Human LifeExtender: An AI-Driven, Multilingual Framework for Personalized and Preventive Healthcare in Resource Diverse Environments

Hrishikesh Khandelwal[1] Khushi Puri[2] Ayurshi Patil[3] Sujal Sayam[4] Dr. Mahendra Pund[5]
[1,2,3,4,5]Department of Computer Science Engineering
[1,2,3,4,5]Prof Ram Meghe Institute of Technology and Research, India

Abstract — The limited accessibility of preventive healthcare in India is attributed to linguistic diversity, fragmented mHealth tools, and limited integration of wearable data with safety-controlled guidance. Human LifeExtender, a multilingual mobile healthcare framework that integrates AI-assisted wellness education, allergy-aware nutrition support, medication reminders and hydration reminders, community fitness engagement, and SOS-based emergency assistance under explicit privacy and safety constraints, is the focus of this paper. The contribution is a literature-gender.

Keywords: Mobile Health (mHealth), Multilingual Healthcare, Wearable-Agnostic Monitoring, Artificial Intelligence, Personalized Nutrition, Preventive Healthcare, Safety-Aware Health Assistance

I. Introduction

Access to timely, understandable, and trustworthy health information remains uneven in India, particularly for users who rely on regional languages, low-bandwidth connectivity, or intermittent access to clinical services. Preventive healthcare—covering routine activity, nutrition, medication adherence, and early awareness of warning signs—can reduce avoidable complications, yet generic guidance often fails to account for individual health context, dietary practices, and linguistic diversity [2], [3], [6]. Mobile health (mHealth) applications have therefore attracted attention as scalable channels for wellness support and preventive education, and recent reviews document their potential across monitoring, adherence, nutrition, and emergency assistance [3], [4], [14], [19]. However, the literature also shows that these functions are usually studied in isolation, under controlled or component-specific conditions, and with limited attention to integrated deployment in multilingual, resource-constrained settings [1], [4], [20].

Three strands of prior work are directly relevant. First, wearable and sensor-based monitoring has advanced considerably. Wireless body area networks, IoT patient-monitoring architectures, and FHIR-based integration pipelines demonstrate that physiological and activity data can be collected, normalized, and transferred to standards-compliant repositories [1], [15], [19]. On-device inference has further shown that lightweight models can process health-related data locally, reducing dependence on cloud transmission [5], [18]. Yet these studies focus primarily on data acquisition, interoperability, or privacy-preserving inference; they do not report an integrated consumer application that combines wearable-agnostic data capture with multilingual wellness guidance, nutrition support, reminders, and emergency assistance. In addition, the FHIR integration case study acknowledged that its data dictionary relied on vendor-specific identifiers rather than clinical coding systems and had not been externally certified [1], while the on-device prototype reported partial implementation and lacked a completed decision-fusion module [18].

Second, conversational and retrieval-augmented systems have been explored for medical question answering and public-health communication. Multilingual chatbots using Rasa, LLM fallback, and rule-based safety layers have achieved high intent-classification accuracy on small test sets in Indian languages [10], and retrieval-augmented generation (RAG) has been used to ground medical chatbot responses in external corpora [11], [16]. However, a systematic review of 137 chatbot health-advice studies found that most evaluated closed-source models, rarely described prompt engineering, and often used subjective performance criteria, while fewer than one-third addressed ethical, regulatory, or patient-safety implications [4]. A separate review highlighted persistent concerns regarding interpretability, bias, data requirements, and legal accountability for LLMs in healthcare [14]. Moreover, evidence indicates that LLM accuracy can decline substantially in lower-resource languages, with reported drops of 10.29 for Arabic, 7.24 for Bengali, and 4.64 for Hindi relative to Chinese/English baselines [2]. These findings caution against assuming that multilingual capability alone ensures equitable or safe health communication.

Third, personalized nutrition and adherence support have been investigated using machine learning and recommender systems. Hybrid approaches combining cosine similarity with random forest classification, stacking classifiers using BMI/BMR/TDEE, and time-aware graph-clustering recommenders have reported promising accuracy and nutritional alignment on food datasets [6], [7], [13], while systematic reviews and comparative studies have catalogued the strengths and weaknesses of SVM, Naive Bayes, decision trees, KNN, and ensemble methods for diet recommendation [12]. Medication-adherence apps have likewise shown measurable improvements in adherence and high user satisfaction in a systematic review [3]. Nonetheless, these studies typically evaluate recommendation accuracy or adherence in isolation; they do not integrate allergy- or medication-aware safety constraints, multilingual interaction, wearable-derived context, or emergency escalation. The recommender literature also acknowledges limitations such as restricted cultural-food coverage, absence of clinical validation, and reliance on offline datasets [6], [7], [13].

II. Literature Review

The reference corpus spans multilingual natural language processing, wearable and sensor integration, personalized nutrition, medication adherence, emergency assistance, community fitness, and mobile-health privacy. Rather than summarizing each study in isolation, this section synthesizes them thematically, compares their methodological choices, and identifies limitations that motivate the proposed Human LifeExtender (HLE) framework. Where authors explicitly acknowledge a limitation, it is reported as such; comparative

All rights reserved by www.ijsrd.com

observations drawn across studies are presented as analytical inference.

A. Multilingual Healthcare Interaction and Linguistic Equity

A first cluster of studies examines how language affects the accessibility and reliability of digital health communication. Han et al. [2] provide direct experimental evidence that LLM performance is uneven across languages. Using 442 medical licensing questions sampled from MedQA and translated into Arabic, Bengali, Hindi, and Portuguese, they reported accuracy declines relative to Chinese/English baselines of 10.29, 7.24, 4.64, and 0.73, respectively. The authors attribute the sharper declines to typological distance and underrepresentation in training corpora, and they acknowledge that translation errors may themselves contribute to the observed disparities [2]. This finding is significant for the Indian context because it indicates that simply adding a language interface does not guarantee equitable performance; language-specific evaluation is necessary.

Medicherla et al. [10] address this concern at the application level by developing a Rasa-based multilingual public-health chatbot supporting Hindi, English, Telugu, and Hinglish, with a rule-based safety layer restricting responses to non-prescriptive guidance. The system reported an overall accuracy of 0.961, precision of 0.948, and F1-score of 0.951 on 35 test queries, with intent accuracy above 85 for English and Hindi and approximately 80 for Telugu and Hinglish. The authors acknowledge that the evaluation used a limited dataset and only four languages, and that large-scale deployment remains future work [10]. Compared with [2], this study demonstrates practical feasibility but does not establish clinical safety or evaluate the safety layer itself. Notably, Marathi—a major language in western India—is absent from both studies, which is directly relevant to HLE's proposed trilingual scope.

The broader chatbot literature reinforces these concerns. Huo et al. [4] systematically reviewed 137 chatbot health-advice studies and found that 99.3 evaluated closed-source models without sufficient version information, 99.3 did not describe a prompt-engineering phase, and 65.0 used subjective means to define successful performance. Fewer than one-third addressed ethical, regulatory, or patient-safety implications. The authors acknowledge that the field is dynamic and that much of the literature used hypothetical patient cases [4]. This review is important methodologically: it shows that multilingual capability claims must be accompanied by transparent reporting of model characteristics, prompts, and evaluation criteria—requirements that HLE must adopt if it is to make defensible claims.

B. LLM-Based Medical Assistance, Retrieval Augmentation, and Safety

A second cluster investigates how large language models can be grounded and governed in medical contexts. Zou and He [14] provide a narrative review of LLM applications in clinical decision support, electronic health record processing, question answering, education, and drug discovery, while identifying challenges relating to training-data requirements, interpretability, ethical and legal accountability, and bias. As a review, it does not report primary empirical results, and it acknowledges the need for multidisciplinary collaboration and clearer regulation [14].

Two studies offer concrete RAG implementations. Ailaparam et al. [11] combined Mistral-7B-Instruct-v0.3 with a Pinecone vector database, retrieving the top three document chunks from WHO, PubMed Central, and CDC sources. Across five common medical queries evaluated by three reviewers, they reported average accuracy of 4.2/5, relevance of 4.4/5, and response time of 6.8 seconds, with substantial inter-rater agreement (Fleiss' $\kappa = 0.79$). The authors explicitly acknowledge latency due to local CPU inference, limited handling of complex queries, dependency on external APIs, residual hallucinations, and retrieval dependency [11]. These limitations are directly relevant to HLE's design: RAG improves grounding but does not eliminate hallucination, and latency must be managed for mobile use.

