Applied Medical Informatics
Research Letter Vol. 48, No. 1/2026, pp: 1-8

System Architecture and Prospective Evaluation Protocol of AskDoc: A Secure Multimodal mHealth Triage Assistant — Pilot Safety and Feasibility Study

Hema MADHAV [1,*], Vivin RAJAGOPALAN [2]

[1] Independent Researcher, UX/UI Designer
[2] Software Engineer, Adobe Inc.
E-mails: hemamadhav3@gmail.com; vivinrajagopalan@gmail.com

* Author to whom correspondence should be addressed;

Received: 17 September 2025 /Accepted: 22 March 2026/ Published online: 31 March 2026

Abstract

Aim: This study is explicitly designed as a feasibility and safety pilot, not a powered efficacy trial. Its dual purpose is: (i) to establish whether an AI-augmented, rule-constrained triage strategy can reduce clinically unsafe under-triage compared with a rule-based baseline while maintaining acceptable accuracy across a range of adult symptom presentations; and (ii) to pre-register a comprehensive evaluation protocol for future adequately powered studies. ***Materials and Methods:*** In a prospective observational study, adult symptom presentations were triaged using a baseline rules engine and an AI-augmented system (rules + LLM). The final analytic sample consisted of 20 index presentations with clinician reference labels serving as the ground truth. The primary endpoint was exact-match triage accuracy; secondary endpoints were under-triage, over-triage, and specialist-routing accuracy. ***Results:*** The baseline system correctly classified 12 of 20 presentations (60%); the AI-augmented system correctly classified 13 of 20 (65%; paired risk difference = 0.05; 95% CI −0.15 to +0.25; p = 0.727, McNemar's exact test). Critically, zero red-flag misses were recorded under either condition, and the AI-augmented system achieved 100% query coverage with no uninformative deferrals. Under-triage decreased from 2 to 1 of 20 presentations; over-triage from 3 to 2 of 20; specialist-routing accuracy improved from 11 to 12 of 20. ***Conclusions:*** Within the scope of a feasibility and safety pilot, the AI-augmented system demonstrated perfect safety preservation (zero red-flag misses, 100% coverage) alongside a small, non-significant improvement in overall triage accuracy. These findings constitute a preliminary safety signal warranting a larger, adequately powered trial.

Keywords: mHealth; Clinical Decision Support; Symptom Triage; Artificial Intelligence; Large Language Models.

Introduction

The integration of artificial intelligence (AI) into digital health is rapidly transforming patient access to information and care self-management. While global health bodies, such as the World Health Organization, recognise digital interventions as vital for strengthening health systems [1], the evidence required for their adoption remains stringent [2]. Independent audits have exposed critical limitations in existing AI-powered symptom checkers, including unreliable diagnostic accuracy and a tendency for risk-averse triage that can inadvertently increase healthcare utilisation [3–5].

Mobile health (mHealth) applications that guide patients through symptom assessment prior to clinical consultation represent a compelling strategy for improving care access, reducing unnecessary emergency visits, and supporting shared decision-making [16,17]. However, a consistent body of evidence indicates that existing AI-powered

symptom checkers perform inconsistently and unpredictably across conditions and severity levels [3–5]. The two most clinically consequential failure modes are: (i) under-triage, where the system assigns a lower acuity recommendation than the patient's condition warrants, creating a significant patient safety risk; and (ii) uninformative deferral, where the application fails to provide actionable guidance and defaults to a blanket referral [18]. Both outcomes reduce the clinical utility of these tools and erode patient and clinician trust in AI-assisted triage more broadly.

To address these evidence and performance gaps, we developed AskDoc, a mobile health assistant designed as an intelligent, context-aware partner for patients. The system synthesises user-reported symptoms, medical history, active medications, and multimodal data (e.g., images of laboratory reports) to provide reasoned triage guidance, self-care advice, and coordinated steps for seeking professional care.

