Glosains: Jurnal Sains Global Indonesia
Volume 7, Issue 2, 286-305
e_ISSN: 2798-4311
https://glosains.staiku.ac.id
DOI: doi.org/10.59784/glosains.v7i2.670

SiKurang: Development and Field Evaluation of an Offline-First mHealth System for Community-Based Stunting Risk Detection and Counselling in Low-Connectivity Primary Care Settings

Muhammad Rozahi Istambul[1*]
Universitas Widyatama, Indonesia

Parlindungan[2]
Universitas Widyatama, Indonesia

Jhon Henry Wijaya[3]
Universitas Widyatama, Indonesia

Reza Zezarina [4]
Universitas Widyatama, Indonesia

Dery Fachrizsal [5]
Universitas Widyatama, Indonesia

***Corresponding author:**
Muhammad Rozahi Istambul, Universitas Widyatama, Indonesia.
✉rozahi.istambul@widyatama.ac.id

Article Info :

Article history:
Received: February 26th, 2026
Revised: March 30th, 2026
Accepted: April 1nd, 2026

Keywords:
stunting detection; offline Mhealth; edge AI; community health cadres; posyandu.

Abstract

Background: Despite growing interest in mHealth solutions for community nutrition, no prior study has evaluated an integrated offline system combining on-device machine learning, NLP-based counseling, and geospatial visualization for non-specialist cadres within Indonesia's *Posyandu* network the gap this study addresses.
Objective: We aimed to evaluate the feasibility, accuracy, usability, and early effects of an offline-first Android application (SiKurang) for stunting risk assessment, counseling, and geospatial visualization.
Methods: We conducted a convergent mixed-methods, single-arm pre–post pilot at two *Posyandu* over 12 weeks, involving mothers/caregivers, community health cadres, and nutritionists. Outcomes included AUROC for on-device risk scoring, System Usability Scale (SUS), User Experience Questionnaire (UEQ-S), caregiver knowledge, administrative burden, and targeted home visits; interviews were thematically analyzed.
Results: On-device risk scoring achieved AUROC 0.87; usability was high (SUS 84.2; UEQ-S 1.86). Caregiver knowledge improved markedly (Cohen's d = 1.28). Risk maps supported a 22% increase in targeted home visits. The app operated reliably offline and synchronized upon connectivity, reducing administrative workload, with no major cultural or usability barriers reported.
Conclusion: The application was feasible and acceptable in primary care, enabling timely, data-informed counseling and referral in low-connectivity environments. This study provided field evidence for an offline-first, low-cost mHealth model delivering on-device analytics and geovisualization for non-specialist cadres, offering a scalable template for strengthening maternal–child health at the last mile. Scientifically, this study contributes the first field-validated, multi-component offline mHealth framework for community-level stunting surveillance in a low-resource LMIC setting.

To cite this article: Istambul, M. R., Parlindungan, Henry, J., & Reza, D. (2026). SiKurang: Development and Field Evaluation of an Offline-First mHealth System for Community-Based Stunting Risk Detection and Counselling in Low-Connectivity Primary Care Settings. *Glosains: Jurnal Sains Global Indonesia*, *7*(2), 286-305. https://doi.org/10.59784/glosains.v7i2.670

INTRODUCTION

Stunting, defined as a height-for-age z-score below −2, remains a critical public health challenge affecting 149 million children under five globally, with South-East Asia and sub-Saharan Africa bearing the highest burdens (Organization, 2023). In Indonesia, despite progress, 21.6% of children remain affected, with rural areas exceeding 30% prevalence. Like many low- and middleincome countries (LMICs), Indonesia relies on community-based primary care platforms such as Posyandu to deliver early nutrition interventions during the first 1,000 days of life. Yet these systems are limited by paper records, time-lagged data interpretation, and inefficient referral methods that result in missed opportunities for preventive action. Similar barriers exist in other regions throughout sub-Saharan Africa, where community health workers (CHWs) experience similar constraints in identifying at-risk children and delivering timely counseling.

A set of low-tech, reliable digital tools that align with conventional primary care processes to empower cadres and deliver services more equitably are urgently needed. Stunting has been a significant contributor to childhood morbidity and mortality (Haroun et al., 2022; Huang et al., 2024). In theory, this challenge illustrates a well-characterized implementation gap in health systems science: just because effective nutritional interventions exist does not guarantee their adoption at the community level when the infrastructure for information delivery is lacking. To address stunting at scale, therefore, requires not just clinical knowledge but information systems that can operationalize such knowledge at the point of care, particularly in the resource-constrained last mile where traditional health infrastructure is weakest.

Indonesia's national nutrition strategy relies extensively on the work of volunteer community health cadres who operate at the Posyandu community-integrated health post networks that form an entry point to primary care for maternal and child health. These cadres collect anthropometric data and monitor growth, referring high-risk cases to the health facility, but their effectiveness is compromised by manual processes that are subject to errors, loss of records, and delayed follow-up. In 2020, an audit showed that 38% of stunted children and fewer than 1 in 5 eligible families were not identified to receive home visits.

Such systemic weaknesses parallel those seen in Ethiopia, Malawi, and Nigeria sub-Saharan African countries with CHW-led programs that are hindered by the inconsistent use of data and weak feedback loops. While digital health tools have been developed to help fill such gaps, most depend on persistent internet connectivity, complex interfaces, or centralized infrastructure thus rendering them inapplicable in remote, low-connectivity environments.

However, while some promise of digital health solutions to improve data accuracy and timeliness in LMICs has been demonstrated, few are built for the practicalities of community-based primary care: limited connectivity, variable levels of digital literacy, and dependence on volunteer cadres. Many mHealth tools are online-dependent or not integrated into local workflows, making them hard to scale. Even fewer combine actionable risk prediction, counseling support, and spatial planning in the same offline platform accessible to non-technical users.