MGPRAg [16] advances this line by introducing a medical knowledge-graph-based RAG framework with task parsing, a Dynamic-Frozen Collaborative Encoder, and a Medical Value-Driven retrieval algorithm. On the CMCQA dataset, it achieved GPT-4o ranking scores of 1.6264 and BERTScore precision/recall/F1 of 0.9137/0.9244/0.9189, while retrieving substantially fewer tokens than TOG and GraphRAG. The authors acknowledge limited generalizability of their entity-matching dataset, annotation bias, incomplete knowledge-graph coverage for rare diseases, and reliance on automated metrics without clinician evaluation [16]. Compared with document-chunk RAG [11], MGPRAg demonstrates that structured retrieval can improve precision and reduce redundancy, but it also shows that knowledge-graph coverage constraints can cause retrieval failure for rare or atypical cases—an important consideration for a general-wellness application that cannot assume comprehensive clinical coverage.

Taken together, [4], [11], [14], and [16] indicate that RAG and knowledge-graph grounding are promising but not sufficient for safe medical advice. They support HLE's decision to restrict AI assistance to general, non-prescriptive wellness education, to include uncertainty communication, and to escalate to professional care rather than generate diagnostic conclusions.

C. Wearable Integration, Sensor Monitoring, and On-Device Prediction

A third cluster addresses data acquisition and processing. Abedian et al. [1] demonstrate a prototype pipeline in which Garmin Vivoactive 4 data flow through the Fitrock hub and API into a Kodjin FHIR server, with Liquid-based mapping to Patient, Observation, and Bundle resources and GDPR-oriented pseudonymization. The authors acknowledge that the Fitrock data dictionary uses vendor-specific identifiers rather than LOINC or SNOMED CT, is not externally certified, and is not integrated with HealthData@EU [1]. This study is directly relevant to HLE's wearable-agnostic objective because it shows both the feasibility and the limitations of vendor-mediated integration. However, it does not address multilingual interaction, nutrition, reminders, or emergency support.

All rights reserved by www.ijsrd.com

Javaid et al. [15] review medical sensors and wireless body area network integration with edge, fog, mobile cloud, and cloud computing. They synthesize sensor types, communication protocols, and architectural layers, and identify energy consumption, interoperability, security, biocompatibility, and complexity as open challenges. As a review, it does not evaluate an application-level system [15]. Mahmmod et al. [19] similarly review IoT patient-monitoring architectures using a five-layer model comprising sensing, communication, edge/fog, cloud, and application/action layers, and identify data and resource management, security and privacy, heterogeneity, data integration, device design, monitoring accuracy, trustworthiness, and transmission delay as open research issues [19]. These reviews collectively justify HLE's layered architecture and its emphasis on resource-aware processing, but they also indicate that integrated consumer-facing evaluation remains scarce.

Two studies examine local inference and mobile prediction. Wang et al. [5] evaluate five small language models on the PMData dataset for stress, readiness, fatigue, and sleep-quality prediction. TinyLlama-1.1B used 4.31 GB of RAM with a time-to-first-token of 0.48 seconds and achieved a mean stress MAE of 0.615, outperforming the LLM mean of 0.639, while readiness MAE was 1.91 versus 2.56 for LLMs. The authors acknowledge that efficiency was evaluated on a standard NLP dataset, that only one health dataset was used, and that only zero-shot performance was examined [5]. Rangi and Song [18] describe an Android proof-of-concept combining TensorFlow Lite, MediaPipe, and Gemma 2B for on-device health assessment using blood pressure, heart rate, and SpO_2 inputs, with rule-based findings and LLM-generated explanations. They acknowledge that the decision-fusion module was not completed, that React Native integration issues forced an Android-only implementation, and that the prototype was partial [18]. Compared with cloud-based approaches [11], these studies support HLE's privacy-conscious local processing option, but neither demonstrates multilingual capability or integration with nutrition, reminders, or emergency workflows.

D. Personalized Nutrition and Dietary Recommendation

A fourth cluster focuses on machine-learning-based nutrition recommendation. Puthran et al. [6] combine cosine similarity with random forest classification using a dataset of more than 1,500 foods and 25 nutritional features across ten health conditions. They report 80.34 accuracy and 83.2 nutritional alignment, with the hybrid approach outperforming content-based, collaborative-filtering, and rule-based baselines. The authors acknowledge that the food database could be expanded for cultural diversity, that temporal recommendations and feedback integration remain future work, and that complete meal recommendations were not implemented [6].

Rostami et al. [7] propose TDLGC, a time-aware hybrid recommender using graph clustering for users and BERT-based deep embedded clustering for foods. On a crawled Allrecipes dataset comprising 68,768 users, 45,630 foods, and 1,093,845 ratings, they reported Precision@10 of 0.0721, Recall@10 of 0.0691, F1@10 of 0.0705, AUC of 0.6812, and NDCG@10 of 0.0497, outperforming LDA, HAFR, and FGCN. The authors acknowledge that side information such as gender, age, weight, height, location, and culture was not yet incorporated and that nutritional characteristics should be used in future work [7].

Abdelmageed et al. [13] apply a stacking classifier to the Food.com dataset, using BMI, BMR, and TDEE alongside recipe features, and report 95.5 accuracy, 93.07 precision, 97.92 recall, 94.5 F1-score, and 0.978 ROC AUC. They acknowledge constraints in the supplied extract without fully detailing them [13]. Shah et al. [12] review SVM, Naive Bayes, decision trees, Extra Trees, KNN, and random forest for diet recommendation and propose a tentative voting-classifier workflow, but do not report primary experimental results [12].

Collectively, these studies demonstrate that nutrition recommendation can be personalized using health conditions, goals, and preferences. However, none of the four integrates allergy- or medication-aware safety constraints, multilingual interaction, or wearable-derived context. Their evaluation is also predominantly dataset-based rather than clinical. These gaps directly motivate HLE's requirement that nutrition guidance be framed as general wellness support, with allergy constraints treated as a design requirement rather than a validated clinical guarantee [6], [7], [13].

E. Medication Adherence and Health Reminders

Pérez-Jover et al. [3] systematically reviewed 11 studies on mobile apps for treatment adherence, with sample sizes ranging from 16 to 99 participants. Seven studies confirmed increased adherence; five reported statistically significant before-and-after improvements ranging from 7 to 40, and average satisfaction was 8.1 out of 10. The authors acknowledge that the review was limited to English and Spanish publications, excluded conference abstracts, and could not control for voluntary nonadherence [3]. Compared with the AI-oriented studies above, this review provides stronger evidence for reminder-based adherence support, but its evidence base predates recent AI and wearable developments and does not address multilingual or integrated emergency workflows.

III. RESEARCH GAP

Within the selected corpus, no study reports an evaluated, integrated, multilingual, wearable-agnostic consumer health framework that combines preventive wellness education, personalized nutrition with safety constraints, medication and hydration reminders, community physical activity, and SOS-based emergency assistance, while explicitly specifying privacy and safety controls for the Indian context. The literature instead provides strong but disconnected evidence for each individual component, often evaluated under controlled or dataset-specific conditions, with acknowledged limitations that become more consequential when the components are combined [1]–[20]. This is a corpus-level gap: it does not establish that no such system exists anywhere, only that the reference literature does not demonstrate or evaluate one. Four specific sub-gaps follow from this synthesis.

- G1 — Integration gap. Wearable integration studies focus on data acquisition and interoperability [1], [15], [19]; chatbot studies focus on conversational accuracy

All rights reserved by www.ijsrd.com

[10], [11], [16]; nutrition studies focus on recommendation accuracy [6], [7], [13]; adherence and emergency studies focus on their respective workflows [3], [9], [17]. The corpus contains no demonstration of how these modules would share user context, coordinate safety rules, or avoid conflicting guidance when deployed together. Cross-module coordination is therefore an open design and evaluation problem.

