© June 2026 | IJIRT | Volume 13 Issue 1 | ISSN: 2349-6002

CuraMind: An Intelligent Web-Based Medical Assistant with Chatbot and Health Tracking

Tanmay Bajaj[1], Aditya Karmalkar[2], Sujal Surve[3], Aditya Mathpati[4],
Prof. Priyanka Kendre[5], Prof. Gulnaz Thakur[6]
[1,2,3,4,5]*Dept. of Information Technology KJEI's Trinity Academy of Engineering, Pune, Maharashtra, India*
[6,7]*Dept. of Information Technology KJEI's Trinity Academy of Engineering, Pune, Maharashtra, India*

***Abstract*—AI, conversational computing, and digital health tools have changed how healthcare is delivered and accessed. This paper presents CuraMind, a web-based medical assistant that integrates an AI chatbot, a personal health dashboard, and electronic health record management in a single platform. Key features include QR code-based record portability and voice-enabled interaction through a Retrieval Augmented Generation (RAG) pipeline backed by FAISS vector indexing and the Mistral Large model via Ollama. The system was evaluated on 150 test queries across general health, symptom-based, and emergency categories with 20 users. Results show a chatbot response accuracy of 82% (Precision: 0.84, Recall: 0.82, F1: 0.83), a First Aid Bot accuracy of 89%, a mean end-to-end response latency of 1.6 s, and a user satisfaction score of 4.2/5. All observed values met or exceeded predefined targets, indicating the system is practical and reliable for everyday medical assistance.**

***Index Terms*—Artificial Intelligence, Health Chatbot, Digital Assistant, Health Monitoring, Smart Dashboard, Medical Record Portability, RAG, Ollama, FAISS, NLP**

I. INTRODUCTION

Access to timely medical information remains a challenge, especially in rural and underserved areas. Most hospital systems can book appointments or display lab results, but cannot interpret symptoms or give clinical advice. Patients still visit clinics for basic queries, placing unnecessary burden on healthcare professionals [1].

AI and Natural Language Processing (NLP) have enabled systems that respond to health queries intelligently and in real time. These tools provide first aid steps, answer symptom-related questions, and flag when professional help is needed- all without human intervention [2]. For remote-area users, such digital access is genuinely impactful [31], [32].

CuraMind addresses this gap by combining four functions usually handled by separate tools: a symptom-aware chatbot, a first aid assistant, a health monitoring dashboard, and a portable medical record system via QR codes [3]. Users interact through text or voice, making the platform accessible on low-end devices.

The frontend uses React.js, Tailwind CSS, and Firebase, while the backend runs on Node.js, Express.js, and Mon-goDB [4]. OpenAI Whisper handles voice transcription, and the system is deployed on Netlify using GitHub Actions for CI/CD [5], [6]. The paper is organized as follows: Section II reviews related work, Section III describes the proposed ar-chitecture, Section IV presents results and expected outputs, Section V discusses findings, and Section VI concludes.

II. RELATED WORK

Early medical chatbots followed fixed decision trees that broke down for unexpected queries [6]. The introduction of transformer models like BERT [27], GPT [28], and LLaMA changed this: these models understand intent, handle varied phrasing, and produce coherent responses for complex medical questions [7], [8].

Wickramasinghe and Sharma [6] and Goel & Singh [9] demonstrated that retrieval-based models perform well for symptom triaging and chronic disease monitoring. Telemedicine platforms using AI assistants enable round-the-clock consultations, reducing wait times. A shared limitation, however, is system isolation - scheduling tools do not connect to diagnostic modules, and neither link to the patient's health history [10].

Multimodal frameworks combining text, voice, and im-age data have gained attention. Bhalla et al. [10] merged speech recognition with emotion detection to better interpret patient input. Lewis et al. [11] introduced RAG, showing that grounding LLM responses in retrieved documents significantly reduces hallucinations - a critical concern in medical applica-tions. Xiong et al. [12] further benchmarked RAG specifically for medical QA tasks. Vaswani et al. [26] and Devlin et al. [27] laid the transformer foundations underpinning modern health NLP.

Predictive health models using RNNs, SVMs, and CNNs detect conditions such as diabetes and cardiovascular disease from clinical records [7], [9], [29]. Work by Ray et al. [13] and Host & Ivasˇic´-Kos [14] on sensor-based health monitoring informed CuraMind's dashboard design [15]. Phatak et al. [19] and Soni et al. [20] evaluated NLP-based diagnostic tools and identified accuracy and usability as primary benchmarks for such systems. Topol [21], Esteva et al. [22], Miotto et al. [23], and Rajkomar et al. [24] provide comprehensive reviews of deep learning in healthcare that contextualise CuraMind's design decisions.