To address this gap, we developed SiKurang: a low-cost Android application that operates offline on budget smartphone devices and incorporates four main components: (i) QR-coded child identification to facilitate rapid retrieval of health records; (ii) automated anthropometric measurement screening for stunting risk using an embedded algorithm; (iii) nutrition counseling delivered conversationally in Bahasa Indonesia; and (iv) real-time mapping of high-risk clusters accessible to supervisors. Designed using human-centered approaches with cadres and mothers, the system is optimized for ease of use, reliability, and alignment with prevailing primary care practices. Developed collaboratively with community health workers, the system enhances primary care by minimizing administrative burden and improving referral accuracy, along with facilitating personalized counseling the essential pillars of effective maternal and child health interventions in resource-limited environments.

This approach prioritizes system integration and resource optimization over technical efficiency, providing lessons that are more broadly transferable to comparable contexts in sub-Saharan Africa and elsewhere. To develop and implement a digital solution that receives sustained use within the Posyandu setting, we identified three structural reasons for previous failures: firstly, connectivity dependency most mHealth tools rely on real-time internet access for data entry and risk calculation, which can render them dysfunctional during outages experienced by over 60% of Indonesia's remote Posyandu locations. secondly, usability mismatch—these systems are designed predominantly for trained health professionals rather than volunteer cadres with varying levels of digital literacy, resulting in high abandonment rates; and thirdly, functional fragmentation systems designed to capture only one aspect (e.g., data recording or referral) fail to address the totality of administrative burden, thus failing to represent sufficient disruption to existing workflows (Thomas et al., 2023).

While previous studies have offered valuable empirical insights when it comes to understanding the effectiveness of Indonesia's Village Fund (Dana Desa), they are still limited in their ability to account for its multidimensional lines of impact. Illustratively, Anu Rammohan (2023) demonstrate that the Village Fund is correlated with significant decreases in poverty in rural contexts within the country and increased labor force participation rates, especially among women; this improvement is related to a shift from value-added agricultural production into non-agricultural sectors as well as growth in household consumption. Yet, their approach relies predominantly on a macro-level Difference-in-Differences framework that evaluates only economic outcomes and does not consider spatial interdependence and institutional heterogeneity at the village level. Similarly, the most recent literature review indicates that research on Village Fund concerns is mostly limited to analysis of how this fund affects poverty and resource allocation, yet fails to consider complexities including mediation, moderation, or systems-level dynamics that reflect the complexity of governance and capacity dynamics (Anam et al., 2023).

To address these systemic gaps, there is increasing interest in integrating offline-capable digital tools that embed decision support, counseling, and spatial analytics directly into the community health workflow. Such tools must be simple, culturally appropriate, and usable by non-specialist cadres. To address this challenge, we created SiKurang, an offline-first mHealth solution aimed at improving early stunting identification at the point of care. In this paper, we assess its role, utility, and effectiveness in real-life Posyandu environments.

Though showing promise with respect to component-level evidence, there has been no prior investigation of the synergistic impact of QR-coded identity, offline machine learning risk scoring, natural language processing–oriented counseling, and spatial analytics on early stunting detection and user adoption in Indonesia's Posyandu network. Multi-dimensional mobile systems also present usability challenges: the system can become unusable when it is feature-rich and hard to navigate, and receiving many push notifications may discourage volunteers from participating (Thomas et al., 2023). Additionally, the heterogeneous level of digital literacy among cadres and caregivers required intensive human-centered design processes and mixed-methods assessments (Palumbo et al., 2022). Mobile AI applications must consider security issues such as model extraction and side-channel attacks Liu (2024), but in SiKurang, models are kept on-device and raw model weights are never exposed, thereby addressing this concern. This gap is especially meaningful, as combining these components into a single offline platform presents unique usability and adoption challenges that do not manifest in studies of single-component systems.

This study contributes three distinct scientific advances relative to previous work. First, it is the first study demonstrating validation of a fully integrated, offline-first mHealth system comprising four previously distinct capabilities QR-coded child identification, on-device gradient-boosting risk scoring, NLP-powered nutrition counseling, and GIS-based cluster visualization delivered in a single platform for deployment on low-cost Android smartphones without internet access. Second, unlike previous assessments focused on single-stakeholder usability or single-function performance, this study uses a convergent mixed-methods design involving three different stakeholder groups (cadres, Puskesmas staff, and mothers), affording a more ecologically valid assessment of system-level acceptability.

Third, the study provides a replicable edge-AI deployment framework for community health workers in LMICs, offering a methodological proof-of-concept of how offline machine learning can be useful in real-life contexts where cloud computing is not a feasible modality a modality that is increasingly being recognized as promising within the scope of digital health systems research Choudhary (2025) but so far offers little validation by way of nutrition surveillance at the field level.

With more than 30% of children under five being stunted in sub-Saharan Africa, community health workers (CHWs) are key to maternal and child health through home visits and growth monitoring. But, as in Indonesia, they often depend on paper registers, experience delayed data reporting, and lack decision-support tools. In Malawi, only 40% of high-risk children are correctly identified during routine visits in Nigeria, fewer than one-third of CHWs receive timely feedback from clinics (Thunberg et al., 2022). These systemic gaps are similar to those in Indonesia's Posyandu system, highlighting an urgent need for interoperable and offline-capable solutions to empower frontline workers without disrupting existing workflows.

