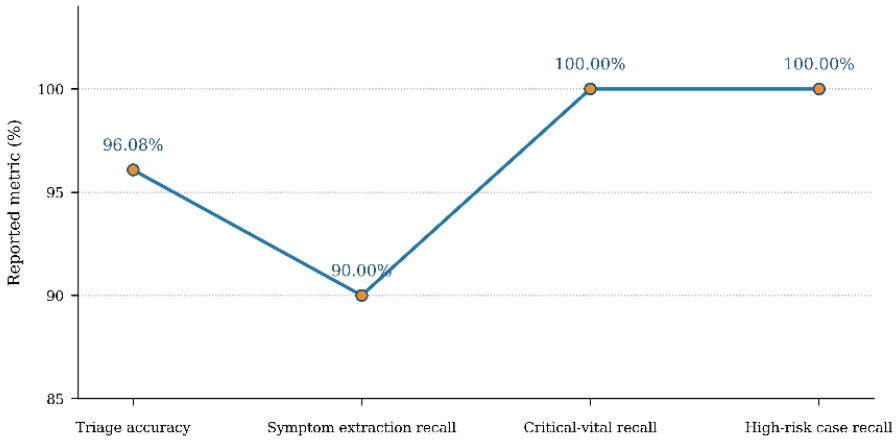


Fig. 4(a). Agent-Level Evaluation Metrics of MedAgentix AI.

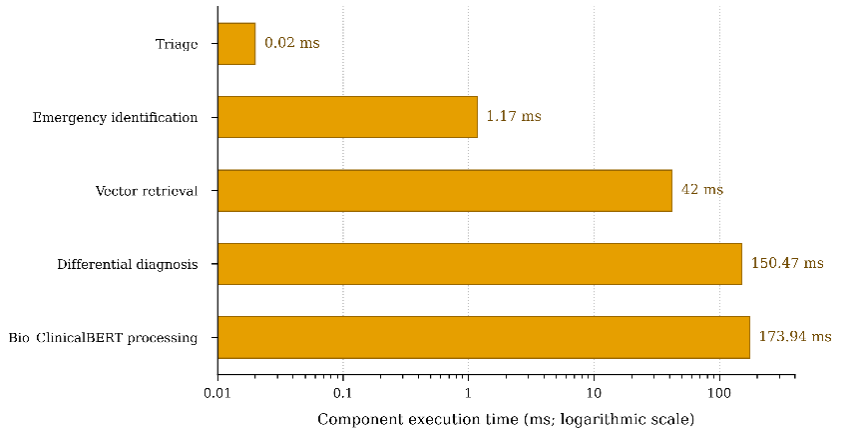


Fig. 4(b). Computational Efficiency of MedAgentix AI

Overall, the results indicate promising benchmark performance and provide an initial assessment of computational efficiency. Further evaluation using independent clinician-reviewed cases is required to establish diagnostic reliability and clinical applicability.



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 11, November 2026
DOI: 10.17148/IJARCCE.2026.151171

VIII. CONCLUSION

MedAgentix AI demonstrates the potential of integrating machine learning, biomedical language processing, retrieval-augmented generation and explainable artificial intelligence within a multi-agent clinical decision support framework. Specialized agents support symptom extraction, diagnostic prediction, emergency triage and risk assessment, while confidence evaluation, SHAP, LIME and biomedical knowledge retrieval contribute to more interpretable outputs.

However, the evaluation is based on open-access benchmark resources and does not include independent clinician-reviewed clinical validation. Model confidence does not guarantee diagnostic correctness, and benchmark performance alone cannot establish clinical safety or generalizability. Further evaluation using independent, clinician-reviewed datasets is necessary to assess reliability and practical applicability. Future work will focus on improving diagnostic consistency, strengthening safety evaluation and validating the system across diverse clinical scenarios.

ACKNOWLEDGMENT

The authors want to say thank you to Prof. Priyanka D. Lanjewar and the faculty of R. C. Patel Institute of Technology for their guidance ongoing support and motivation during this research work. The authors also appreciate the staff, for giving the needed help and resources for building and testing MedAgentix AI.

REFERENCES

[1]. S. A. Thorakkattil, S. B. Sridhar, S. Abdulsalim, M. S. Karattuthodi, P. Chandra, and M. K. Unnikrishnan, "Clinical decision support systems in Indian healthcare settings: Benefits, barriers, and future implications," *Healthcare*, vol. 13, p. 2220, 2025. https://doi.org/10.3390/healthcare13172220

[2]. S. Gupta and S. Mukhopadhyay, "Multi agent system based clinical diagnosis system: An algorithmic approach," *International Journal of Engineering Research and Applications*, vol. 2, no. 5, pp. 1474–1477, 2012.

[3]. S. Chakraborty and S. Gupta, "Medical application using multi agent system—a literature survey," *International Journal of Engineering Research and Applications*, vol. 4, no. 2, pp. 528–546, 2014.

[4]. Y.-J. Chen, C.-S. Chen, and A. Albarqawi, "Reinforcing clinical decision support through multi-agent systems and ethical AI governance," arXiv preprint arXiv:2504.03699, 2025. https://arxiv.org/abs/2504.03699

[5]. Ç. U. Öğdü, K. Arslanoğlu, and M. Karaköse, "An adaptive multi-agent LLM-based clinical decision support system integrating biomedical RAG and web intelligence," *IEEE Access*, vol. 13, pp. 167390–167404, 2025. https://doi.org/10.1109/ACCESS.2025.3613340

[6]. X. Wu, H. Zhang, N. E. Garduno-Rapp, J. F. Rousseau, M. Thakkallapally, Y. Ji, S. Visweswaran, Y. Peng, and Y. Wang, "Orchestrator multi-agent clinical decision support system for secondary headache diagnosis in primary care," *Journal of the American Medical Informatics Association*, vol. 33, no. 10, pp. 1910–1920, 2026. https://doi.org/10.1093/jamia/ocag111

[7]. T. Mugambiwa and B. Ndlovu, "Multi-agent retrieval augmented generation for clinical decision support: A systematic review and integrative conceptual framework," *Journal of Applied Informatics and Computing*, vol. 10, no. 1, pp. 171–183, 2026. https://doi.org/10.30871/jaic.v10i1.11900

 This work is licensed under a Creative Commons Attribution 4.0 International License