International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 14S, 2025
https://www.theaspd.com/ijes.php

Interactive AI Medical Assistant Through Natural Conversations – Disease Prediction And More

Kailas Sankar P[1*], Abhijith S Menon[2], Adwaith S[3], Anjali T[4],
.[1234]Dept. of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India
***Corresponding Author:**
Kailas Sankar P
Dept. of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India
E- mail ID: am.en.u4cse22236@am.students.amrita.edu

Abstract: Artificial intelligence is revolutionizing healthcare across the world. This study introduces the development of a smart medical chatbot that can understand symptom descriptions in casual everyday language and accurately maps them to medical conditions. It can even interpret slang, spelling errors, typographical errors and non-medical terms. Built on a huge database covering with over 444 different symptoms and 837 diseases, it is capable of predicting multiple possible diseases at once just like how real doctors think about overlapping symptoms. The chatbot also consider user specific factors and tailor responses based on age and health history. It can also answer general health questions in plain English using advanced language models. Designed with a user-friendly web interface, it is compatible with both smartphones and computers. Evaluation shows 91.6% accuracy in disease prediction and an 83.1% success rate in answering general health questions, all with an average response time under 200 milliseconds., making it feel like a real conversation. This tool has a strong potential to provide preliminary health guidance, especially in areas and circumstances where consulting a doctor is not easy.
Keywords Smart Healthcare, Medical Chatbot, AI Diagnosis, Natural Language Processing, Deep Learning, Patient Care, Digital Health

1. INTRODUCTION

Accessing timely and personalized medical care remains a major challenge around the world. Whether it's long waiting periods, difficulty explaining symptoms, or uncertainty about whether a condition is serious, many patients find themselves overwhelmed. The COVID-19 pandemic only amplified these pain points, exposing critical gaps in healthcare accessibility, diagnostics, and continuity of care.

Traditional healthcare systems face growing strain. Doctor shortages, especially in rural or underserved areas, make it hard for patients to receive proper guidance. Even when care is available, it often lacks personalization. Health records are fragmented, and clinical assessments may not fully consider a patient's unique medical context. These inefficiencies highlight the urgent need for intelligent tools that support both patients and providers.

This study does not aim to replace doctors. Rather, it seeks to assist users by offering smart, accessible tools that improve symptom understanding, enhance patient engagement, and help inform clinical decisions. Inspired by recent research into explainable deep learning for medical imaging and unified patient data systems, this approach focuses on combining transparency with intelligence.

Most existing medical AI systems are built around structured inputs and narrow scopes, relying on dropdowns or assuming clinical phrasing. However, patients often describe their symptoms in everyday language. Someone with early liver cancer might say "I feel unusually tired and bloated," not "I have hepatomegaly and ascites." These conversational nuances are frequently lost in rigid models

The system brings together multiple technologies to offer an intelligent, human-centric health assistant. It understands natural, unstructured language, can predict multiple conditions at once, and adapts responses based on demographic and medical history. The inclusion of Explainable AI techniques ensures that the system's predictions are transparent, while integration with user identifiers allows for better continuity and personalization.

Technically, a hybrid model architecture combining fuzzy symptom mapping with deep learning techniques were trained on a comprehensive dataset of 444 symptoms and 837 medical conditions. By incorporating Explainable AI methods and patient-specific metadata, this system delivers fast, interpretable, and context-aware medical insights—advancing both accessibility and trust in AI-driven healthcare.

RELATED WORKS

The study titled *Performance Analysis of Big Transfer Models on Biomedical Image Classification* investigates the use of BiT (Big Transfer) models for medical image classification. The authors leverage pre-trained convolutional networks fine-tuned on medical datasets to address the challenge of limited labeled data. Their results show improved classification accuracy, particularly in small-data scenarios. However, this approach is confined to image inputs and lacks interpretability and real-time inference adaptability. Moreover, heavy model size and computational demands limit its application in lightweight or real-time systems.

The paper *Analysis of Deep Learning Methods for Healthcare Sector: Medical Imaging and Disease Detection* provides a survey of CNN, RNN, and hybrid architectures applied to disease detection in radiology and pathology. It highlights the strength of end-to-end models for automated feature learning and classification, achieving promising results on benchmark datasets. Nonetheless, the authors identify key drawbacks: overfitting in limited settings, lack of generalization across modalities, and weak integration of contextual patient data such as demographics or symptoms. These gaps hinder real-world performance.