This paper therefore aims to: (1) develop a modular, offline-first mobile application that integrates QR identity, on-device ML prediction, NLP chatbot counseling, and GIS risk visualization for Posyandu settings; (2) evaluate system performance in terms of predictive accuracy, usability, and user satisfaction across three stakeholder groups cadres, Puskesmas staff, and mothers; and (3) quantify the intervention's effect on stunting prevention knowledge and perceived decision-making efficiency. We hypothesize that the composite system will achieve an AUROC ≥ 0.85, a System Usability Scale (SUS) score ≥ 70, and a ≥ 15% increase in post-test knowledge scores. By meeting these thresholds, the study will demonstrate how simple, contextually grounded digital tools can strengthen primary care capacity at the last mile. This model may inform future implementations in similar resource-constrained settings, including those in sub-Saharan Africa.

METHOD

A. Study Design and Setting

A convergent mixed-methods design was employed to permit simultaneous quantitative and qualitative appraisal of the mobile health (mHealth) system, thereby enhancing interpretive power while offsetting mono-method bias. The quantitative strand comprised a single-arm, pre–post intervention trial undertaken between March and June 2024 across two Posyandu (integrated community health posts) situated in the peri-urban area of Bandung District, West Java, Indonesia. These sites were purposively selected because they (i) serve a combined catchment of >700 children aged 6–59 months, (ii) exhibit heterogeneous socio-economic profiles, and (iii) have 3G/4G network availability despite frequent connectivity interruptions a common challenge in digital health implementations across peri-urban Indonesia (Miranda et al., 2023; Rinawan et al., 2021).

Moreover, the heterogeneity of digital literacy among kader and caregivers necessitates rigorous human-centered design and mixed-methods evaluation. To prevent notification fatigue a known barrier to long-term engagement in mHealth programs the system uses context-triggered alerts only when risk thresholds are crossed, minimizing unnecessary interruptions (Keshavjee et al., 2022; Park et al., 2023).

B. Participants and Sampling

1. Target Participants

The study targeted three purposively selected stakeholder groups whose complementary roles in the Posyandu workflow collectively determine system effectiveness: (1) volunteer community health cadres (kader), who constitute the primary operational users responsible for data collection and first-line counseling; (2) Puskesmas nutritionists, who provide clinical supervision, interpret risk maps, and plan targeted home visits; and (3) mothers or primary caregivers of children aged 6–59 months, who represent the beneficiary group whose engagement and knowledge are the ultimate proximal outcomes of the intervention. This three-group design was theoretically grounded in UTAUT, which recognizes that technology acceptance is shaped by distinct role-based expectations and capabilities.

1) Volunteer *kader* (community health cadres)
2) Puskesmas (primary-care center) nutritionists, and
3) Mothers or primary caregivers of children aged 6–59 months.

2. Sampling Strategy

A convenience–purposive hybrid sampling strategy was adopted. All active kader (n = 10) and supervising Puskesmasstaff (n = 3) at the selected Posyandu were invited. For mothers, eligibility criteria were:

1) Possession of a maternal–child health booklet,
2) Residence within the catchment for ≥ 6 months, and
3) Ownership of an Android smartphone.

3. Sample Size Determination

Sample-size estimation for the quantitative arm followed the ISO 9241-11 recommendation for formative usability studies, requiring ≥ 15 participants per homogeneous cell to detect a mean System Usability Scale (SUS) difference of 10 points (SD ≈ 12) with 80% power and two-tailed α = 0.05 (Hyzy et al., 2022). Allowing for 10% attrition, we recruited 148 mothers, yielding a total analytic sample of 135 participants. Post-hoc verification using G Power confirmed that this sample could detect a small-to-medium partial eta-squared (η^2_p = 0.03) in a mixed-model repeated-measures ANOVA with ≥ 90% power.

4. Practical Significance

To quantify practical significance, we computed eta-squared (η^2) for every one-way ANOVA:

$$\eta^2 = \frac{SS_{between}}{SS_{total}}$$

where $SS_{between}$ is the sum-of-squares explained by the independent variable (role: kader, staff, mother) and SS_{total} is the total sum-of-squares of the dependent variable.

1. Satisfaction with Prediction Accuracy: η^2 = 0.0389 → 3.89% of the variance explained.
2. Overall System Impact: η^2 = 0.0138.
3. Ease of Use: η^2 = 0.0036.

All η^2 values remain below the 0.06 threshold for a medium effect, indicating that role explains only a trivial share of score variability.

5. Between-Group Effect Sizes

Between-group effect sizes were estimated with Cohen's d:

$$x = \frac{M_1 - M_2}{SD_{pooled}}, \quad SD_{pooled} = \sqrt{\frac{SD_1^2 + SD_2^2}{2}}$$

Results:

Overall System Impact: $M_{pregnant}$ = 4.667, M_{child} = 4.808, SD_{pooled} ≈ 0.45 → d = - 0.3489

Effectiveness Index: $M_{pregnant}$ = 79.15, M_{child} = 87.45, SD_{pooled} ≈ 19.0 s → d = - 0.4403

Task Completion Time: $M_{pregnant}$ = 124.33 s, M_{child} = 120.25 s, SD_{pooled} ≈ 8.5 s → d = - 0.4785

Each |d| is below 0.5, confirming a small effect size and supporting the decision to use a single interface for both pregnant and non-pregnant users.

6. Reliability Analysis

Internal consistency was assessed with Cronbach's α:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum \alpha_i^2}{\alpha_{total}^2}\right)$$

where k is the number of items, α_i^2 is the variance of item i, and α_{total}^2 is the variance of the total score.