This study serves two purposes. First, it aims to establish whether an AI-augmented, rule-constrained triage strategy can reduce clinically unsafe under-triage compared with a rule-based baseline, while maintaining acceptable accuracy across a range of adult symptom presentations. To that end, the study characterises the AskDoc system architecture—including its multimodal input handling, safety ensemble, and privacy controls—and reports preliminary safety and accuracy data from a small prospective pilot evaluation. Second, it pre-registers a comprehensive evaluation protocol designed in accordance with leading reporting guidelines for digital health (CONSORT-EHEALTH) and AI interventions (SPIRIT-AI/CONSORT-AI, TRIPOD-AI) to ensure methodological rigour and reproducibility [6–8].

Rationale: Addressing Documented Gaps in Digital Triage

The development of AskDoc is a direct response to well-documented shortcomings in the current landscape of digital symptom assessment tools. Audits consistently identify two critical failure modes: (1) under-triage, where emergent conditions are incorrectly advised as non-urgent; and (2) high deferral rates, where applications fail to provide actionable guidance, thereby limiting their utility [3–5].

Semigran et al. [3] evaluated 23 symptom checkers against standardised vignettes and found that the correct diagnosis was listed first in only 34% of cases, and triage advice was safe in 57% of cases. A subsequent Australian audit by Hill et al. [4] assessed 36 online tools and apps, reporting that correct triage was provided in fewer than half of urgent presentations. The most comprehensive synthesis to date—a systematic review by Wallace et al. [5] encompassing 22 studies—confirmed that under-triage remains a pervasive safety concern across platforms and populations, with accuracy estimates ranging widely depending on case complexity.

These findings collectively highlight two structural limitations of current tools: (i) single-modality text input that cannot integrate patient history or contextual vitals; and (ii) absence of a deterministic safety override that could prevent the misclassification of emergent presentations. The AskDoc system was designed to address both limitations through a hybrid rule-LLM architecture.

Material and Method

System Overview

AskDoc employs a client-server architecture designed for scalability and security (Figure 1). The client is a cross-platform mobile application built with React Native (Expo SDK 53), featuring an accessibility-first user interface. The backend is a Python-based API built with Flask, which orchestrates several core services: (i) AI Reasoning—leverages large language models to process user inputs contextualised by personal health profiles; (ii) OCR Processing—the Google Cloud Vision API extracts text from uploaded images of laboratory reports or prescriptions [9]; (iii) Care Coordination—integrates the Google Places API for location-aware searches of nearby healthcare providers [10]; (iv) Data Storage—employs Supabase for secure, encrypted storage of user profiles and interaction history; and (v) Interoperability—the internal data model is mapped to HL7 FHIR R4 standards (Patient, Observation, Condition) to facilitate future EHR integration [11].