- G2 — Linguistic coverage gap. Hindi and English are represented in the multilingual literature [2], [10], but Marathi is absent from the corpus. Han et al. [2] demonstrate that accuracy can decline substantially in lower-resource languages, and Medicherla et al. [10] report lower intent accuracy for Telugu and Hinglish than for English and Hindi. The corpus therefore does not establish whether Marathi preventive-health interaction can be delivered with acceptable reliability, nor does it provide a language-specific evaluation protocol for it.
- G3 — Safety-constraint gap. Nutrition studies personalize by health condition, goals, and preferences but do not integrate allergy- or medication-aware constraints [6], [7], [12], [13]. RAG studies improve grounding but acknowledge residual hallucination and retrieval dependency [11], [16], and systematic reviews report limited attention to ethics, regulation, and patient safety in chatbot studies [4], [14]. The corpus therefore does not specify how safety rules should constrain AI-generated wellness guidance or nutrition recommendations, nor how uncertainty and escalation should be communicated to users.
- G4 — Evaluation gap. Evaluation practices in the corpus are heterogeneous: small test sets [9], [10], [17], dataset-level metrics without clinical validation [6], [7], [13], automated metrics without clinician involvement [16], and subjective performance definitions in a majority of chatbot studies [4]. Privacy reviews note that few studies involve end users or healthcare professionals [20]. Consequently, the corpus provides little guidance on how an integrated multilingual health application should be evaluated for safety-constraint violations, multilingual correctness, latency, or privacy characteristics.

The gap motivates the following research questions. These are formulated as questions to be investigated through design and planned evaluation, not as findings already established.

- RQ1 (Integration): How can wearable-agnostic data capture, multilingual wellness assistance, nutrition guidance, reminders, community activity, and emergency assistance be coordinated within a single mobile architecture without duplicating or conflicting with one another?
- RQ2 (Linguistic equity): What design and evaluation measures are required to support English, Hindi, and Marathi preventive-health interaction with acceptable and measurable reliability, given documented language-dependent performance disparities [2], [10]?
- RQ3 (Safety constraints): How should allergy-aware and medication-aware constraints, retrieval grounding, uncertainty communication, and escalation to professional care be specified so that AI assistance remains general and non-prescriptive [4], [11], [16]?
- RQ4 (Privacy): Which privacy-by-design measures—data minimization, local or on-device processing where feasible, consent management, and access control—are appropriate for a multilingual consumer health application handling wearable and self-reported data [18], [20]?
- RQ5 (Evaluation): What evaluation criteria and procedures are suitable for assessing such an integrated framework, including multilingual correctness, recommendation relevance, safety-constraint violations, latency, privacy characteristics, and usability [4], [20]?

IV. Proposed Human Life-Extender Framework

Human LifeExtender (HLE) is proposed as a modular, multilingual, wearable-agnostic mobile healthcare framework intended for general wellness, preventive health education, and emergency assistance in the Indian context. It is described here as a design specification and, where applicable, as prototype scope; the architecture does not assume that any specific module has been clinically validated, and no module is presented as a diagnostic or prescribing component. The framework is positioned as an integration layer that coordinates capabilities which the reference corpus has demonstrated only in isolation: wearable and sensor data capture [1], [15], [19], multilingual conversational assistance [2], [10], retrieval-grounded AI guidance [11], [16], personalized nutrition recommendation [6], [7], [13], reminder-based adherence support [3], community fitness engagement [17], and SOS-based emergency assistance [9]. Five design principles govern the architecture.

- P1 — Integration without conflation. Modules share a common user context but retain distinct responsibilities and safety boundaries. Multilingual wellness education, nutrition guidance, reminders, and emergency assistance are not collapsed into a single generative component, because the corpus shows that component-level reliability varies substantially and that conflating functions can propagate errors across workflows [4], [11], [16].
- P2 — Wearable-agnostic abstraction. Device data are accessed through an abstraction layer rather than hard-coded to a single vendor, reflecting both the feasibility and the vendor-specific limitations reported in FHIR-based integration work [1] and the heterogeneity identified in sensor and IoT reviews [15], [19].
- P3 — Safety-bounded AI assistance. Generative assistance is restricted to general, non-prescriptive wellness education, supported by retrieval grounding where appropriate, uncertainty communication, and escalation to professional care [4], [10], [11], [14], [16].
- P4 — Privacy-conscious processing. Data minimization, consent management, access control, and local or on-device processing where feasible are treated as architectural requirements rather than optional features [18], [20].
- P5 — Graceful degradation. Because connectivity, device capability, and API availability vary, each module

All rights reserved by www.ijsrd.com

must define a reduced-functionality mode that preserves safety and does not fabricate results [5], [18], [19].

A. Layered Architecture

The HLE architecture adopts a layered structure informed by the three-layer mobile crowd-sensing model [8], the edge/fog/cloud integration model for wireless body area networks [15], and the five-layer patient-monitoring architecture comprising sensing, communication, edge/fog, cloud, and application/action layers [19]. HLE adapts these into six logical layers.

Layer	Responsibility	Representative corpus basis
L1 Sensing and device abstraction	Access smartphone sensors and compatible wearables through a normalized interface; handle permission and availability states	[1], [15], [19]
L2 Communication and synchronization	Local transport (e.g., Bluetooth-class links) and wide-area synchronization; offline buffering and retry	[15], [19], [8]
L3 Edge/on-device processing	Local rule evaluation, lightweight inference, caching, and pre-screening before any remote call	[5], [18]
L4 Application services	Multilingual routing, AI assistance with safety layer, nutrition constraint engine, reminder engine, activity/community services, emergency workflow	[3], [6], [9], [10], [11], [16], [17]
L5 Data and governance	Profile store, consent and preference store, audit log, encrypted storage, retention policy	[1], [18], [20]
L6 Presentation and interaction	Trilingual user interface, accessibility, disclaimers, escalation prompts, permission explanations	—

Table I: Proposed HLE Layered Architecture Proposed HLE Layered Architecture

B. Core Modules

The core modules are described below:

- M1 — Device and Sensor Abstraction. M1 exposes a normalized data schema (for example, timestamped step, sleep, heart-rate, and activity-intensity records) and attaches provenance (source device, API version), quality flags (missing, implausible, low-confidence), and availability state. The abstraction is necessary because integration studies report that vendor dictionaries may rely on proprietary identifiers rather than clinical coding systems [1], and because sensor heterogeneity and interoperability are recurring challenges [15], [19]. M1 does not claim clinical-grade measurement validity; it records what a device reports and marks uncertainty explicitly.
- M2 — User Profile and Consent. M2 maintains the profile fields required by downstream modules: language preference, declared health conditions, medications, allergies, dietary preference, activity goals, and emergency contacts. It also stores consent state per data category (for example, activity data, location for SOS, nutrition preferences) so that a user can enable or disable categories independently. This design responds to privacy reviews that emphasize consent management, data minimization, and transparency [20], and to integration work that implemented pseudonymization and participant-controlled sharing [1].
- M3 — Multilingual Interaction and Language Routing. M3 detects or receives the user's language, routes the request to the appropriate language-specific processing path, and applies localized response templates. The design follows evidence that LLM accuracy can differ substantially across languages [2] and that a language-routing pipeline with separate language components is practically feasible [10]. M3 is deliberately separated from M4 so that a language-specific failure can be contained and reported rather than silently degrading the entire interaction.
- M4 — AI Wellness Assistant with Safety Layer. M4 is the generative component, constrained to general wellness and preventive education. Its pipeline has four stages: (i) intent and constraint extraction from the routed query; (ii) retrieval of relevant general-wellness content from a curated, versioned corpus; (iii) constrained generation using retrieved context and the user's non-sensitive profile attributes; and (iv) a safety layer that screens output for prescriptive or diagnostic language, unsupported claims, and unsafe content before display. The RAG design is motivated by reported grounding benefits [11], [16], while the safety layer and non-prescriptive scope respond to documented hallucination and reporting weaknesses [4], [11], [14]. M4 must also communicate uncertainty and provide an escalation prompt when the query appears to require clinical assessment.
- M5 — Nutrition Recommendation and Constraint Engine. M5 produces meal-category suggestions using user goals, dietary preference, and declared conditions, and applies a constraint layer that filters items conflicting with declared allergies or medication-related restrictions where such rules are available. The recommendation core draws on hybrid cosine-similarity and random-forest approaches [6], time-aware recommender designs [7], ensemble and stacking classifiers [12], [13]. The constraint layer is a proposed extension: the corpus's nutrition studies do not integrate allergy- or medication-aware safety constraints [6], [7], [12], [13], so M5 must be explicit that its constraints are rule-based safeguards rather than clinically validated dietary advice.