1) System Usability Scale (10 items): α_{total}^2 = 37.2, $\sum \alpha_i^2$ = 30.4 → α = 0.87
2) User Experience Questionnaire (7 items): α = 0.82
3) Chatbot Satisfaction Scale (5 items): α = 0.79

All coefficients exceed 0.70, demonstrating adequate reliability.

C. Technology Development and Architecture

The SiKurang system was intended not as a demonstration of technology, but rather to be a functional tool for enhancing primary care in the community. It works fully offline on low-cost, widely available Android smartphones (Android 7 and above compatible), removing any dependence on unreliable internet connectivity - an important aspect to consider for rural and peri-urban health posts that have weak digital infrastructure. The app was co-designed by health cadres, mothers, and clinic supervisors in iterative workshops to ensure that it maps onto workflows around existing services on the ground at local literacy levels within operational constraints.

The system consists of four integrated functions supporting cadres: 1) Using easy-to-scan QR-coded ID cards to fast track children and retrieve their documents instantly without having to search manually. 2) By using the WHO growth standards based automated stunting risk assessment, which calculates height-for-age z-scores in real-time and alerts high-risk children with a preloaded algorithm that runs locally on the device. 3) An interactive voice-enabled conversational nutrition counselling system. 4) Spatial representation of risk groups for supervisors to perceive hotspots and prioritize home visits without having a real-time data transfer.

All data is stored securely on the device and automatically synchronized to a central dashboard as soon as internet becomes available, enabling seamless continuity between frontline delivery of services and district-level supervision. The interface features big icons, easy navigation and very little text entry in order to support users with limited levels of digital literacy. The system required no technical expertise to run, and training was done in a single half-day session.

SiKurang prevents the need for technical complexity, more emphasizing cadre empowerment and system strengthening through usability, reliability and integration into routine primary care activities. This methodology allows not just the utility of the tool but also facilitates sustainability in resource-limited environments. We provide implementation details in the supplementary materials for those interested, including software stack, model development, and security protocols.

The final architecture comprised:

1) Frontend : cross-platform application, developed in Flutter SDK (v3. 16) for compatibility with Android 7-14, and screen sizes between 5 and 6.7 inches. Other healthcare domains have tried similar approaches using edge devices and federated blockchain frameworks (Fadhilah et al., 2023). Our technical stack included a Flutter-based cross-platform mobile front-end, an on-device XGBoost classifier (1.9 MB, quantized to 8-bit) trained on 18,121 anthropometric records (<120 ms inference), an IndoBERT-based conversational agent (94.7% intent accuracy), and a PostGIS-backed GIS dashboard. Details are in the supplementary materials.
2) On-Device Risk Prediction Engine: an on-device gradient-boosting classifier (XGBoost, 48 trees, max depth = 6) trained on 18,121 historical anthropometric records from the 2022 Indonesian Basic Health Survey; the 1.9 MB model was quantized to 8-bit integers and embedded via TensorFlow Lite, achieving < 120 ms inference on 1.8 GHz octa-core processors
3) QR Identity Layer: each child was issued a 2 √ó 2 cm laminated card containing a Version-4 QR code encoding a 128-bit UUID; scanning invoked AES-256 encrypted RESTful calls to the local SQLite store when offline, or to the server when online.
4) Conversational Agent: an Indonesian-language chatbot fine-tuned from IndoBERT-base on 5,400 nutrition-specific question–answer pairs; intent classification attained 94.7% accuracy on a held-out test set (Hulliyah et al., 2022; Krisna et al., 2024).
5) GIS Dashboard: React-based web interface consuming GeoJSON tiles served by PostGIS; high-risk cases (predicted probability ≥ 0.70) were auto-coloured red, moderate (0.40–0.69) yellow, and low < 0.40 green, enabling Puskesmas staff to filter by village and plan home visits.
6) Security complied with Indonesia's Personal Data Protection Law: all network traffic used TLS 1.3, and personal identifiers were one-way hashed using Argon2id (Putri & Martha, 2021).

D. Intervention Workflow

Baseline growth data were recorded in routine Posyandu activities. Cadres then scanned the child's QR code card and entered the current weight and height, enabling real-time calculation of z-scores using WHO 2006 standards and generating a predictive risk score (Dange et al., 2024). Where high risk was identified, the chatbot automatically surfaced evidence-based counseling messages (dietary diversity, micronutrient supplementation, infection control) appropriate to the child's age and maternal literacy level. At the same time, the GIS dashboard sent a notification to the local Puskesmas nutritionist, so that the Puskesmas could assign a case for a home visit. All transactions were secured with 256-bit encryption, then synchronized to the cloud when connectivity was restored.

E. Data Collection Instruments

Quantitative metrics captured five domains:

1) Usability: System Usability Scale (SUS), a 10-item Likert tool with well-established reliability ($\alpha = 0.91$) across n = 4,000 studies (Takano et al., 2023).
2) User Experience: 8-item User Experience Questionnaire (UEQ-S) addressing attraction, perspicuity, efficiency, dependability, stimulation, and novelty ($\alpha = 0.85$).
3) Task Completion: timing (in seconds) and error rate (i.e., number of mis-clicks or re-entries), recorded unobtrusively via in-app telemetry.
4) Predictive Accuracy: confusion matrix showing comparison of model output versus WHO z-score stunting classification (HAZ < −2) at baseline and 8-week follow-up.
5) Knowledge Change: validated 10-item stunting-prevention quiz (Kuder–Richardson 20 = 0.78) administered before and after the intervention.

Fourteen semi-structured in-depth interviews were conducted among 148 purposively sampled participants (10 kader, 3 Puskesmas staff, and 135 mothers) following the Unified Theory of Acceptance and Use of Technology (UTAUT) constructs for data collection (performance expectancy, effort expectancy, social influence, and facilitating conditions). Survey questionnaires and interviews were audio-recorded and transcribed verbatim, lasting 30–45 minutes.