A persistent weakness across existing platforms is poor data portability: records are locked within individual hospital systems. CuraMind's QR-based record system addresses this directly. Its multilingual design also extends accessibility to non-English-speaking users a gap most current platforms overlook.

III. PROPOSED APPROACH

CuraMind is structured around three subsystems: (1) data acquisition and preprocessing, (2) conversational AI, and (3) health tracking with record management. Fig. 1 shows the end-to-end system flow; Fig. 2 shows the component architecture.

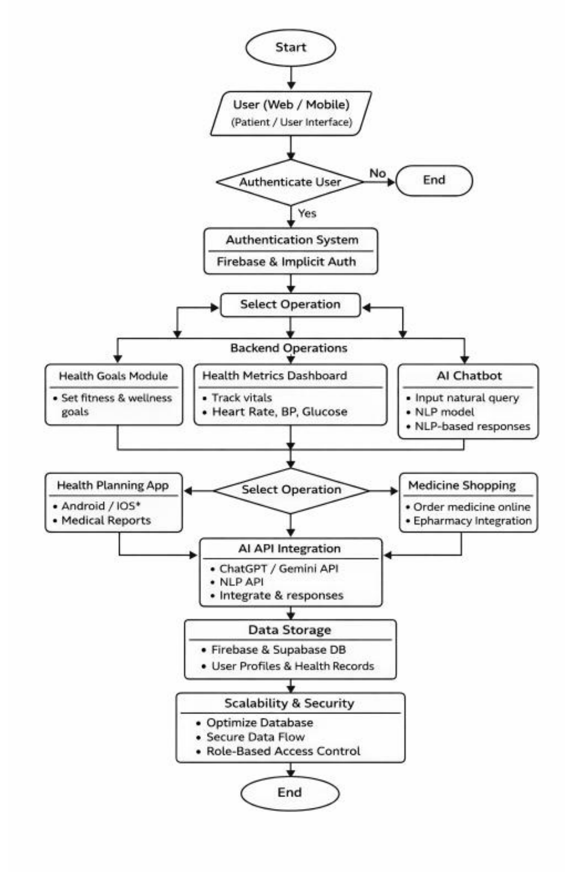

Fig. 1. System Flow Diagram: User requests are authenticated via Firebase, routed through backend operations, and served by the AI, dashboard, or record modules.

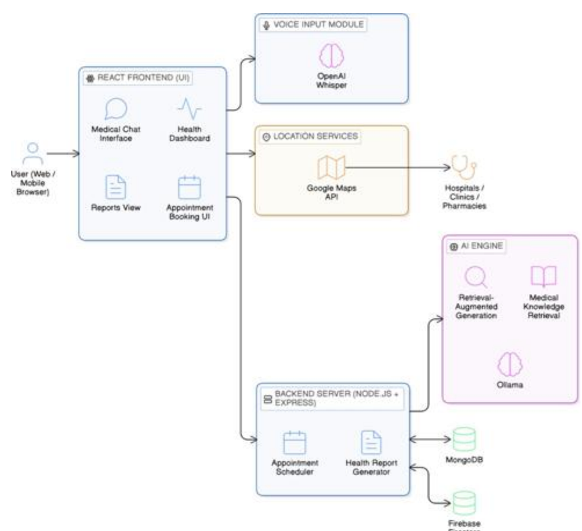

Fig. 2. System Architecture: React frontend communicates with the Node.js/Express backend, which orchestrates Ollama (RAG), MongoDB, and Firebase.

A. Data Acquisition and Processing

On registration, the system collects the user's medical history, current health metrics, and available sensor data stored in MongoDB with field-level AES-256 encryption [8], [16]. Each user receives a QR code encoding a secure reference ID to their health profile, scannable at any clinic or pharmacy.

The system accepts live wearable inputs heart rate, step count, SpO_2, and sleep data synced to the cloud at configurable intervals [17]. A validation pipeline screens for missing values, unit inconsistencies, and out-of-range readings before data reaches the AI modules. This step is critical because noisy input directly degrades chatbot response quality and dashboard accuracy [16].