All rights reserved by www.ijsrd.com

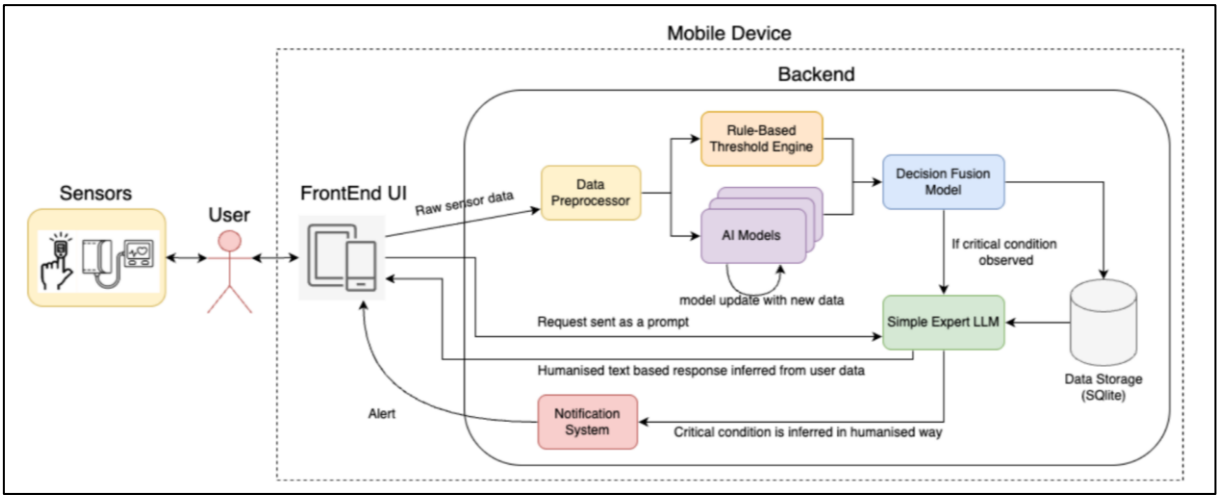


Fig. 1. Proposed Architecture of the On-device Mobile Health Monitoring System.

- M6 — Reminder and Adherence. M6 manages medication and hydration schedules, confirmation actions, missed-dose summaries, and optional caregiver notification with user consent. Its design is grounded in adherence evidence and satisfaction findings [3], while its notification and consent behavior is constrained by M2.
- M7 — Activity and Community Engagement. M7 supports activity goals, progress visualization, optional location-based challenges, and community leaderboards, drawing on gamification and fitness-synchronization precedent [17] and on crowd-sensing considerations regarding recruitment, incentives, and compliance [8]. Community features are opt-in and bounded by location-permission consent stored in M2.
- M8 — Emergency Assistance and Facility Access. M8 implements a user-initiated SOS flow with a cancellation window, an alert payload containing a location link and a user-approved medical summary, notification of stored emergency contacts, and a nearby-facility query subject to location permission and service availability. The design follows the SOS and false-positive-mitigation concepts reported in accident-rescue work [9], while deliberately avoiding claims of automatic crash detection accuracy.
- M9 — Data Governance and Synchronization. M9 manages encrypted local storage, synchronization policy, audit logging, retention, and standards-oriented export where supported. It reflects FHIR-oriented mapping practice [1], on-device storage and local processing [18], and the privacy-evaluation concerns identified in the scoping review [20].

C. *Data Flow*

The end-to-end data flow is described below:

1) Capture: M1 collects device and smartphone data; M2 collects or updates profile and consent entries.
2) Local processing: At L3, deterministic checks run: plausibility filtering on sensor records, constraint extraction from the profile, and language detection for incoming text or voice input.
3) Routing: M3 assigns the language path and forwards the request to M4, M5, or M6 according to intent.
4) Assisted generation or retrieval. M4 retrieves grounded content and generates a constrained response; M5 computes similarity-based candidates and applies the constraint filter; M6 evaluates schedule state.
5) Safety screening: M4 output and M5 explanations pass through the safety layer, which blocks prescriptive, diagnostic, or unsafe content and attaches uncertainty or escalation language.
6) Presentation: L6 renders the localized response with disclaimers and, where relevant, escalation prompts.
7) Persistence: M9 writes audit entries and, subject to consent, persists records locally and synchronizes where configured.
8) Emergency flow: An SOS trigger from L6 is validated for cancellation, assembled into a payload, and dispatched through M8 without passing through the generative pipeline, so that an AI service outage cannot block emergency notification. This separation is a direct architectural response to the reliability expectations of emergency workflows [9] and to the availability constraints of remote services [11], [19].

All rights reserved by www.ijsrd.com

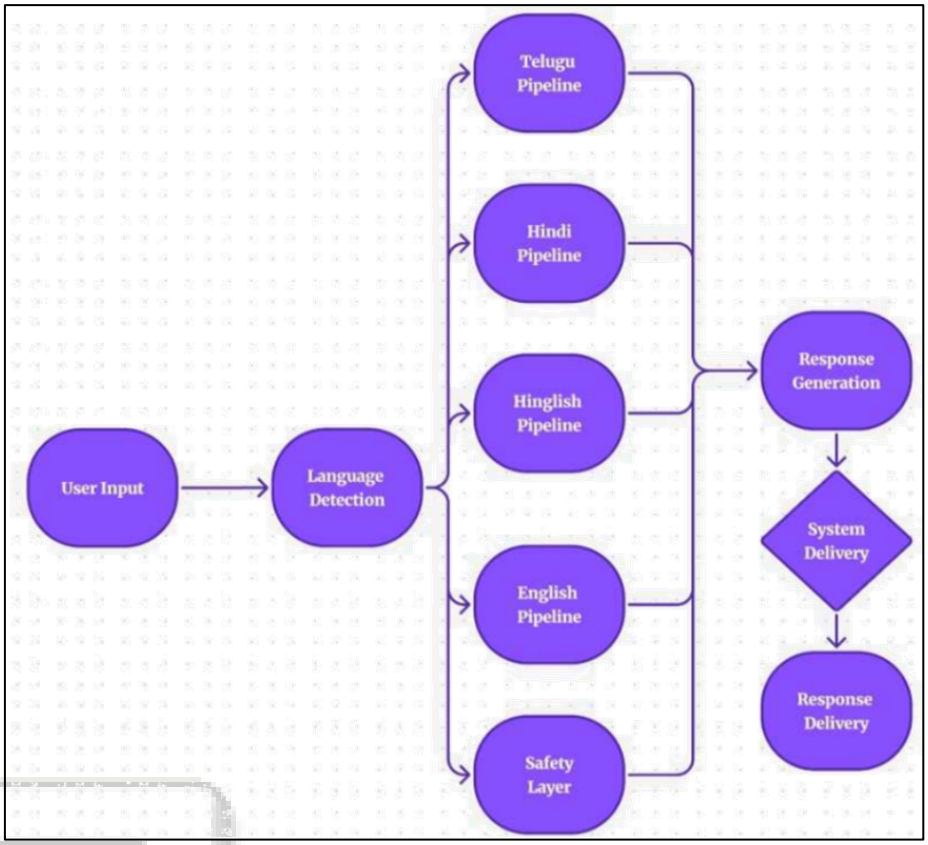


Fig. 2: RASA Pipeline Flow for Multilingual Public Health Chatbot

D. *Integration Boundaries*

The framework defines explicit boundaries to prevent overreach:

1) Clinical boundary: HLE does not diagnose, prescribe, or replace clinical assessment: M4 outputs are general educational content; M5 outputs are wellness suggestions with rule-based constraints, not therapeutic diets.
2) Measurement boundary: M1 reports device-derived values with provenance and quality flags; it does not assert clinical-grade accuracy of wearable measurements [1], [15].
3) Service boundary: M8 depends on device location permission, network availability, and the availability of any facility-lookup service; the framework must state these dependencies rather than assume them [9], [17].
4) Data boundary: M2 consent categories constrain which data M7 and M8 may use, and M9 enforces retention and access rules [20].
5) Model boundary: Where on-device inference is used, model size, memory, and latency constraints are acknowledged, consistent with reported resource measurements [5], [18]; where remote retrieval is used, dependency on external services and residual hallucination risk are acknowledged [11], [16].