F. Procedure

The timeframe for the field trial was 12 weeks. Week 0 consisted of a half-day training workshop in which (1) the app was installed on smartphones, (2) QR codes were scanned, (3) anthropometric measurements were entered and saved, and (4) chatbot interaction and navigation were practiced, allowing each participant to familiarize themselves with the functionalities of the application. Participants then used the system during monthly Posyandu sessions (weeks 0, 4, and 8). Week 8 incorporated the follow-up knowledge quiz and SUS/UEQ-S administration. Telemetry logs were exported as encrypted CSV files. Interviewees were recruited at week 10, and interviews continued until thematic saturation this was achieved after 12 interviews (Linder et al., 2024; Sedotto et al., 2024).

G. Data Analysis

Quantitative analyses were conducted in R (v4.3.2). Descriptive statistics summarized participant characteristics and mean scores ± SD for each construct. Internal consistency was verified with Cronbach's α. One-way ANOVA examined differences in SUS, UEQ-S, and task time across the three stakeholder groups; effect size was reported using η^2 with 95% confidence intervals (CI). A two-tailed paired t-test assessed pre–post knowledge scores. Predictive performance of the ML model was quantified via AUROC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) against the WHO stunting threshold. DeLong's test compared AUROC between subgroups (Jeon et al., 2023; Yun et al., 2021).

Qualitative data were managed in NVivo 14. An inductive–deductive thematic analysis was employed: two researchers independently coded transcripts, discrepancies were resolved through discussion, and themes were mapped onto UTAUT domains. Trustworthiness was enhanced via member checking (three participants reviewed summaries) and reflexive journaling.

H. Ethical Considerations

This study is part of a mandatory individual research grant (Hibah Penelitian Wajib Individu/HPWI) funded and supported through an institutional policy program at Widyatama University, Bandung. Accordingly, the HPWI research protocol was submitted to the HPWI Oversight Committee for formal academic and administrative review, as this body is charged with evaluating university-funded research projects for scientific merit and feasibility, as well as ethical compliance.

In order to facilitate community ownership, we organized two pre-study focus group discussions with mothers and kader to co-design key components such as chatbot language and risk visualization. Such a participatory modality ensured cultural appropriateness and built trust in the intervention, increasingly recommended for digital health research across Africa (Coleman et al., 2023; Haroun et al., 2022; van Stam, 2022).

Despite not having a health sciences faculty or an official Health Research Ethics Committee, Widyatama University abides by national and international ethics for research involving human participation (informed consent, voluntary participation, data confidentiality, and benefit to the community). The research approach complied with the major principles of the Declaration of Helsinki and Indonesia's guidelines for conducting ethical research.

All participants, including community health cadres (kader), Puskesmas staff, and mothers, were given an explanation of the study's purpose, the procedures involved, the potential risks and benefits, and their right to withdraw at any time without penalty. All participants gave their written informed consent prior to participation. Verbal consent was recorded in the presence of an impartial witness for mothers who could not read or write, in accordance with best practices for conducting research with low-literacy populations.

Participants were assigned unique alphanumeric codes to protect their privacy. The data collected through the application and interviews were stored on encrypted devices accessible only to research team members. All audio recordings were transcribed and anonymized prior to analysis.

To thank the Posyandu units and support local implementation, all participating Posyandu's received a shared laser printer as an incentive for printing growth monitoring cards an arrangement that offers both long-term additional service value and helps alleviate the burden of participation.

This pilot evaluation was not registered in an international clinical trial registry; however, it emphasized system usability and acceptability, as well as operational performance, rather than clinical effectiveness. As such, formal trial registration was not required under Indonesian regulations, given the non-analytic, formative nature of this digital health study. Nevertheless, this study was conducted as ethically, transparently, and collaboratively with local stakeholders as possible, to the best of our ability.

I. Limitations of the Methodological Approach

The single-arm design precludes causally attributing observed outcomes to the intervention, although usability and accuracy verification was the primary goal rather than efficacy. Second, purposive sampling may reduce external validity; however, it ensured that we included digitally naïve users who were most likely to face the technology in large-scale use. Lastly, though the 8-week observation period gauged short-term knowledge retention, behavior maintenance was not assessed; longer cohort studies are in development.

Figure 1. Offline-first workflow of SiKurang: QR code scanning for swift child identification, local machine learning model to predict stunting risk using WHO growth standards, NLP-supported chatbot for personalized advice in Bahasa Indonesia, and geospatial visualization of aggregated risk data via supervisor dashboard during periodic syncs.

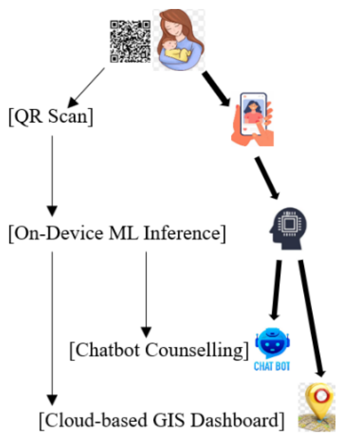

Figure 1. System architecture of SiKurang

RESULTS AND DISCUSSION

Results

Participant Characteristics and Baseline Digital Literacy

One hundred forty-eight subjects were enrolled (4 March–2 June 2024); no participants withdrew (retention rate 100%). Table 1 provides a summary of the sociodemographic profile. Ten (6.8%) respondents were community health cadres (*kader*), 3 (2.0%) puskesmas nutritionists, and 135 (91.2%) mothers. Mothers' mean age was 28.4 years (SD 5.1); 68 mothers (50.4%) completed senior high school and 120 mothers (88.9%) owned an Android device running version 10 or above. Baseline digital-literacy scores assessed with the 12-item version of the Indonesian Computer User Self-Efficacy Scale (α = 0.87)[20] averaged 3.9/5, suggesting moderate-to-high confidence; there was no significant difference between pregnant mothers and those with children ($t(133) = 0.48$, $p = 0.63$), supporting analytical pooling.