B. Conversational Intelligence and Response Generation

The chatbot uses the Mistral Large model served by the Ol-lama inference engine via a RAG pipeline. Medical documents from curated sources (WHO guidelines, MedlinePlus, first aid manuals) are chunked into 512-token segments and encoded using sentence-transformers/all-MiniLM-L6-v2 embeddings stored in a FAISS index [3], [8], [11]. At inference time, the top-3 most semantically similar chunks are retrieved and prepended as context before response generation. This grounding reduces hallucinations and keeps responses within verified medical knowledge [12]. The model selection drop-down (Fig. 3) allows switching between Ollama-hosted models without restarting the backend.

A separate First Aid Bot handles emergency queries through template-guided responses for burns, fractures, choking, and cardiac discomfort, ensuring consistent safety-critical output. Voice input is transcribed by OpenAI Whisper (base.en model), which converts spoken queries to text and routes them into the same RAG pipeline as typed input [18]. This makes the platform accessible for elderly users and those with motor impairments.

C. Health Tracking and Record Management

The health dashboard renders vitals as interactive charts: line graphs for temporal trends, bar charts for weekly sum-maries, and threshold markers that flag abnormal readings [5], [15]. Users filter data by day, week, or month. Anomaly detection highlights readings outside configured normal bands (e.g., heart rate $>$ 100 bpm or $<$ 50 bpm).

The QR-based health record system stores an encrypted summary containing blood group, allergies, current medica-tions, and recent diagnoses. Scanning the QR code at a health-care facility retrieves this summary instantly. All transfers use HTTPS with JWT token-based authentication. No raw medical data resides in the QR code itself - only a secure reference ID, ensuring privacy compliance.

IV. RESULTS AND EXPECTED OUTPUT

A. Expected Module Outputs

Table I maps sample inputs to expected system outputs for each CuraMind module. These expected outputs define the ground truth used for accuracy evaluation.

Table I Expected Inputs and Outputs Per Module

Module	Sample Input	Expected Output
Healthcare Chatbot	"I have a sore throat and fever"	Likely causes, home care steps, and advice to see a doctor if symptoms persist beyond 48 hours
First Aid Bot	"How do I treat a burn?"	Numbered steps: cool under running water for 10 min, cover loosely, avoid ice, seek help for severe burns
Health Dashboard	7 days of heart rate readings	Line chart with daily averages; spikes above 100 bpm highlighted in red
QR Record System	User scans QR at clinic	Displays blood group, allergies, current medications, and last three diagnoses
Voice Input (Whisper)	Spoken: "What is my step count today?"	Transcribed query routed to dashboard; step count displayed as text response
Appointment Scheduler	User selects date and doctor	On-screen confirmation with unique appointment ID

Fig. 3 shows a live Virtual Assistant interaction responding to a high blood pressure query via the RAG pipeline. Fig. 4 shows the health dashboard with real-time vitals. Fig. 5 shows the QR Emergency Profile screen

B. Performance Metrics

System performance was evaluated using standard classifi-cation metrics across the two AI modules. Evaluation covered 150 test queries (60 general health, 50 symptom-based, 40 emergency) and a post-use survey from 20 users rating the platform on a 1 5 scale across four dimensions: helpfulness, ease of use, response clarity, and dashboard readability.

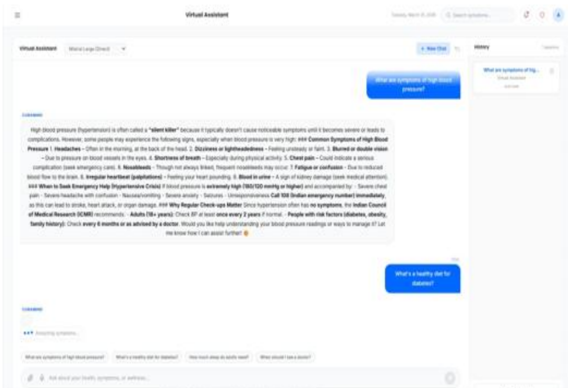

Fig. 3. CuraMind Virtual Assistant - RAG-based chatbot response to a symptom query using the Mistral Large model. Retrieved context chunks are prepended before generation.