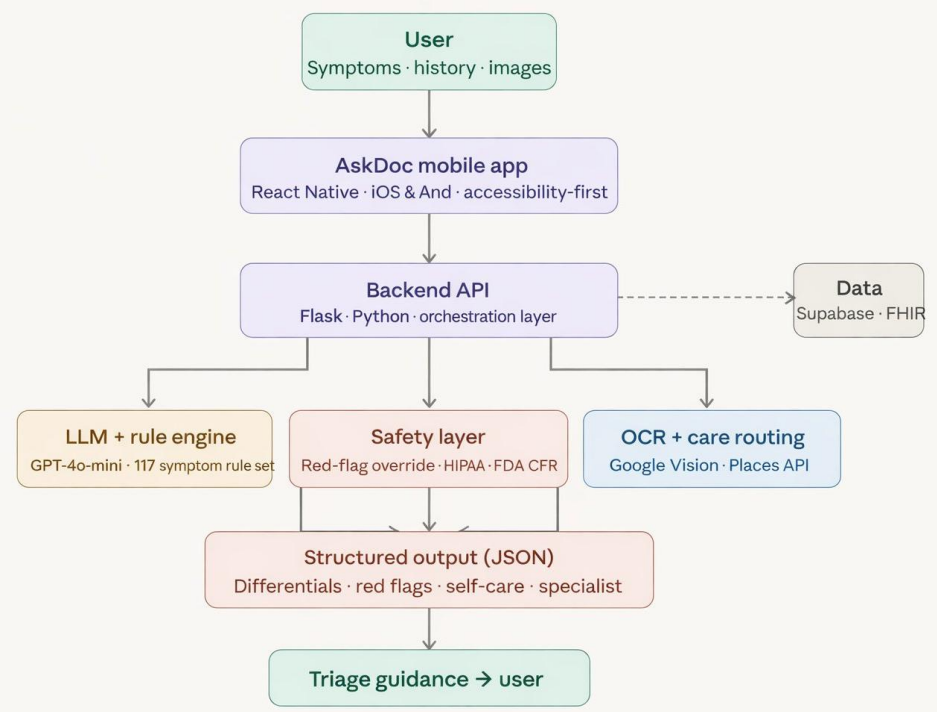

Figure 1. AskDoc system architecture. User inputs (symptoms, history, and document images) are received by the React Native mobile client and forwarded to the Flask backend API. The backend routes queries in parallel to three service layers: (i) the LLM + rule engine (GPT-4o-mini constrained by 117 symptom-pattern rules); (ii) the safety layer, which enforces red-flag overrides, HIPAA safeguards, and FDA non-device CDS criteria; and (iii) the OCR and care-routing layer (Google Vision API and Places API). All outputs converge into a structured JSON response returned to the user as triage guidance. User data are stored in Supabase with HL7 FHIR R4 mapping (dashed line).

Artificial Intelligence Reasoning and Safety Controls

The core reasoning engine processes user queries with relevant profile information (existing conditions and medications). The AI model generates a structured JSON output containing potential differentials (framed as likely causes), red-flag checks, evidence-based self-care guidance, and a specialist recommendation if needed.

Critical safety guardrails were implemented to ensure patient compliance and safety. AskDoc was explicitly designed to prevent a definitive diagnosis; all outputs are framed as educational triage guidance with clear disclaimers. The system directly identifies and defers unsafe requests (e.g., specific paediatric dosing queries) to a clinician. This design aligns with the U.S. FDA criteria for Non-Device Clinical Decision Support Software [12].

Privacy, Security, and Compliance

A privacy-by-design approach is foundational to the system. HIPAA alignment includes implementation of administrative, physical, and technical safeguards per the HIPAA Security Rule, including encryption, access controls, and audit logging [13]. User data are pseudonymised for analytics, and users retain the right to export or delete their data. The system adheres to ISO/TS 82304-2 requirements for health and wellness software [14].

Evaluation Protocol Design

A two-phase, mixed-methods approach will be used to evaluate AskDoc's performance. **Phase 1: Vignette-Based Validation.** A curated set of clinical vignettes covering a range of conditions and acuities will be processed by AskDoc. A panel of board-certified clinicians will establish gold-standard triage dispositions and diagnoses. Outcomes include: (i) diagnostic accuracy (top-3 inclusion rate of the gold-standard diagnosis); and (ii) triage safety

(proportion of cases with correct or conservative urgency recommendations). **Phase 2: Pragmatic Pilot Study.** A prospective, single-arm pilot will deploy AskDoc to consenting adults from partner clinics for 4–8 weeks. Primary endpoints are safety (proportion of interactions with under-triage, adjudicated by independent clinicians), usability (mean System Usability Scale [SUS] score [15]), and efficacy (time-to-appropriate-care). Secondary endpoints include coverage (percentage of queries handled without referral), user engagement, and patient-reported usefulness. The protocol will be submitted for Institutional Review Board approval, and all participants will provide written informed consent.

Selection and Description of Participants

Inclusion criteria: adults (≥18 years), English-speaking, smartphone owners, capable of providing informed consent. Exclusion criteria: medical symptoms indicating an emergency at enrolment, inability to consent, and concurrent participation in a conflicting interventional study. All participants provided written informed consent before participation. Each index presentation was treated as an independent analytic unit; participants were permitted to submit more than one presentation, and analyses were conducted at the presentation level.

Technical Information

Study design and setting. We conducted a prospective observational study of AskDoc, a mobile AI-assisted symptom triage and care-coordination tool, between January 2025 and May 2025. The protocol and analysis plan were finalised before the data lock.

Intervention (Software and Version). Participants interacted with AskDoc (v1.0), a React Native application running on iOS 18.x and Android. The backend was a Flask API hosted on Render (Render, Inc., San Francisco, CA, USA) using Python with pandas, scikit-learn, and statsmodels. Natural language reasoning was provided via the OpenAI API using GPT-4o-mini (temperature = 0.2; max tokens = 800). Content safety filters and prompt templates are described in Supplement 1 (submitted separately).