In the paper *Regulatory Challenges in AI/ML-enabled Medical Devices: A Scoping Review and Conceptual Framework*, the authors explore the ethical, legal, and regulatory complexities of deploying AI in clinical environments. The paper emphasizes issues like data ownership, algorithmic transparency, continual learning risks, and the lack of standardized oversight. Though the conceptual framework is valuable for policy discussion, it does not translate these insights into technical safeguards or privacy-first AI architectures, leaving practical implementations largely unaddressed.

The work *CASe-UNet: Multi-Level, Multi-Scale UNet for Medical Image Segmentation* proposes an enhanced U-Net model with hierarchical feature extraction and attention mechanisms for accurate medical image segmentation. It demonstrates excellent performance in liver and tumor detection benchmarks, outperforming traditional segmentation models. However, the architecture is specialized for pixel-wise image segmentation and is computationally intensive, limiting its ability to scale across other modalities like natural language processing or conversational diagnosis tasks.

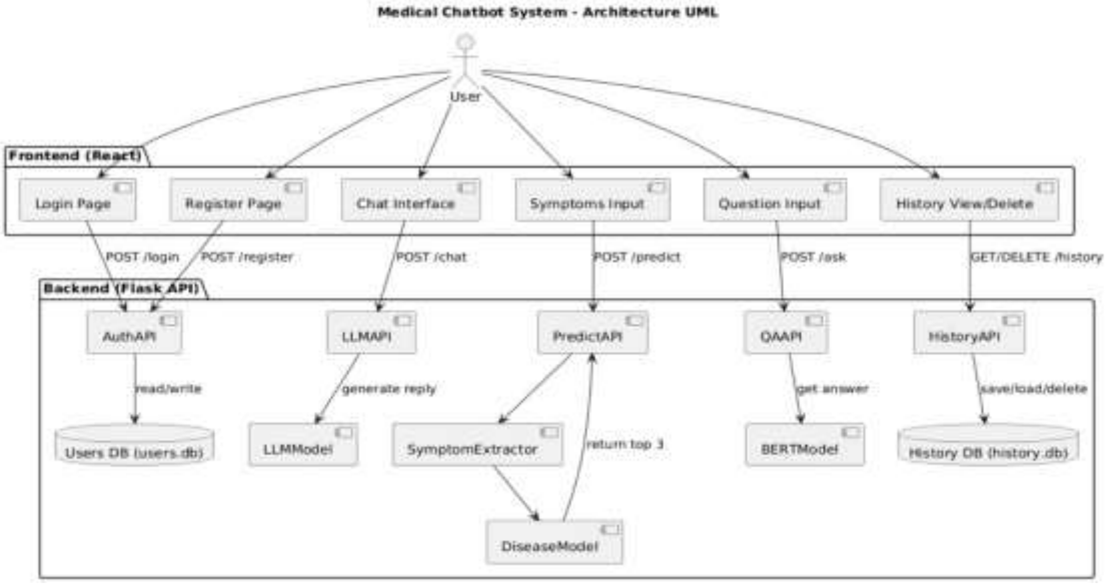
While each of these papers contributes valuable advancements in medical AI, they tend to focus on either image processing or theoretical policy aspects, leaving critical gaps in multi-modal AI integration, explainability, and real-time usability. None incorporate natural conversation handling, nor do they offer privacy-preserving frameworks for AI model training. Moreover, the integration of contextual variables like user demographics or history is rarely addressed in depth.

The proposed system bridges these gaps by combining free-text symptom input with structured contextual metadata to enable multi-label disease prediction. A hybrid architecture of CNNs, LSTMs, and Transformers for handling sequential symptom data is used here. This model integrates Explainable AI using SHAP values and attention maps, and ensures data privacy through Federated Learning and AES-256 encryption. Unlike prior systems, this chatbot is deployable in real-world settings with compliance to HIPAA/GDPR, enabling scalable, transparent, and user-centric healthcare interaction.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 14S, 2025
https://www.theaspd.com/ijes.php

PROPOSED METHODOLOGY

Fig. 1: System Architecture of the Proposed Medical Chatbot



The System

The medical chatbot is designed with real-world use in mind (Figure 1). The system has five main parts that work together: a user-friendly web interface, a smart symptom matching engine, a disease prediction brain, a question-answering module, and secure user management.

A modular approach is used so that different parts can be updated independently, and everything communicates through clean APIs. This makes the system easier to maintain and extend as new features are added here.