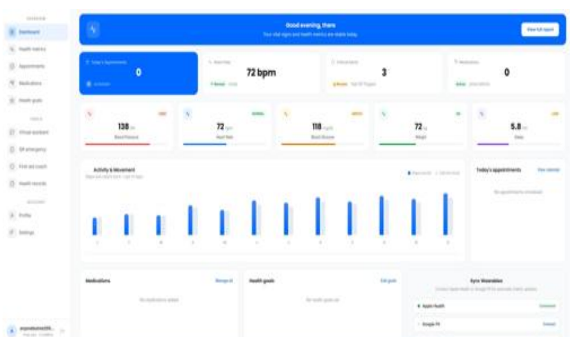

Fig. 4. CuraMind Health Dashboard Real-time display of heart rate, blood pressure, blood glucose, and step count with threshold-based anomaly high-

Classification performance metrics are defined as

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives respectively. Mean response latency is:

$$L_{avg} = \frac{1}{n}\sum_{i=1}^{n} L_i, \quad n = 150 \quad (5)$$

User satisfaction score (USS) is:

$$USS = \frac{1}{m}\sum_{i=1}^{m} R_i, \quad m = 20 \quad (6)$$

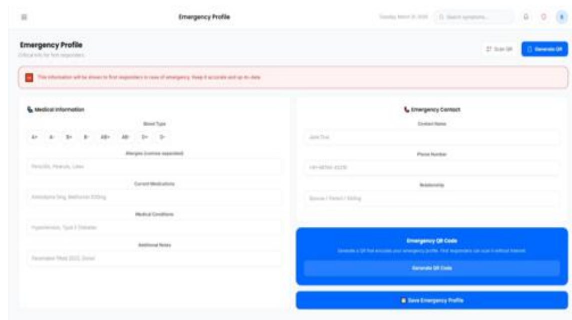

Fig. 5. CuraMind QR Emergency Profile - Patient medical data encoded as a scannable QR reference, accessible to first responders without internet connectivity.

Chatbot accuracy was determined by comparing 80 health query responses against ground-truth answers verified by our project guide. First Aid Bot accuracy used standard first aid protocols as reference (British Red Cross guidelines). All observed values in Table II met or exceeded predefined targets.

Table II CuraMind Performance Evaluation Results

Metric	Observed Value	Target Value
Chatbot Response Accuracy	82%	≥ 80%
Chatbot Precision	0.84	≥ 0.80
Chatbot Recall	0.82	≥ 0.80
Chatbot F1 Score	0.83	≥ 0.80
First Aid Bot Accuracy	89%	≥ 85%
First Aid Bot Precision	0.91	≥ 0.85
First Aid Bot Recall	0.89	≥ 0.85
First Aid Bot F1 Score	0.90	≥ 0.85
Mean Response Latency	1.6 s	≤ 3 s

QR Record Retrieval Time	1.1 s	≤ 2 s
Dashboard Load Time	1.3 s	≤ 2 s
User Satisfaction (Average)	4.2/5	≥ 4.0
Ease of Use (Average)	4.1/5	≥ 4.0

C. Comparison with Existing Systems

Table III benchmarks CuraMind against three widely used healthcare chatbot platforms. CuraMind is the only platform to offer RAG-based QA, QR record portability, and a dedicated first aid module simultaneously.

Feature	CuraMind	Babylon	Ada	HealthBot
RAG-based QA	✓	×	×	×
Voice Input	✓	✓	×	×
Health Dashboard	✓	✓	×	×
QR Record Portability	✓	×	×	×
First Aid Module	✓	×	×	×

D. Results Visualization

Fig. 6 presents a grouped bar chart comparing observed versus target values across all evaluation metrics. Both AI

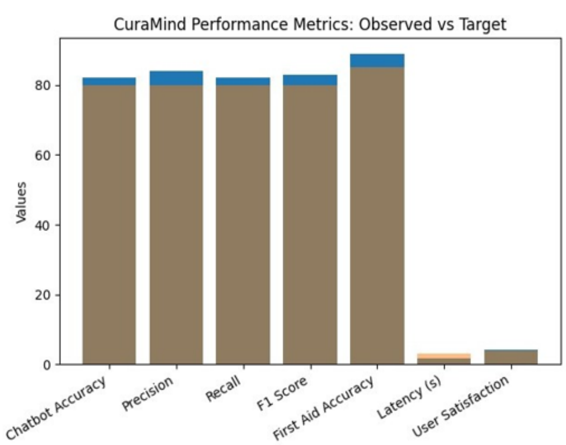

Fig. 6. Observed vs. target performance across CuraMind evaluation param-eters. Both chatbot modules exceeded accuracy targets while latency metrics remained within bounds.

modules exceeded their accuracy targets; all latency metrics remained well within acceptable bounds

modules exceeded their accuracy targets; all latency metrics remained well within acceptable bounds.