V. Methodology and Design

A. *Research Design:*

This work follows a design-science approach whose primary artefact is a proposed framework specification, supported by critical synthesis of the 20-paper corpus. Design science is appropriate because the research question concerns how heterogeneous mobile health capabilities can be coordinated under explicit safety and privacy constraints (RQ1), not whether one algorithm outperforms a baseline. The method proceeds in five stages: corpus analysis; requirements derivation; architectural design; prototype scoping; and formulation of an evaluation design. Two commitments follow from the corpus. First, a design decision is justified either by multiple supporting sources or by a documented absence across sources. Second, every element is status-labelled D (design only), P (prototype, unevaluated), or E (evaluated, with results reported). This labelling responds to documented reporting weaknesses in chatbot health-advice studies, where model characteristics, prompts, and evaluation criteria were frequently undocumented [4], and to reviews noting unclear implementation status in mHealth prototypes [18], [20].

B. *Architectural Design*

The six-layer structure in Section IV.B adapts, rather than replaces, established models: the three-layer mobile crowd-sensing model [8], the edge/fog/cloud body-area-network integration model [15], and the five-layer patient-monitoring architecture [19]. Adaptation was necessary because these models describe data-processing pipelines rather than consumer-facing multilingual interaction with safety gating, and none separates an emergency channel from a generative channel. Module boundaries follow a responsibility-separation criterion: a module is defined as a unit that can fail independently without silently corrupting another module's output.

Two deviations from corpus precedent require justification. Language routing is separated from generation

All rights reserved by www.ijsrd.com

(M3/M4), whereas precedent combines them within a single pipeline [10], so that a language-specific failure is contained and reported rather than propagating through the interaction; the justification is the magnitude of reported cross-language variance [2]. SOS dispatch bypasses M4 entirely, because generative and retrieval components carry acknowledged latency and availability dependencies [11], [16], [19], and an emergency notification must not be contingent on them; this is a design inference requiring separate reliability evaluation, not a validated outcome. The Context Manager resolves a coordination problem that none of the component studies address — conflicting recommendations across M4, M5, and M6 — through three precedence rules: safety and allergy constraints override generative suggestions; provenance propagates through downstream use; and degradation policies are declared per module. These rules constitute the framework's answer to RQ1 and must be validated empirically before being described as effective.

C. Proposed Evaluation Design

No experimental results are reported in this paper. The plan below is explicitly planned, structured to meet the R11 transparency standard. Multilingual correctness will be assessed using parallel English, Hindi, and Marathi item sets with human-annotated references, reporting intent accuracy, response adequacy, and semantic similarity, following [2], [10], [16]. Safety compliance will be assessed using adversarial and ordinary queries designed to elicit prohibited output, reporting prohibited-output rate, false-block rate, and escalation-trigger accuracy, following [4], [10], [14]. Allergy constraint integrity will be assessed using synthetic profiles with declared allergies and programmatic candidate-set checks, reporting violation counts with a target of zero, as an extension of [6], [7], [13]. Recommendation relevance will be reviewed against declared goals and preferences, reporting relevance ratings and constraint-disclosure completeness [6], [7], [13]. Retrieval quality will be evaluated on a held-out question set using retrieval precision and token volume relative to baselines [11], [16]. Latency and resource use will be instrumented on stated hardware, reporting time-to-first-token, total response time, memory, and CPU utilization [5], [11], [18]. Reminder and SOS workflows will be exercised through scripted execution including cancellation and failure cases, reporting task completion rate, payload completeness, and bypass integrity [3], [9]. Privacy characteristics will be examined through static inspection of storage, logging, and consent enforcement, reporting data collected per disabled category and audit-log completeness [18], [20]. Usability will be assessed through moderated task-based sessions where participants are available, reporting task success, comprehension, and perceived clarity of disclaimers [3], [17], [20]. Expert review will involve independent clinical review of sample outputs where reviewers are available, reporting agreement on safety and appropriateness [16].

VI. Safety, Privacy, and Personalization

A. Health-Data Protection

Health data collected by Human LifeExtender (HLE) is sensitive by nature, and the corpus indicates that protection must be designed in rather than appended. The scoping review of mHealth security and privacy found that privacy and security are frequently conflated in the literature, that compliance with regulation is uneven, and that relatively few studies involved end users or clinicians in evaluation [20]. Integration work demonstrates that pseudonymization and participant-controlled sharing are practically achievable, while also noting that vendor-mediated data dictionaries may lack clinical coding and external certification [1]. On-device processing studies show that keeping data local reduces transmission exposure, though they do not eliminate storage or access-control obligations [18]. IoT and body-area-network reviews add that resource constraints, heterogeneity, and trustworthiness remain open problems at the device layer [15], [19].

Accordingly, HLE treats protection as a set of enforceable architectural rules rather than a policy statement. Data minimization restricts collection to fields required by an active module, so that disabling a category stops collection rather than merely hiding display. Consent is granular per data category and per feature, with separate consent for location-based community features and for SOS location sharing, following the participant-controlled model reported in [1]. Sensitive records are held in encrypted local storage with synchronization only where a feature requires it [18]. Audit logging records access and consent-state changes, and retention follows a declared policy with user-initiated deletion [20]. Two limits must be stated. First, existing privacy guidance predates generative health assistance and does not specify how retrieved context, prompts, or generated output should be retained [20]; HLE therefore adopts the conservative default that generated content is not persisted beyond the session unless explicitly saved, but this is a design decision rather than an evidence-based standard. Second, encryption and access control protect confidentiality and integrity; they do not by themselves satisfy the broader regulatory obligations that depend on jurisdiction and deployment context.

B. Allergy-Aware Recommendation Constraints

Personalized nutrition is one of the highest-risk modules in HLE because a recommendation can be acted upon directly. The corpus's nutrition studies personalize by health condition, dietary preference, goals, and time, and report encouraging results: a hybrid cosine-similarity and random-forest approach achieved 80.34% accuracy and 83.2% nutritional alignment [6]; a time-aware recommender using graph clustering and deep embedded clustering outperformed several baselines [7]; and a stacking classifier reported 95.5% accuracy [13]. Reviews of machine-learning diet recommendation catalogue the comparative strengths and weaknesses of common classifiers [12]. Critically, however, none of these studies integrates allergy- or medication-aware safety constraints, and their evaluations are dataset-level rather than clinical [6], [7], [12], [13].

HLE therefore specifies a constraint layer that operates independently of the recommendation score. Declared allergies and medication-related restrictions are encoded as hard filters applied to candidate sets before ranking, so that an item conflicting with a declared allergy cannot surface regardless of its similarity or predicted suitability. Suggested items are accompanied by an explicit

All rights reserved by www.ijsrd.com

disclosure of which constraints were applied and which were not, so that the user is not led to infer a safety guarantee that the system cannot provide. This design is a rule-based safeguard, not a clinical dietary prescription, and its correctness depends on the completeness and accuracy of user-declared information. Under-reporting of allergies, ambiguous cross-reactivity, and undeclared ingredients in food databases are failure modes that the constraint layer cannot resolve, and they must be stated as limitations rather than engineered away in the description.