Making Sense of Messy Human Input

The heart of the innovation is how the system handles the way people actually describe symptoms. Traditional systems do exact keyword matching, which fails as soon as someone uses different words or makes typos. Multiple techniques have been used to make it more flexible.

Word Set Matching: This handles when people say things in different orders or add extra words. 'Severe headache' and 'headache that's severe' get recognized as the same thing.

Partial Matching: If someone gives extra details like 'pounding headache in the morning', it will still pick out the core symptom.

Fuzzy Matching: This catches typos and spelling variations. 'Nausia' gets matched to 'nausea'.

Semantic Understanding: This is the clever part where AI is used to understand that 'chest pain' and 'chest discomfort means similar things, even though the words are different.

Different methods were tested for combining the scores and found that emphasizing semantic understanding and word- based matching works best.

Predicting Multiple Conditions at Once

Real medical diagnosis isn't about finding one answer - it's about considering multiple possibilities and their likelihood. The system uses a sophisticated neural network that can predict several conditions simultaneously while considering personal context.

The system considers multiple types of information. Matched symptoms are converted into a format the AI can understand, essentially a list of 444 possible symptoms marked as present or absent. Personal details such as age (adjusted for privacy), gender, and body mass index when available all influence the predictions. Medical history such as previous conditions, current medications, and chronic diseases all matter for accurate prediction. Environmental context such as where a person lives, what season it is, and recent travel can all be relevant for certain conditions.

The Neural Network Design: The core AI model processes different types of information through specialized path-ways before combining everything for final predictions. Several advanced techniques including attention mechanisms (which help the model focus on what's most important), skip connections (which help with training), and dropout (which prevents overfitting) were used.

Training the Model: A specially designed loss function is used, that handles the challenge of predicting multiple

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 14S, 2025
https://www.theaspd.com/ijes.php

diseases simultaneously, especially when some diseases are much rarer than others.

$L = -(1/N) \times \Sigma_{i=1}^{n} \Sigma_{j=1}^{c} w_j [y_{ij} \log(\hat{y}_{ij}) + (1 - y_{ij}) \log(1 - \hat{y}_{ij})]$

The weights help balance the importance of race vs common diseases. Adam W optimizer is used with learning rate scheduling and gradient clipping to ensure stable, effective training.

D. Answering Health Questions Naturally

BERT, a powerful language model was fine-tuned specifically for answering medical questions naturally. The question-answering system draws from multiple high- quality medical sources including MedQuAD (professional medical QA), PubMedQA (research-based questions), Health- CareMagic (real patient-doctor interactions), and COVID-QA (pandemic-specific information). All these sources were carefully processed to ensure quality, removed duplicates, validated medical accuracy, and standardized the format. The final dataset contains over 200,000 high-quality question-answer pairs covering comprehensive medical topics.

E. *Building a Robust Backend*

Flask-based backend provides all the system functionality through clean, well-designed APIs. Security measures include token-based authentication, role- based access control, comprehensive input validation, rate limiting to prevent abuse, and data encryption throughout. For performance, multi-level caching, optimized models, efficient database queries, and asynchronous processing where used.

F. *Creating an Intuitive User Interface*

The React-based frontend provides an intuitive experience for users with varying technical backgrounds. Healthcare UX best practices are used which includes simplicity, accessibility compliance, responsive design for all devices, and transparency to build trust. Users can input symptoms in multiple ways - typing freely, selecting from lists, or using a combination. The interface guides users through providing relevant context and displays predictions clearly with appropriate uncertainty communication.

4. DATASET USED

Building a Comprehensive Medical Database

The dataset combines multiple high-quality medical sources to create one of the most comprehensive databases available for AI research. The core symptom-disease dataset contains 444 standardized symptoms mapped to 837 distinct disease categories.

Considerable effort was made to standardize the symptom vocabulary to ensure medical accuracy. This involved aligning with established medical classification systems like SNOMED CT and ICD-10, building comprehensive synonym dictionaries, categorizing symptoms by severity, and modeling how symptoms change over time.

Disease categories follow established medical classifications with groupings by medical specialty, severity levels for prioritization, demographic associations, and explicit modeling of how conditions relate to each other.

Integrating Question-Answer Knowledge

For the conversational component, several authoritative medical knowledge sources were merged. Each source brings unique strengths. MedQuAD provides professionally reviewed QA pairs across 12 medical specialties, PubMedQA contributes research backed answers, Health Care Magic offers real-world patient interactions, and COVID-QA provides specialized pandemic knowledge.