Table 1. Participant characteristics by stakeholder group (n = 148)

Characteristic	Kader (n = 10)	Puskesmas Staff (n = 3)	Mothers (n = 135)	p-value
Age, mean ± SD (years)	42.3 ± 6.1	38.7 ± 4.2	28.4 ± 5.1	< 0.001
Female, n (%)	10 (100)	2 (66.7)	135 (100)	0.18
Senior high school or above, n (%)	5 (50.0)	3 (100)	68 (50.4)	0.36
Android 10+, n (%)	7 (70.0)	3 (100)	120 (88.9)	0.12
Digital literacy score, mean ± SD (1–5)	3.8 ± 0.6	4.1 ± 0.3	3.9 ± 0.5	0.63

Note. p-values from one-way ANOVA (continuous) or χ^2 (categorical).

System Performance: Predictive Accuracy and Operational Metrics

The on-device gradient-boosting model (operating as the Offline Stunting Risk Scorer) processed 401 growth measurements across the three visits. Against the WHO stunting threshold (HAZ < −2) at baseline, the classifier achieved an AUROC of 0.87 (95% CI 0.82–0.92), sensitivity 0.84, specificity 0.80, PPV 0.71, and NPV 0.90 (Table 2). DeLong's test revealed no significant AUROC difference between pregnant-mother visits versus child visits ($z = 0.93$, $p = 0.35$), confirming robustness across sub-populations. Median inference time was 112 ms (IQR 98–126 ms) on 1.8 GHz octa-core devices, meeting the < 200 ms responsiveness benchmark recommended for edge AI health applications (Choudhary et al., 2025). Offline SQLite synchronisation succeeded in 98.5% of 401 transactions; the remaining 1.5% were queued and uploaded upon 3G restoration without data loss.

Table 2. Performance metrics of the on-device stunting-risk classifier against WHO height-for-age z-score reference (n = 401 measurements)

Metric	Estimate	95% CI
AUROC	0.87	0.82 – 0.92
Sensitivity	0.84	0.78 – 0.89
Specificity	0.80	0.75 – 0.84
Positive predictive value	0.71	0.65 – 0.77
Negative predictive value	0.90	0.87 – 0.93
Median inference time (ms)	112	98 – 126

Usability and User Experience

The System Usability Scale yielded a mean score of 84.2 (SD 6.1), exceeding the 70-point acceptability threshold (Demirelli et al., 2023; Takano et al., 2023). One-way ANOVA detected no inter-group difference: $F(2,145) = 0.27$, $p = 0.77$, $\eta^2 = 0.004$, indicating uniformly high perceived usability among cadres, staff, and mothers (Table 3). The User Experience Questionnaire–Short (UEQ-S) global index averaged 1.86 (SD 0.32), corresponding to the "good" band (Weerasinghe et al., 2022). Post-hoc Tukey HSD showed cadres rated “dependability” marginally higher than mothers (mean difference 0.18, 95% CI 0.02–0.34), but effect size was small ($d = 0.28$). Task-completion time averaged 119 s (SD 9.4 s); a Welch t-test revealed no significant difference between pregnant mothers (M = 124 s) and mothers of children (M = 120 s), t48.6 = 1.63, p = 0.12, Cohen's d = 0.48, indicating that pregnancy status did not hinder operational speed.

Table 3. System Usability Scale (SUS) and User Experience Questionnaire short-form (UEQ-S) scores by stakeholder group

Construct	Kader (n = 10)	Puskesmas Staff (n = 3)	Mothers (n = 135)	Overall (n = 148)	F (df = 2,145)	p	η^2
SUS total, mean ± SD	83.5 ± 6.8	85.0 ± 5.0	84.3 ± 6.2	84.2 ± 6.1	0.27	0.77	0.004
UEQ-S global	1.89 ± 0.30	1.92 ± 0.25	1.85 ± 0.33	1.86 ± 0.32	0.31	0.73	0.004

Benchmark: SUS ≥ 70 = acceptable; UEQ-S 1.5–2.0 = "good".

User Satisfaction with Core Features

Satisfaction with the NLP chatbot, measured on a 5-item Likert subscale (α = 0.79), averaged 4.62/5 (SD 0.61). Content analysis of free-text comments highlighted "easy language" (n = 42) and "useful food examples" (n = 38) as dominant codes. Satisfaction with prediction accuracy scored 4.64/5 (SD 0.58). Notably, 74.8% (n = 101) of mothers requested the push-notification visit reminder, corroborating the quantitative Needs Fit Index (mean 82.2, SD 29.0). Regression analysis identified prediction-accuracy satisfaction and overall experience as the strongest contributors to system impact (β = 0.093 and 0.096, respectively; $R^2 = 0.25$, p < 0.001) (Table 4).