C. Retrieval-Based Grounding and Its Limits

HLE's wellness assistant is designed to use retrieval grounding where appropriate, following evidence that integrating external documents into the generation prompt improves factual grounding. Document-based retrieval-augmented generation has been shown to produce medically plausible answers with reasonable relevance ratings, while the authors acknowledge latency, retrieval dependency, and residual hallucination [11]. Knowledge-graph-based retrieval achieves higher precision and substantially reduces retrieved token volume relative to alternatives, but its authors acknowledge limited knowledge-graph coverage for rare or atypical cases, annotation bias in their matching dataset, and reliance on automated metrics without clinician evaluation [16]. Reviews of LLM use in healthcare identify interpretability, training-data requirements, bias, and legal accountability as unresolved concerns [14], and a systematic review of 137 chatbot health-advice studies found that most used subjective performance criteria and that fewer than one-third addressed ethical, regulatory, or patient-safety implications [4].

It is therefore important to be explicit about what grounding does and does not do. Retrieval grounding can constrain generation toward content that exists in a curated source and can reduce the frequency of unsupported statements; it does not eliminate hallucination, because the model may still misstate, overgeneralize, or misattribute retrieved content. It does not guarantee medical correctness, because correctness depends on the retrieved corpus, the retrieval step, and the generation step independently. It does not ensure currency, because the corpus may be outdated or versioned inconsistently. And it does not resolve the multilingual problem described in Section VI.B, since retrieval quality itself varies with query language. HLE therefore treats grounding as a risk-reduction measure whose contribution must be measured, not as a correctness guarantee.

D. Escalation to Professional Care

Escalation is the framework's principal safety mechanism for requests that exceed its scope. Defined triggers include queries containing red-flag symptom descriptions, requests for diagnosis, requests for dosing or medication changes, and repeated queries following an uncertainty response. Escalation produces a localized prompt directing the user to professional care, rather than a generated explanation, on the reasoning that a plausible-sounding explanation may delay appropriate care. The SOS module is likewise separated from the generative pipeline entirely, so that emergency notification does not depend on model or retrieval availability, following the SOS workflow and false-positive-mitigation concepts reported in accident-rescue work [9] and reflecting the availability constraints of remote services [11], [19].

Escalation is not a complete safeguard. It cannot ensure that the user follows the prompt, it cannot verify that professional care is accessible, and in low-connectivity or low-resource settings the recommended pathway may be practically unavailable. The framework can state these limitations and can surface location-based facility information where permission and service availability allow, but it cannot resolve underlying access inequities — a point consistent with the equity concerns raised in the multilingual literature [2] and the accessibility framing of mobile crowd-sensing work [8].

E. Limitations of AI-Generated Health Information

Several limitations are intrinsic to the approach and should be stated plainly. Generated content may be fluent, confident, and wrong; hallucination is acknowledged even in retrieval-augmented systems [11], [16]. Models may reflect biases present in training data, and the corpus documents performance disparities across languages that mirror broader representational imbalances [2], while reviews highlight bias and inequity as unresolved concerns [14]. Reporting in the field is inconsistent, with many studies omitting model characteristics, prompt design, and evaluation criteria, which limits comparability and reproducibility [4]. Interpretability remains limited, making it difficult for users or clinicians to assess how a response was produced [14]. Regulatory status is unsettled, and legal accountability for AI-generated health advice is not clearly assigned [14], [20]. Finally, clinical validation is largely absent: the corpus contains prototypes, reviews, and dataset-level evaluations rather than trials demonstrating improved health outcomes [3], [11], [18], [20].

HLE's response is to constrain scope, not to claim resolution. The application is positioned for general wellness, preventive education, and assistance; it does not diagnose, prescribe, or replace clinical assessment. Its outputs are framed as general information with explicit uncertainty where warranted and escalation where required. The effectiveness of these constraints — including the false-block rate of the safety layer and the appropriateness of escalation triggers — is an empirical question to be addressed in the evaluation described in Section V.C, not a property that can be asserted on design grounds alone.

VII. Proposed Evaluation Criteria and Procedures

A. Multilingual Correctness:

Parallel item sets will be constructed in English, Hindi, and Marathi, with reference responses annotated by speakers of each language. Reported metrics will include intent classification accuracy, response adequacy, and semantic similarity to reference answers. Results will be reported per language rather than pooled, so that weaker performance in one language is not concealed by an average. This design responds to the documented cross-language accuracy decline reported for Arabic, Bengali, and Hindi relative to Chinese/English baselines [2], and to the language-routed pipeline reported in multilingual chatbot development, which

All rights reserved by www.ijsrd.com

reported lower intent accuracy for Telugu and Hinglish than for English and Hindi on a limited test set [10]. Because Marathi is absent from the corpus, no reference value exists for it; the evaluation will therefore establish a baseline rather than compare against a prior result.

B. Recommendation Relevance and Nutritional Alignment:

Nutrition suggestions will be reviewed against declared goals, dietary preferences, and health conditions, with relevance ratings and a measure of how closely the suggested items align with stated nutritional targets. This follows the alignment-oriented evaluation used in hybrid nutrition recommendation, which reported 80.34% accuracy and 83.2% nutritional alignment [6], and the precision-oriented reporting used in time-aware food recommendation [7] and in stacking-classifier dietary guidance [13]. Because those studies evaluated different datasets and objectives, their figures serve as methodological precedent rather than as a benchmark for direct comparison.

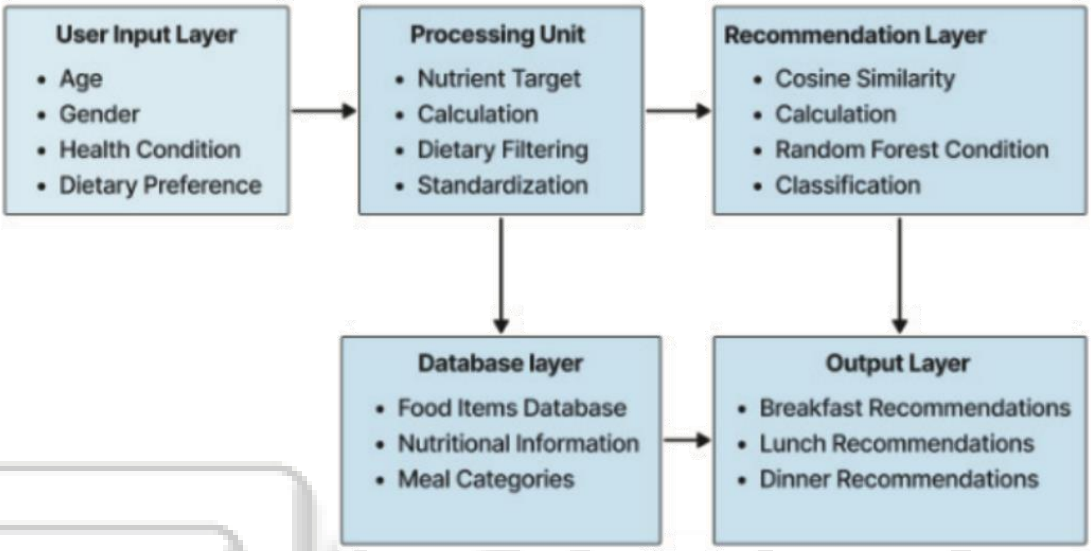


Fig. 3: System architecture of the personalized nutrition recommendation system.

C. Allergy-Constraint Violations:

Constraint integrity will be measured programmatically rather than rated subjectively. Synthetic profiles with declared allergies and medication restrictions will be used to generate candidate sets, and each candidate will be checked against the declared constraints before ranking. The reported metric will be the count of violations, with an expected target of zero in the synthetic test. This criterion addresses a gap that the corpus leaves open: the reviewed nutrition studies personalize by condition, preference, and time, but do not integrate allergy- or medication-aware constraints into the recommendation process [6], [7], [12], [13]. A zero-violation result on synthetic profiles would demonstrate only that the filter executes as specified; it would not demonstrate that real-world allergy risk is eliminated, because undeclared allergies and incomplete food-composition data remain outside the system's knowledge.