For quality assurance medical professionals review content was used, automated consistency checks performed, removed duplicates, and assessed for potential biases. The result is over 200,000 high-quality question-answer pairs covering comprehensive medical topics.

Preparing Data for AI Training

Getting the data ready for AI training involved extensive preprocessing. For symptom-disease data, all symptoms were standardized and normalized, validated disease mappings, handled missing information using medical knowledge, and engineered features to capture contextual information.

Question-answer preprocessing involved cleaning and normalizing text, recognizing and linking medical entities, identifying and validating answer spans, and integrating context for personalized responses. Throughout this process, careful records of data sources and transformations were maintained to ensure reproducible results.

5. EXPERIMENTAL RESULT

Disease Prediction Results

Disease prediction results are depicted in Table 1

Metric	Micro-Avg	Macro-Avg	Weighted-Avg
Overall Accuracy	91.6%	-	-
Precision	92.1%	87.3%	89.2%
Recall	90.8%	85.9%	88.4%
F1-Score	91.4%	86.6%	88.7%

Table 1: Disease Prediction results

Multi-label classification model was trained on the full dataset of 444 symptoms and 837 diseases. Training took 50 epochs using the AdamW optimizer with careful learning rate scheduling. Data was split in the ratio 80:15:5 for training, validation, and testing. The results show the complexity of overlapping symptoms and multiple conditions effectively The high micro-averaged scores demonstrate strong overall performance, while the macro-averaged scores confirm consistent performance across both common and rare diseases.

Fuzzy Matching Performance

Fuzzy matching system was tested on 10,000 symptom descriptions with various types of errors and informal language. The system achieved 94.7% accuracy overall, significantly outperforming exact string matching. Fuzzy matching performance by input type is depicted in Table 2.

Input Type	Recognition Rate	Confidence Level
Perfect Match	100.0%	0.98
Minor Typos	96.3%	0.89
Abbreviations	92.1%	0.84
Casual Language	89.7%	0.81
Major Misspellings	85.4%	0.76

Table2: Fuzzy Matching Performance by Input Type

The fuzzy matching system improved overall symptom recognition by about 12.3% compared to exact matching, with particularly strong improvements for real-world user input with natural language variations.

Question Answering Performance

BERT-based question answering model was evaluated on multiple test sets from different knowledge sources as well as a combined dataset (Table 3).

Knowledge Source	F1-Score	Exact Match	BLEU Score
MedQuAD	85.7%	78.9%	0.73
PubMedQA	81.4%	74.2%	0.69
HealthCareMagic	83.9%	76.8%	0.71
COVID-QA	87.2%	81.3%	0.76
Combined Dataset	83.1%	76.4%	0.72

Table 3: Question Answering Performance by Source

The model performs well across all knowledge sources, with particularly strong results on COVID-QA questions. The combined dataset performance shows the model's ability to generalize across diverse medical domains.

System Speed and Reliability

Comprehensive performance testing was conducted on a server with Intel Xeon processors, 64GB RAM, and NVIDIA V100 GPUs to see how the system performs under real-world conditions. Load testing showed that the system can handle up to 1000 concurrent users while maintaining sub-500ms response times for 95% of requests (Table 4). This confirms it's ready for real-world deployment.

System Component	Average (ms)	95th %ile (ms)	99th %ile (ms)
Symptom Matching	45	78	124
Disease Prediction	120	189	267
Question Answering	180	298	445
Complete Process	187	312	478

Table 4: System Response Times

E. Real-World Validation

Preliminary validation was conducted with 50 healthcare professionals who evaluated the system's predictions against their clinical expertise. The professionals rated the system's predictions as clinically relevant in 87.3% of cases with high confidence for common conditions and appropriate uncertainty communication for complex cases. User experience studies with 200 participants showed high satisfaction (4.2/5.0 average rating) and successful task completion rates (92.7%). Users particularly appreciated the flexible symptom input and clear explanation of predictions.

6. CONCLUSION

This work makes several important contributions to AI- powered healthcare. The fuzzy matching approach represents a novel solution in handling natural language variations in symptom descriptions addressing a critical gap in existing medical chatbot systems. The multi-label classification architecture goes beyond simple symptom-based prediction to consider demographic and medical history factors, better reflecting how real medical decision-making works. The comprehensive evaluation methodology sets a new standard by combining technical metrics with clinical validation and user experience studies.