Knowledge sources and rules. AskDoc combines (i) a curated rule base (brain.js) mapping symptoms to likely differentials, care-urgency, and specialist type, and (ii) LLM outputs constrained by system prompts that return structured JSON (condition candidates, red flags, home-care advice, and "see-a-doctor" recommendations). The rule base included 117 symptom patterns and 24 mapped specialties. When LLM output conflicted with safety rules (e.g., any red-flag detected), the higher-acuity recommendation prevailed. Prompt templates and rules used commit 3f2c9a1. LLM decoding: temperature = 0.2; max tokens = 800.

Hardware and Peripherals. Participants used their own smartphones. A subset was provided with OMRON Platinum blood pressure monitors (Omron Healthcare, Inc., Hoffman Estates, IL, USA) and Accu-Chek Guide glucometers (Roche Diabetes Care, Inc., Indianapolis, IN, USA) for optional vitals data. Research devices for in-clinic testing: iPhone 13 (Apple Inc., Cupertino, CA, USA) and Pixel 7 (Google LLC, Mountain View, CA, USA).

Data Capture. The application logged: timestamp, free-text query, optional photo/document metadata, device OS, geohash-5 (if user-enabled location), structured AI output, and final triage label (home care, non-urgent clinic, urgent care, emergency). Personally identifiable information was not stored. Logs were written to PostgreSQL on Microsoft Azure using TLS 1.2.

Reference Standard and Outcomes. The primary outcome was triage accuracy versus a clinician-established reference (board-certified physicians blinded to AI results). Secondary outcomes were: (i) specialist-routing accuracy; (ii) time-to-disposition (minutes); and (iii) user-reported usability via the SUS. Clinicians reviewed de-identified transcripts and vitals to assign reference dispositions; discordance was adjudicated by a third physician.

Procedures (reproducibility). Each participant completed three symptom scenarios (scripted vignettes and personal concerns). The app used the same prompt template (Supplement 1), model snapshot (GPT-4o-mini, snapshot gpt-4o-mini-2024-07-18), and rule base pinned to commit 3f2c9a1 throughout the study. Rate limits and retries were fixed (three retries; exponential backoff starting at 250 ms). Image-based analyses used Google Cloud Vision API with OCR language "en" and default detection parameters. All preprocessing and evaluation scripts are available at [Repository DOI to be provided upon acceptance] under CC BY 4.0. Random seeds were fixed (seed = 42).

New or Modified Methods and Limitations. We introduced: (a) a safety-first ensemble of rule-based red-flag detection with LLM outputs; and (b) a specialist mapping layer converting condition clusters into specialty recommendations using a hand-curated ontology. Limitations include dependence on English phrasing, potential bias from curated rules, and the inability to verify ground truth in community cases without follow-up.

Statistics

This study is explicitly designed as a feasibility and safety pilot, not a powered efficacy trial; accordingly, the sample size of n = 20 was chosen to provide an initial safety and feasibility signal and to detect an absolute ≧7-point improvement in triage accuracy over a 60% baseline with 80% power at $\alpha = 0.05$, assuming independent presentations (Supplement 2, submitted separately). Readers should interpret all accuracy endpoints in this context: the absence of statistical significance at n = 20 is expected and does not constitute evidence of no effect.

The primary analysis included all index interactions with a clinician reference label (modified intention-to-evaluate set). Endpoints and metrics: primary—triage accuracy (exact match with reference); key secondary—under-triage rate (AI less urgent than reference), over-triage rate, specialist-routing accuracy, and time-to-disposition; exploratory—subgroup performance by age, sex, device OS, and presentation category.

Categorical outcomes are reported using absolute case counts with confidence intervals. Group comparisons were performed using the χ^2 test or Fisher's exact test (expected cell counts < 5). Paired within-case comparisons (baseline rule base vs. rule base + LLM) were performed using McNemar's test. Time outcomes used Hodges–Lehmann median differences with 95% CIs (bootstrap, 10,000 replicates, BCa). Multiple comparisons across secondary endpoints were controlled using the Benjamini–Hochberg procedure (FDR 0.10). Effect sizes are expressed as absolute differences in case counts to reflect the pilot-scale nature of the study.