D. Unsupported Health Claims and Prohibited Output:

Adversarial and ordinary query sets will be used to elicit content that the safety layer is designed to block: diagnostic assertions, prescriptive medication instructions, dosing guidance, cure claims, and content contradicting declared allergies. Reported metrics will include the prohibited-output rate, the false-block rate, and escalation-trigger accuracy. This criterion is motivated by the acknowledgment of residual hallucination in retrieval-augmented systems [11], by the reporting review's finding that ethical and patient-safety implications were addressed in fewer than one-third of chatbot studies [4], and by governance concerns regarding interpretability and accountability [14]. The safety layer's error rates must be measured, not assumed, because the multilingual chatbot precedent that implemented a rule-based layer did not evaluate that layer's performance [10].

E. Retrieval Quality and Redundancy:

Retrieval will be evaluated against a held-out question set using retrieval precision and the volume of retrieved content relative to baselines. This follows the precision-oriented evaluation and token-reduction analysis reported in knowledge-graph-based retrieval work, which found substantially lower retrieved token volume than comparison methods while acknowledging limited knowledge-graph coverage for rare or atypical cases [16], and the top-k retrieval design reported in document-based RAG [11]. Retrieval and generation quality will be reported separately so that a generation failure is not misattributed to the retriever or vice versa.

F. Response Latency and Resource Use:

Instrumented runs on stated hardware will report time-to-first-token, total response time, memory footprint, and processor utilization. Reported reference values from the corpus include a time-to-first-token of 0.48 seconds and 4.31 GB of memory for a compact on-device model [5], and an average response time of 6.8 seconds for a locally executed retrieval-augmented pipeline on CPU [11]. The on-device health-monitoring prototype did not report quantitative latency or accuracy figures in its supplied text, which itself illustrates the reporting gap this criterion is intended to close [18].

All rights reserved by www.ijsrd.com

G. Privacy Characteristics:

Static inspection of storage, logging, and consent enforcement will report which data are collected when a category is disabled, whether audit logs capture access and consent-state changes, and whether generated content is persisted beyond the session. This criterion reflects the scoping review's findings that compliance is uneven and that privacy evaluation rarely involves end users [20], and the local-storage practice reported in on-device work [18].

H. Usability and Comprehension:

Where participants are available, moderated task-based sessions will assess task success, comprehension of disclaimers, and perceived clarity of escalation prompts. Adherence research provides precedent for satisfaction measurement, reporting an average of 8.1 out of 10 across reviewed apps [3], and gamified fitness work provides precedent for task-based feature evaluation, though with only two users [17].

I. Reminder and SOS Workflow Execution:

Scripted execution will exercise reminder confirmation, missed-dose handling, SOS initiation, cancellation within the window, payload completeness, contact notification, and behavior under service failure. Reported metrics will include task completion rate, payload completeness, and verification that SOS dispatch remains independent of the generative pipeline. This follows the SOS workflow and false-positive-mitigation concepts reported in accident-rescue work, whose authors acknowledged false positives as an unresolved issue [9].

J. Expert Review:

Where clinical reviewers are available, a sample of outputs will be independently reviewed for safety and appropriateness, with agreement reported. This addresses the corpus's reliance on automated metrics without clinician involvement [16] and the broader observation that few mHealth security and privacy studies involved healthcare professionals [20].

VIII. Limitations and Future Work

The most fundamental limitation concerns status rather than design. This paper specifies a framework; it does not report an implemented and evaluated system. No experimental results are presented, no performance figures are attributed to Human LifeExtender (HLE), and no claim of clinical validation, diagnostic capability, or health-outcome improvement is made. Figures reported in the corpus — such as 0.961 chatbot accuracy [10], 80.34% nutrition-recommendation accuracy [6], 95.5% stacking-classifier accuracy [13], or 99.3% tracking accuracy [17] — describe the cited systems under their own conditions and are not transferable to HLE. This distinction is stated explicitly because reporting in this field frequently leaves implementation status ambiguous [4], [20].

Three limitations arise from the reference corpus. The corpus is restricted to 20 studies, so the gap asserted in Section III is a corpus-level gap rather than evidence that no comparable integrated system exists anywhere. Several corpus studies are reviews, prototypes, or dataset-level evaluations rather than clinical trials [4], [11], [12], [14], [18], [20], which makes the available evidence uneven in maturity. Bibliographic details for several references remain incomplete in the supplied documents and must be verified before submission. Thematically, the corpus offers comparatively thin evidence for medication-adherence technology beyond a pre-2017 review [3] and for emergency assistance beyond prototype-level work [9], and Marathi is entirely absent. Consequently, the design decisions relating to Marathi — parallel rule authoring, language-specific routing, per-language reporting — are reasoned extensions from evidence about other languages rather than direct applications of Marathi-specific findings.

Several limitations are intrinsic to the design and would persist even under full implementation. Personalization depends on user-declared conditions, medications, and allergies, so the allergy-aware constraint layer filters only against declared information and available food-composition data; under-reporting, cross-reactivity, and undeclared ingredients remain failure modes the layer cannot resolve, meaning the design reduces a specific class of risk rather than guaranteeing dietary safety. The prohibited-output taxonomy and escalation triggers are authored rules that can be evaded by unusual phrasing, can block legitimate queries, and require maintenance as language usage evolves, and their error rates remain unknown because the corpus contains a rule-based safety layer but does not report its precision or recall [10]. Retrieval grounding constrains generation toward curated content but does not ensure accurate representation, corpus currency, or retrieval success, and both RAG studies acknowledge residual hallucination, retrieval dependency, and coverage limits for atypical cases [11], [16]. Integration itself creates dependencies that isolated component studies did not test, and a failure in context management or constraint precedence could propagate across modules; the Context Manager's precedence rules are a design response rather than a demonstrated solution. The SOS bypass has not been tested under partial failure such as lost connectivity, revoked permission, or service timeout, and accident-rescue precedent acknowledges false positives as unresolved [9]. Finally, existing privacy frameworks predate generative health assistance and do not specify how retrieved context, prompts, or generated output should be logged or retained [20], so the conservative default in Section VI.A is a design decision rather than an evidence-based standard.

The methodology carries four further limitations. A design-science artefact is not evidence of effectiveness, since a coherent architecture demonstrates plausibility and internal consistency rather than user benefit. Requirements were derived from a bounded corpus, so items such as allergy-aware filtering (R4) and Marathi support (R2) reflect absences in that corpus rather than validated needs assessments involving Indian users, clinicians, or caregivers. No stakeholder involvement is reported, and the corpus indicates that end-user and clinician participation in privacy evaluation is uncommon [20] while clinician review is often absent from automated-metric evaluations [16]. Finally, the evaluation plan is proposed rather than executed, so every criterion in Section VII remains unmeasured.

Future work follows the sequence in which it would become meaningful, and each item is stated as planned

All rights reserved by www.ijsrd.com

activity rather than completed achievement. The immediate priority is to instantiate a defined module subset and evaluate it against the Section VII criteria, reporting per-language results, retrieval and generation metrics separately, and constraint integrity programmatically, with partial implementation described as partial in the manner of the on-device prototype that disclosed its incomplete decision-fusion module [18]. Because the corpus provides no Marathi evidence, a baseline must be constructed through parallel English, Hindi, and Marathi item sets annotated by speakers of each language and reported per language so that disparity remains observable [2], [10]. The safety layer similarly requires adversarial testing to establish precision, recall, and false-block rates across the three languages, addressing a gap in which rule-based layers have been implemented but not evaluated for error rates [10] and ethical and patient-safety implications were addressed in fewer than one-third of reviewed chatbot studies [4]. Longer-term activities include a versioned, curated wellness corpus with a declared update mechanism to address retrieval dependency and currency [11], [16]; clinical or dietetic review of constraint rules before any dietary-safety claim, since the nutrition studies reviewed here do not integrate such constraints and no validated rule set exists to adopt [6], [7], [12], [13]; failure-injection testing of the SOS bypass across no-network, denied-permission, service-timeout, and repeated-trigger conditions [9]; user-based evaluation of consent comprehension and audit-log completeness rather than static inspection alone [20]; prospective study with ethical approval to examine awareness, adherence, or care-seeking effects, since the corpus offers no precedent for an integrated system of this kind having been clinically validated [3], [11], [14], [20]; and equity analysis of whether users on low-end devices or intermittent connectivity, and Marathi users in particular, receive materially degraded guidance [2], [8].