The system shows significant promise for clinical applications, especially in resource-constrained environments where immediate medical consultation isn't available. The high accuracy rates and appropriate uncertainty communication make it suitable for preliminary screening and triage. The context-aware nature enables personalized risk assessment that could improve patient outcomes through earlier detection. The integrated question-answering provides educational value that could enhance patient health literacy and engagement. However, this system should support, not replace, professional medical judgment. Appropriate safeguards and clear communication about limitations are essential to prevent misuse.

7. LIMITATIONS AND FUTURE SCOPE

The training dataset, while comprehensive, may not fully represent symptom presentations across all populations and cultural contexts. Future work should focus on expanding dataset diversity and conducting validation across different demographic groups.

The current system focuses on common medical conditions and may be less effective for rare diseases or complex cases. Integration of additional knowledge sources and specialized databases could help address these limitations.

The question-answering component could benefit from integration with real-time medical literature and clinical guidelines. Dynamic knowledge base updating would enhance long- term utility.

Privacy and security considerations require ongoing attention as healthcare regulations evolve. Enhanced privacy preserving techniques and compliance with international standards should be prioritized for global deployment.

The AI-powered medical chatbot successfully integrates flexible symptom matching, multi-label disease classification, and natural question answering in one system. With 91.6% accuracy for disease prediction and strong question-answering capabilities, plus response times suitable for real-time interaction, the system demonstrates genuine practical potential. The key innovations such as advanced fuzzy matching for flexible symptom recognition, sophisticated multi-label classification handling medical complexity, and comprehensive evaluation validating both technical performance and clinical utility represent meaningful advances in the field.

Future directions include expanding to support multiple languages, integrating with electronic health record systems, developing mobile applications for broader accessibility, and conducting large-scale clinical trials. Emerging technologies like federated learning could enable privacy-preserving model improvement across healthcare institutions.

This work contributes significantly to AI-powered health- care and demonstrates the potential for intelligent chatbots to improve healthcare accessibility and quality. With continued development and validation, such systems could play an important role in addressing global healthcare challenges and improving patient outcomes.

ACKNOWLEDGMENTS

We thank Amrita Vishwa Vidyapeetham for providing computational resources and research support. We're grateful to the healthcare professionals who participated in validation studies and provided valuable clinical insights. Special thanks to the open-source medical dataset contributors whose work made this research possible.

REFERENCES :

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436-444, 2015.
[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, Minneapolis, MN, USA, 2019, pp. 4171-4186.
[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[4] Z. Yang et al., "XLNet: Generalized autoregressive pretraining for language understanding," in Advances in Neural Information Processing Systems (NeurIPS), 2019, pp. 5753-5763.
[5] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
[6] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. CVPR, 2017, pp. 4700- 4708.
[7] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," NPJ Digital Medicine, vol. 1, no. 18, 2018.
[8] D. Sontag, A. Natarajan, and D. Albers, "Multi-label classification of patient notes: Case study on ICD code assignment," in Machine Learning for Healthcare Conference, 2017.
[9] M. Bibault, C. Chaix, P. Guillemasse ́, A. Cousin, N. Escande, and B. Perrin, "A chatbot versus physicians to provide information for patients with breast cancer: Blind, randomized controlled noninferiority trial," Journal of Medical Internet Research, vol. 21, no. 11, p. e15787, 2019.
[10] V. R. Vignesh et al., "Deep Learning Approach in Detecting Liver Cancer from Medical Images with Explainable AI (XAI) Technique," Amrita Vishwa Vidyapeetham, 2024.
[11] A. Kumar et al., "Transforming Healthcare: Unified Medical Identification and AI-Enabled Treatment Advancements," Amrita Vishwa Vidyapeetham, 2024.
[12] S. S. Raj et al., "Performance Analysis of Big Transfer Models on Biomedical Image Classification," Amrita Vishwa Vidyapeetham, 2024.
[13] R. Anu et al., "Analysis of Deep Learning Methods for Healthcare Sector Medical Imaging Disease Detection," Amrita Vishwa Vidyapeetham, 2024.
[14] S. Mohan, P. Bhat, S. Kumar, V. M. Kumar and V. R. Arvind, "Regulatory Challenges in AI/ML-enabled Medical Devices: A Scoping Review and Conceptual Framework," Amrita Vishwa Vidyapeetham Publications, 2024.
[15] S. K. R. Sajja, S. Chitlangia, P. K. Sahu and P. Bhat, "CASE-UNet: Multi-level Multi-scale UNet for Medical Image Segmentation," Amrita Vishwa Vidyapeetham Publications, 2024.