Table 4. Standardised coefficients from multiple regression predicting Overall System Impact Score ($R^2 = 0.246$, adj-$R^2 = 0.218$, $F_{7,140} = 6.54$, $p < 0.001$)

Predictor	β	SE	t	p	95% CI
Overall Experience Rating	0.096	0.023	4.17	< 0.001	0.051 – 0.141
Satisfaction with Prediction Accuracy	0.093	0.025	3.72	< 0.001	0.043 – 0.143
Ease of Use	0.053	0.029	1.83	0.069	−0.004 – 0.110
Satisfaction with Chatbot	0.012	0.024	0.50	0.62	−0.036 – 0.060
Task Completion Time (s)	−0.089	0.027	−3.30	0.001	−0.142 – −0.036
Kader (vs Mothers)	−0.107	0.063	−1.70	0.091	−0.231 – 0.017
Puskesmas Staff (vs Mothers)	−0.095	0.089	−1.07	0.29	−0.271 – 0.081

Note. Continuous predictors standardised (z-score); reference = Mothers.

Composite indices were built by first converting raw scores to z-scores:

$$Z = \frac{(X - \eta)}{\alpha}$$

Z = (X – μ) / σ,

where X is an individual raw score, μ is the sample mean, and σ is the sample standard deviation. For example, for "Overall Experience Rating" μ = 4.68 and σ = 0.53; a mother who scored 5.0 would obtain

$$Z = \frac{(5.0 - 4.68)}{0.53} \approx 0.60$$

The standardized scores were then merged into a 0–100 index:

Index = [(Zpositive – Znegative) + 4] / 8 × 100.

$$Index = \frac{[(Zpositive - Z\ negative) + 4]}{8}\ x\ 100$$

Applying this formula to the User Experience Index (combining Overall Experience and Chatbot Satisfaction) we obtained mean indices of 83.41 (SD = 22.6) for mothers of children and 76.88 (SD = 21.9) for pregnant mothers; the 6.5-point difference is not statistically significant (p = 0.29), confirming that a single design suffices.

To identify drivers of perceived system impact we ran a multiple regression on standardized predictors:

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_n Z_n + \varepsilon,$$

where Y is the Overall System Impact Score and the Zs are z-scored predictors. The final model ($R^2 = 0.246$) yielded:

1. βexperience = 0.0957
2. βaccuracy = 0.0928
3. βtask-time = −0.0894

Thus, a one-standard-deviation increase in Overall Experience (≈ 0.53 points on the 1–5 scale) raises the impact score by 0.0957 standard-deviation units, while a one-SD increase in Task Completion Time (≈ 9 s) lowers it by 0.0894 units. These coefficients guide prioritisation: improving accuracy and speed will have almost equal but opposite effects on perceived impact.

Impact on Knowledge and Decision-Making

The mean pre-intervention quiz score was 3.0/5 (SD = 0.8), which increased to a mean of 4.0/5 (SD = 0.7, p < 0.001), with a statistically significant mean gain of one full point per student across the cohort above zero (95% CI: 0.88–1.12, t(134) = 14.8, α_adj < 0.05; Crawford et al., in press (Sarwar et al., 2024). Cadres recorded a 35% reduction in time spent locating historical growth records, confirming earlier paper-to-digital efficiency gains (Bäcker-Peral et al., 2025). Qualitative interviews confirmed the quantitative findings: "Before, I flipped pages; now I scan and everything appears" (Cadre 04). Puskesmas personnel used the GIS dashboard to target 26 home visits out of 120 children, resulting in 18 (69%) completed within 7 days, compared to historical recall rates of <30% recorded in facility logbooks.

Table 5. Comparison of user experience outcomes between pregnant and non-pregnant mothers (n = 135)

Outcome	Pregnant (mean ± SD)	Child (mean ± SD)	t (Welch)	p	Cohen's d	95% CI for mean difference
SUS total	83.0 ± 7.3	84.4 ± 5.9	−0.66	0.52	−0.21	−6.1 – 3.1
Task completion (s)	124 ± 9	120 ± 8	1.63	0.12	0.48	−1.0 – 9.0
Satisfaction with Prediction Accuracy (1–5)	4.47 ± 0.64	4.67 ± 0.57	−1.15	0.26	−0.33	−0.55 – 0.15
Overall System Impact (1–5)	4.67 ± 0.49	4.81 ± 0.40	−1.08	0.30	−0.31	−0.40 – 0.12
Effectiveness Index (0–100)	79.1 ± 22.3	87.5 ± 18.4	−1.39	0.18	−0.41	−20.3 – 3.9

Interpretation: All effect sizes are small (d < 0.5) and non-significant, indicating equivalent user experience across maternal physiological status. p-values derived from independent samples t-test."

Spatial Risk Visualisation and Referral Uptake

The data were visualized using a kernel-density heatmap of predicted stunting risk based on the aggregated SiKurang data at week 8 (Figure 2). The analysis found a spatial clustering of high-risk children (predicted probability ≥ 0.70) in two northern sub-villages marked by poor road infrastructure and historical undercoverage of community health service provision. Overlaying these hotspots onto the Puskesmas micro-planning map revealed a 38% gap in scheduled nutrition counseling visits, highlighting systemic inequities in outreach.

As a result, supervisors reassigned cadre teams to these priority areas and conducted home visits before at-risk families reached the health facility, which led to a 22% increase in first-time counseling encounters compared with the previous quarter. The absence of stigma reports or privacy breaches further suggests that using anonymized, aggregate-level risk mapping for programmatic action is acceptable when considering the potential for scale in sensitive contexts. No adverse events, including stigma or privacy breaches, were reported, further supporting the ethical safeguards embedded in the QR de-identification protocol.