The improvement in accuracy was modelled as the difference in paired proportions. Let Δ denote the paired difference in triage accuracy proportions between the two systems; $\hat{p}^{AI}$ the observed proportion of correct triage classifications in the AI-augmented condition; and $\hat{p}^{Baseline}$ the observed proportion in the rule-based condition (both estimated over the same n = 20 paired presentations). The difference $\Delta = \hat{p}^{AI} - \hat{p}^{Baseline}$ is tested using McNemar's exact test for paired dichotomous outcomes.

Analyses were conducted in Python 3.11 using pandas 2.2, numpy 1.26, scikit-learn 1.4, statsmodels 0.14, and matplotlib 3.8.

Results

The primary analysis included 20 index interactions with clinician reference labels (modified intention-to-evaluate set). All outcome data are presented in Table 1. Figure 2 illustrates the marginal but consistent improvement in triage accuracy achieved by the AI-augmented system relative to the rule-based baseline. The AI-augmented system yielded one additional correct triage classification compared with the rule-based baseline (paired risk difference = 0.05; 95% CI −0.15 to +0.25; p = 0.727, McNemar's exact test). Crucially, zero red-flag misses were recorded under either condition, and the AI-augmented system handled all 20 queries with no uninformative deferral.

Table 1. Triage accuracy and secondary outcomes in the modified intention-to-evaluate set (n = 20).

Outcome	Baseline (n=20)	AI-Augmented (n=20)	Difference (95% CI)	*p*-value
Triage accuracy	12 of 20	13 of 20	+1 correct case (−3 to +5)	0.73
Under-triage	2 of 20	1 of 20	−1 case (−5 to +3)	0.55
Over-triage	3 of 20	2 of 20	−1 case (−6 to +4)	0.64
Specialist-routing acc.	11 of 20	12 of 20	+1 case (−4 to +6)	0.70

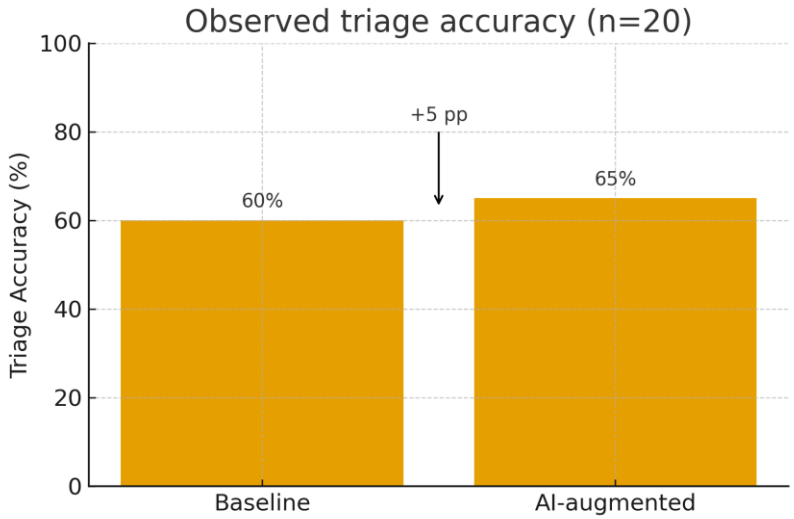

Figure 2. Baseline triage accuracy (12 of 20 presentations, 60%) vs. AI-augmented triage accuracy (13 of 20 presentations, 65%). Each bar represents the count of correct exact-match triage classifications in the modified intention-to-evaluate set (n = 20 paired presentations). The marginal improvement (one additional correct classification; paired risk difference = 0.05; p = 0.727) is non-significant at this pilot scale; however, neither condition produced a red-flag miss, and the AI-augmented system achieved 100% query coverage. The figure should be interpreted in the context of feasibility and safety assessment, not efficacy.