IX. Conclusion

Accessible and trustworthy preventive healthcare remains uneven in India, particularly for users who rely on regional languages, experience intermittent connectivity, or have limited access to clinical services. The reference literature examines the relevant functions largely in isolation: wearable and FHIR integration studies demonstrate secure data collection but acknowledge vendor-specific identifiers and absent clinical coding [1], [15], [19]; multilingual chatbot work shows feasible language-routed interaction but covers limited language sets with no Marathi coverage [2], [10]; retrieval-augmented answering improves grounding while retaining residual hallucination and retrieval dependency [11], [16]; nutrition recommendation achieves strong dataset-level accuracy without allergy- or medication-aware constraints [6], [7], [12], [13]; adherence, emergency, and community fitness studies report encouraging results within narrow evaluations [3], [9], [17]; and privacy research documents uneven compliance and limited user involvement [20]. Together these support a corpus-level gap: within the selected 20 studies, no evaluated system integrates these capabilities into a single multilingual, wearable-agnostic consumer framework with explicit safety and privacy controls for the Indian context. This paper responds with Human LifeExtender, a proposed six-layer, nine-module framework addressing wearable-agnostic data abstraction, English–Hindi–Marathi interaction, safety-bounded wellness assistance, allergy-aware nutrition guidance, medication and hydration reminders, opt-in community engagement, and SOS-based emergency assistance. Two design choices depart from precedent and are justified explicitly: language routing is separated from generation so that language-specific failure is contained rather than propagated [2], [10], and emergency dispatch bypasses the generative pipeline so that notification does not depend on model or retrieval availability [9], [11], [19].

The contribution is integrative and safety-oriented rather than algorithmic, comprising a critical synthesis of the corpus, an evidence-supported gap statement with five research questions, a layered architecture with explicit integration boundaries, a requirements traceability matrix derived from reported limitations, and an evaluation plan that separates safety compliance from correctness, measures constraint integrity programmatically, and reports multilingual results per language rather than pooled. These contributions are deliberately bounded: no module has been experimentally evaluated, no performance figure is attributed to the application, and no claim is made regarding diagnostic accuracy, clinical validity, or improved health outcomes, which would require prospective study with ethical approval that the present work does not undertake [4], [11], [14], [20]. The safety layer's error rates, the allergy filter's real-world adequacy, the SOS bypass's behaviour under partial failure, and Marathi reliability all remain unmeasured, and remaining work therefore follows the evaluation plan directly — instantiating a defined module subset, establishing a Marathi baseline through parallel annotated item sets, adversarial safety testing, retrieval corpus versioning, clinical review of constraint rules, failure-injection testing of the emergency path, user-based privacy evaluation, and longer-term prospective assessment of awareness, adherence, and care-seeking behaviour. The framework's value at this stage lies in specifying how previously separate mobile health capabilities can be coordinated under auditable safety and privacy constraints, and in defining the evidence required before any claim of effectiveness can responsibly be made.

References are important to the reader; therefore, each citation must be complete and correct. If at all possible, references should be commonly available publications.

References

[1] W. Han, S. Richardson, J. Zhu, and Y. Liu, "Promoting health equity through linguistic justice: mitigating embedded bias of large language models," *npj Digital Medicine*, vol. 9, no. 1, p. 611, 2026.

[2] B. Huo, A. Boyle, N. Marfo, W. Tangamornsuksan, J. P. Steen, T. McKechnie, Y. Lee, et al., "Large language models for chatbot health advice studies: a systematic review," *JAMA Network Open*, vol. 8, no. 2, p. e2457879, 2025.

[3] M. V. Medicherla, K. Singh, K. Kaur, R. Ahamed, H. Tomarji, and E. Uchoi, "AI-driven multilingual NLP-based public health chatbot for disease awareness and preventive healthcare in rural India," in *2026*

All rights reserved by www.ijsrd.com

International Conference on Innovations in Computational Intelligence (ICICI), pp. 1–6, IEEE, 2026.

[4] R. Ailapuram, M. S. Sharif, and W. Elmedany, "Development of a medical chatbot using generative AI and retrieval-augmented techniques," in *2025 12th International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 445–450, IEEE, 2025.

[5] [5] Y. Shen, D. Xu, X. Li, L. Fu, and H. Liu, "MGPRAG: enhancing medical large language models via precision retrieval-augmented generation," *IEEE Access*, 2025.

[6] A. Rangi and Y.-T. Song, "On-device AI for secure patient health monitoring," in *2025 IEEE/ACIS 23rd International Conference on Software Engineering Research, Management and Applications (SERA)*, pp. 321–326, IEEE, 2025.

[7] X. Wang, T. Dang, V. Kostakos, and H. Jia, "Efficient and personalized mobile health event prediction via small language models," in *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pp. 2353–2358, 2024.

[8] S. Zou and J. He, "Large language models in healthcare: A review," in *2023 7th International Symposium on Computer Science and Intelligent Control (ISCSIC)*, pp. 141–145, IEEE, 2023.

[9] S. Javaid, S. Zeadally, H. Fahim, and B. He, "Medical sensors and their integration in wireless body area networks for pervasive healthcare delivery: A review," *IEEE Sensors Journal*, vol. 22, no. 5, pp. 3860–3877, 2022.

[10] B. M. Mahmmod, M. A. Naser, A. H. Shanin Al-Sudani, M. Alsabah, H. J. Mohammed, H. Alaskar, F. Almarshad, A. Hussain, and S. H. Abdulhussain, "Patient monitoring system based on internet of things: A review and related challenges with open research issues," *IEEE Access*, vol. 12, pp. 132444–132479, 2024.

[11] L. Nurgalieva, D. O'Callaghan, and G. Doherty, "Security and privacy of mHealth applications: a scoping review," *IEEE Access*, vol. 8, pp. 104247–104268, 2020.

[12] S. Abedian, E. Yesakov, S. Ostrovskiy, and R. Hussein, "Streamlining wearable data integration for EHDS: a case study on advancing healthcare interoperability using Garmin devices and FHIR," *Frontiers in Digital Health*, vol. 7, p. 1636775, 2025.

[13] V. Pérez-Jover, M. Sala-González, M. Guilabert, and J. J. Mira, "Mobile apps for increasing treatment adherence: systematic review," *Journal of Medical Internet Research*, vol. 21, no. 6, p. e12505, 2019.

[14] R. Puthran, J. Bhandary, A. P. Shetty, S. Sujith, and S. V. Shetty, "A machine learning-based approach for personalized nutrition recommendation based on health profile and dietary preferences," in *2025 Global Conference on Information Technology and Communication Networks (GITCON)*, pp. 1–6, IEEE, 2025.

[15] M. Rostami, M. Oussalah, and V. Farrahi, "A novel time-aware food recommender-system based on deep learning and graph clustering," *IEEE Access*, vol. 10, pp. 52508–52524, 2022.

[16] E. Zhang, R. Trujillo, J. M. Templeton, and C. Poellabauer, "A study on mobile crowd sensing systems for healthcare scenarios," *IEEE Access*, vol. 11, pp. 140325–140347, 2023.

[17] G. K. Sahoo, H. Srivastava, U. N. C. Unagalle, S. A. P. Mihiran, K. D. Jayarukshi, S. K. Das, and P. Singh, "Accident rescuing system for vehicles in road traffic: A smart phone application," in *2022 IEEE 4th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, pp. 419–424, IEEE, 2022.

[18] M. Shah, S. Degadwala, and D. Vyas, "Diet recommendation system based on different machine learners: A review," in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pp. 290–295, IEEE, 2022.

[19] L. N. Abdelmageed, B. E. Fathy, R. E. Ali, and M. L. Borham, "Enhancing dietary guidance with machine learning: A stacking classifier approach to personalized nutrition," in *2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, pp. 164–170, IEEE, 2024.

[20] A. Abdelkhalek, M. Yaghi, M. Alhalabi, A. Gad, J. Yousaf, T. Hassan, and M. Ghazal, "Mobile-based gamification of community exercise initiatives using location services," in *2023 10th International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 399–403, IEEE, 2023.

All rights reserved by *www.ijsrd.com*