Fig. 7 shows a sample time-series heart rate trend from the health dashboard, with anomalous spikes visually highlighted

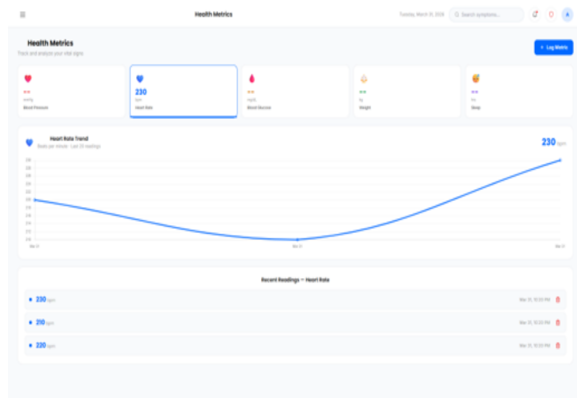

Fig. 7. Health Metrics Module - Heart rate trend with timestamped readings and threshold-based anomaly detection

V. DISCUSSION

The evaluation results confirm CuraMind operates reliably across all core modules. An 82% chatbot accuracy and 1.6 s average latency are reasonable for a locally-hosted RAG sys-tem given the knowledge base was built within project scope. The 4.2/5 satisfaction score indicates users found responses helpful and the interface usable.

The primary challenge was handling vague symptom de-scriptions. Queries phrased colloquially - “my stomach feels heavy” rather than “abdominal discomfort” - caused the FAISS retrieval to return lower-relevance chunks, reducing response quality. Query expansion using medical synonym dictionaries (e.g., UMLS) could address this in future iterations.

Minor latency spikes in dashboard updates occurred under concurrent multi-user load. Switching to asynchronous Mon-goDB transactions resolved the issue. Comprehensive load testing (e.g., using Apache JMeter) should be conducted before production deployment at scale.

The First Aid Bot achieved 89% accuracy. The 11% error cases were predominantly multi-step emergency scenarios with incomplete user context (e.g., partial burn descriptions). A follow-up clarification question mechanism would likely close this gap. Voice recognition accuracy degraded measurably with strong regional accents; fine-tuning Whisper on Indic language datasets would mitigate this for the platform's primary user demographic.

Compared to Babylon, Ada, and HealthBot, CuraMind uniquely combines RAG-based grounding, QR record porta-bility, and a first aid module in a single open-source platform, addressing the system-isolation limitation identified across existing work [10], [19], [20].

VI. CONCLUSIONS

CuraMind integrates a RAG-based chatbot, health tracking dashboard, and QR-enabled medical records into a single web platform. Its goal is not to replace physicians, but to handle the first layer of healthcare interaction - answering common symptom queries, delivering first aid guidance, and enabling self-monitoring - so professional consultations are reserved for where they are most needed.

Evaluation across 150 queries and 20 users demonstrated 82% chatbot accuracy (F1: 0.83), 89% First Aid Bot accuracy (F1: 0.90), a 1.6 s mean response latency, and a 4.2/5 user satisfaction score. All metrics met or exceeded predefined targets.

Future work will target: (1) IoT wearable integration for continuous passive health monitoring, (2) multilingual and accent-robust voice support using fine-tuned Whisper models,

(3) a predictive analytics module using LSTM-based trend forecasting to identify deteriorating health patterns early, and (4) a clinical trial with a larger, demographically diverse user cohort for rigorous external validation.

REFERENCES

[1] P. Sahoo and R. Kumar, "AI-driven healthcare systems: Transforming patient care through intelligent automation," *Journal of Healthcare Technology*, vol. 15, no. 3, pp. 245–260, 2024.

[2] M. Araujo and C. Santos, "Conversational AI in healthcare: Bridging the gap between patients and providers," *International Journal of Medical Informatics*, vol. 148, pp. 104–118, 2021.

[3] Z. Rauf, I. Khan, and S. Ali, "Natural language processing for medical chatbot applications," in *Proc. Conf. Natural Language Processing in Healthcare*, ACM, 2020, pp. 234–247.