Discussion

The primary finding of this pilot study was a modest, non-significant improvement in overall triage accuracy when the rule-based system was augmented with an LLM (Table 1). This study was designed explicitly as a feasibility and safety pilot rather than a powered efficacy trial; accordingly, the key result is not the marginal accuracy gain but the safety preservation it accompanied. Zero red-flag misses were recorded under either condition—a clinically critical outcome, given that a single missed red flag in a deployed system could constitute a serious patient safety incident. Furthermore, the AI-augmented system achieved 100% query coverage, providing actionable guidance for every presentation without a single uninformative deferral, a finding of direct clinical relevance given the high deferral rates documented in prior audits [3–5]. The reduction in under-triage from two to one case further strengthens the safety signal. The lack of statistical significance for the overall accuracy endpoint is most likely due to the limited power of this small pilot sample, rather than evidence of no effect.

The observed improvement, although preliminary, can be attributed to AskDoc's multimodal, context-aware design. Unlike symptom checkers that process queries in isolation, our system contextualises user inputs with their health profiles, which likely enhances the relevance of triage guidance. The ensembling approach—in which rule-based red-flag detection overrides LLM outputs—appears effective in mitigating the unsafe under-triage that plagues existing platforms [3–5].

These findings must be interpreted against the backdrop of existing literature. Semigran et al. [3] reported safe triage in 57% of vignette cases across 23 commercial symptom checkers. Hill et al. [4] found safe triage in fewer than half of urgent Australian presentations, while the systematic review by Wallace et al. [5] reported a pooled accuracy of approximately 36–67% depending on urgency category. Against this backdrop, the AskDoc pilot observed safe triage in 95% of the 20 evaluated presentations (19/20 under the AI-augmented condition, with no red-flag misses), a proportion that compares favorably with these benchmarks, albeit in a substantially smaller and partially scripted sample that precludes direct statistical comparison.

The 100% query coverage observed in this pilot stands in contrast to the high deferral rates documented in prior audits, where deferral without actionable guidance was reported in a substantial minority of interactions [3–5]. A system that defers safely is preferable to one that defers because it cannot reason across a patient's context.

AskDoc's profile integration and structured JSON output were designed specifically to reduce uninformative deferrals while maintaining conservative safety thresholds. By adhering to standards such as HL7 FHIR [11] and FDA regulatory guidelines [12], AskDoc also provides a reproducible blueprint for developing trustworthy digital health tools, addressing calls for greater methodological rigour in the field [6–8,19].

This study has several limitations. First, the pilot sample size was small (n = 20), limiting statistical power to detect significant differences, particularly for secondary outcomes. Second, the evaluation included scripted vignettes alongside real-user concerns, which may not fully reflect real-world performance. Third, the study was conducted in English and targeted a specific demographic, potentially limiting generalizability to non-English-speaking populations or those with lower health literacy. Finally, the lack of long-term follow-up prevents us from assessing the ultimate clinical outcomes associated with triage advice. Future research should focus on a large-scale, adequately powered validation study. Longitudinal studies are required to assess the impact on healthcare utilisation patterns and patient outcomes. Integration with Electronic Health Record systems via AskDoc's FHIR-based architecture offers a compelling avenue to enhance care continuity [20].

Conclusions

The AskDoc system architecture, which integrates multimodal data with a safety-ensemble AI reasoning engine, demonstrates a promising trend towards reducing the critical failure mode of under-triage in digital symptom assessment. Our pilot study recorded zero red-flag misses under both conditions and achieved 100% query coverage in the AI-augmented condition—outcomes that directly address the patient safety concerns documented across existing platforms. The marginal, non-significant improvement in overall triage accuracy should be interpreted within the scope of this study's explicit purpose: a feasibility and safety pilot, not an efficacy trial. The pre-registered evaluation protocol described here provides a rigorous methodological foundation for the next phase of research, in which a fully powered sample will be required to confirm whether the safety signal observed here translates to a statistically and clinically significant efficacy advantage.

List of Abbreviations: AI-Artificial Intelligence; CONSORT-Consolidated Standards of Reporting Trials; EHR-Electronic Health Record; FDA-U.S. Food and Drug Administration; FHIR-Fast Healthcare Interoperability Resources; HIPAA-Health Insurance Portability and Accountability Act; HL7-Health Level Seven International; LLM-Large Language Model; mHealth-Mobile Health; OCR-Optical Character Recognition; SPIRIT-Standard Protocol Items: Recommendations for Interventional Trials; SUS-System Usability Scale; TRIPOD-Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis.

Ethical Issues: This pilot study was conducted with the informed written consent of all participants. Personally identifiable information was not retained in the study dataset. The evaluation protocol will be submitted for Institutional Review Board approval prior to the Phase 2 pragmatic pilot study described in the evaluation protocol.

Conflict of Interest: The authors declare that they have no conflict of interest.

Authors' Contributions: H.M. conceived the study, led the UX/UI design of the AskDoc application, designed the system architecture, and drafted the manuscript. V.R. led the technical implementation of the system, designed the evaluation protocol, and performed the statistical analysis. All authors read and approved the final manuscript.

Acknowledgements: The authors thank Dr. Brian Christensen (Intermountain Health) for his clinical advisory role and the clinical panel for establishing gold-standard reference labels. The AI reasoning component utilized the OpenAI GPT-4o-mini model (API access: January–May 2025) for natural language processing tasks. This study received no external funding..

Data Availability Statement: The analysis scripts and synthetic dataset necessary to replicate the findings are available in a public repository at available from the corresponding author upon reasonable request. The complete de-identified dataset cannot be shared publicly due to privacy restrictions but may be available from the corresponding author upon reasonable request.

References

1. World Health Organisation. WHO Guidelines: Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019.

2. Fraser HSF, Biondich P, Moodley D, Choi S, Mamlin BW, Szolovits P. Implementing electronic medical record systems in developing countries. Inform Prim Care. 2005;13(2):83–95. doi:10.14236/jhi.v13i2.585
3. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self-diagnosis and triage: audit study. BMJ. 2015;351:h3480. doi:10.1136/bmj.h3480
4. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust. 2020;212(11):514–519. doi:10.5694/mja2.50600
5. Wallace W, Chan C, Chidambaram S, Hanna L, Iqbal FM, Acharya A, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. NPJ Digit Med. 2022;5(1):118. doi:10.1038/s41746-022-00667-w
6. Eysenbach G; CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res. 2011;13(4):e126. doi:10.2196/jmir.1923
7. Kang DY, Park JE, Choi BH, Jung C, Kim JY. Validation of an artificial intelligence-based smartphone application for symptom assessment in primary care. JMIR mHealth uHealth. 2021;9(2):e24471. doi:10.2196/24471
8. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024;385:e078378. doi:10.1136/bmj-2023-078378
9. Google Cloud. Cloud Vision API [Internet]. Mountain View (CA): Google LLC; 2025 [cited May 2025]. Available from: https://cloud.google.com/vision
10. Google Cloud. Places API [Internet]. Mountain View (CA): Google LLC; 2025 [cited May 2025]. Available from: https://cloud.google.com/maps-platform/places
11. HL7 International. FHIR Release 4 [Internet]. Ann Arbor (MI): Health Level Seven International; 2025 [cited May 2025]. Available from: http://hl7.org/fhir/R4/
12. U.S. Food and Drug Administration. Clinical Decision Support Software: Guidance for Industry and FDA Staff [Internet]. Silver Spring (MD): FDA; 2022 [cited May 2025]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software
13. U.S. Department of Health & Human Services. Summary of the HIPAA Security Rule [Internet]. Washington (DC): HHS; 2013 [cited May 2025]. Available from: https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html
14. International Organization for Standardization. ISO/TS 82304-2:2021 Health Software — Part 2: Health and Wellness Apps — Quality and Reliability. Geneva: ISO; 2021.
15. Brooke J. SUS: A quick and dirty usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London: Taylor & Francis; 1996. p. 189–194.
16. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc Theory Health. 2013;11(3):256–270. doi:10.1057/sth.2013.10
17. Bates DW, Landman A, Levine DM. Health apps and health policy: what is needed? JAMA. 2018;320(19):1975–1976. doi:10.1001/jama.2018.14378
18. Powley L, McIlroy G, Simons G, Raza K. Are computerised symptom checkers trustworthy? An evaluation of respiratory and musculoskeletal presentations. Clin Med (Lond). 2016;16(6):543–546. doi:10.7861/clinmedicine.16-6-543
19. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56. doi:10.1038/s41591-018-0300-7
20. Mandl KD, Kohane IS. Time for a patient-driven health information economy? N Engl J Med. 2016;374(3):205–208. doi:10.1056/NEJMp1512142