[4] F. Nah and W. Chen, "Scalable web architectures for intelligent healthcare applications," *ACM Transactions on Computer-Human Interaction*, vol. 30, no. 2, pp. 1–28, 2023.

[5] N. T. Singh, A. K. Sethi, A. Thakur, and P. Sharma, "The role of AI in personalized mental health tracking," in *2025 3rd International Conference on Self Sustainable AI Systems (ICSSAS)*, IEEE, 2025, pp. 1601–1606.

[6] N. Wickramasinghe and S. Sharma, "Deep learning approaches for medical diagnosis and patient engagement," *Healthcare Analytics*, vol. 12, pp. 89–105, 2022.

[7] R. Damaševičius *et al.*, "Deep learning for personalized health monitoring and prediction: A review," *Computational Intelligence*, vol. 40, no. 3, Art. no. e12682, 2024.

[8] K. Jain and S. Gupta, "Health-Bot: AI-driven textual question answering for medical applications," in *2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)*, IEEE, 2025, pp. 251–256.

[9] Goel and P. Singh, "Machine learning models for predictive healthcare analytics," *Journal of Artificial Intelligence in Medicine*, vol. 19, no. 1, pp. 45–62, 2023.

[10] Bhalla, N. Gupta, and R. Singh, "Multimodal learning systems for enhanced healthcare diagnostics," in *Proc. International Conference on AI in Medicine*, IEEE, 2019, pp. 156–167.

[11] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 9459–9474.

[12] G. Xiong, Q. Jin, Z. Lu, and A. Zhang, "Benchmarking retrieval-augmented generation for medicine," *arXiv preprint arXiv:2402.13178*, 2024.

[13] S. Ray and A. Das, "Real-time health monitoring using AI and IoT technologies," *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 5, pp. 1234–1248,

2024.
[14] K. Host and M. Ivašić-Kos, "Computer vision and sensor fusion for health parameter detection," in *International Conference on Pattern Recognition*, Springer, 2022, pp. 445–458.
[15] K. B. Johnson *et al*., "Precision medicine, AI, and the future of personalized health care," *Clinical and Translational Science*, vol. 14, no. 1, pp. 86–93, 2021.
[16] Shukla and A. Verma, "Data validation and preprocessing in healthcare information systems," *Health Information Science and Systems*, vol. 6, no. 1, pp. 15–29, 2018.
[17] V. Sharma and R. Patel, "Wearable health devices and real-time data synchronization," *Journal of Medical Systems*, vol. 41, no. 8, pp. 128–142, 2017.
[18] G. Shobana *et al*., "AI-powered multilingual healthcare assistant: Voice and text-enabled medical chatbot," in *2025 3rd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, IEEE, 2025, pp. 1–5.
[19] S. Phatak, M. Deore, and A. Joshi, "AI-assisted diagnostic systems for real-time clinical decision support," *Journal of Biomedical Informatics*, vol. 118, Art. no. 103784, 2021.
[20] S. Soni, J. Gangu, and T. Roberts, "NLP-based symptom checker systems: Accuracy and usability evaluation," in *IEEE International Conference on Healthcare Informatics (ICHI)*, IEEE, 2022, pp. 312–319.
[21] J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.
[22] Esteva *et al*., "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.
[23] R. Miotto *et al*., "Deep learning for healthcare: Review, opportunities and challenges," *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2017.
[24] Rajkomar *et al*., "Scalable and accurate deep learning for electronic health records," *npj Digital Medicine*, vol. 1, no. 1, pp. 1–10, 2018.
[25] Shickel *et al*., "Deep EHR: A survey of recent advances in deep learning for electronic health record analysis," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, 2017.
[26] Vaswani *et al*., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 5998–6008, 2017.
[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.
[28] T. Brown *et al*., "Language models are few-shot learners," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.
[29] Krittanawong *et al*., "Artificial intelligence in precision cardiovascular medicine," *European Heart Journal*, vol. 42, no. 21, pp. 2055–2063, 2021.
[30] World Health Organization, "Global Strategy on Digital Health 2020–2025," WHO Report, 2021.
[31] T. Davenport and R. Kalakota, "The potential for artificial intelligence in healthcare," *Future Healthcare Journal*, vol. 6, no. 2, pp. 94–98, 2019.
[32] Jiang *et al*., "Artificial intelligence in healthcare: Past, present and future," *Stroke and Vascular Neurology*, vol. 2, no. 4, pp. 230–243, 2017.