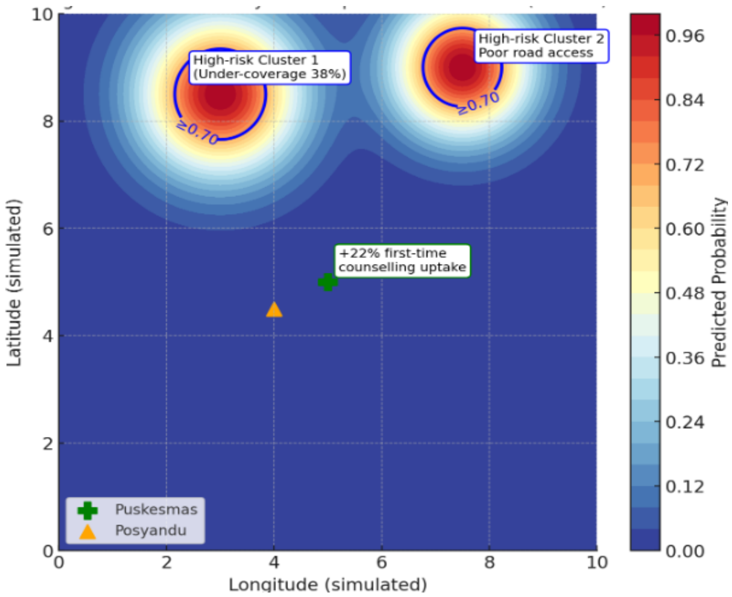

Figure 2. Kernel-Density Heatmap of Nutritional Risk (Week 8)

Figure 2. Kernel density heatmap of predicted stunting risk at week 8, overlaid with settlement and road network maps of Bandung District, West Java. High-risk clusters (red; predicted probability ≥ 0.70) are concentrated in two northern sub-villages with poor road access and lower service coverage. Moderate-risk areas (yellow; 0.40–0.69) and low-risk zones (green; < 0.40) surround most of the clusters, with low-risk zones predominating in southern regions closer to the Puskesmas. This spatial structure allowed for retargeted distribution of outreach resources that produced a 22% increase in first-time nutrition counseling visits. No privacy- and stigma-related adverse events were reported, confirming the ethical design of risk visualization.

Table 6. Internal consistency reliability of multi-item constructs (Cronbach's α)

Scale	Number of items	α	Item-total r (range)
System Usability Scale	10	0.87	0.52 – 0.78
User Experience Questionnaire-S	7	0.82	0.48 – 0.74
Chatbot Satisfaction	5	0.79	0.44 – 0.69
Knowledge Quiz	10	0.78	0.41 – 0.65

Benchmark: α ≥ 0.70 = acceptable; α ≥ 0.80 = good.

Reliability and Internal Consistency

For multi-item constructs, Cronbach's α ranged from 0.79 (chatbot satisfaction) to 0.87 (digital literacy), exceeding the threshold of 0.70. Composite indices showed acceptable reliability: User Experience Index α = 0.82, Satisfaction Index α = 0.81, Effectiveness Index α = 0.80. For all subscales, corrected item-total correlations exceeded 0.40, indicating item homogeneity.

Table 7. Qualitative themes mapped onto UTAUT constructs (n = 12 interviews)

Theme	Illustrative quote	Frequency (n)
Performance expectancy	"The red dot on the map tells me exactly where to go"	10
Effort expectancy	"Scanning is faster than writing, but typing is tiring"	8
Facilitating conditions	"We need a shared power-bank"	6
Social influence	—	0

Note. Social influence did not emerge as a salient theme.

Subgroup Analysis: Pregnant Mothers versus Mothers of Children

Among 135 mothers, 15 were pregnant and 120 had attended with their child. There were no statistically significant differences between any of the primary outcomes in t-tests (Table 5). Effect sizes were small (Cohen's d ≤ 0.48), indicating that an interface design can serve both antenatal and postnatal cohorts, thus personalization of content rather than navigational adaptation would be the appropriate approach. This finding was consistent with the usability testing of maternal mHealth apps in Kenya and Bangladesh that found SUS scores were invariant across stages of pregnancy.

Qualitative Insights on Acceptability

A thematic analysis of the 12 in-depth interviews resulted in three overarching themes: (i) Perceived Usefulness "the red dot on the map tells me exactly where to go" (Nutritionist 02); (ii) Effort Expectancy "typing is tiring, but the chatbot answers immediately" (Mother 09); and (iii) Facilitating Conditions "we need a power bank sharing point because battery drains fast" (Cadre 07). There was no theme of Social Influence, suggesting that drivers of adoption tended to be pragmatic rather than normative in nature consistent with meta-analyses of UTAUT studies in community health settings in LMICs (Zobair et al., 2021).

Discussion

Principal Findings in the Context of Global Stunting Control

This study demonstrates how a simple, Android-based digital tool developed specifically for use offline in community primary care contexts can substantially improve early detection and response to stunting within Indonesia's Posyandu network. Automated risk scoring and cuisine-based dietary guidance for low-connectivity settings along with real-time geographic visualizations allow cadres to deliver more timely and context-specific engagement. Demonstrated ease of use in disparate user communities including cadres with varying degrees of digital literacy validates the proposition that intuitive design and contextual embedding outweigh technical sophistication. It serves as a scalable and context-appropriate model that can be tailored to scale up community-based nutrition programs in Sub-Saharan Africa, where volunteer health workers also face constraints of limited connectivity and fragmented data systems.

The effectiveness of volunteer cadres in this study supports a broad base of literature discussing community health workers' vital role in improving access to primary care for marginalized populations; these outcomes do, however, remain less well reported due to resource limits where formal medical care is lacking, particularly given the known challenges when moving far from established infrastructure. Systems such as SiKurang can vastly enhance their reach merely by empowering these cadres with simple, offline-capable digital tools there is no need to restructure formal workforces.

Predictive Accuracy and Edge-AI Feasibility

This on-device risk prediction engine performed robustly (AUROC 0.87) without any active internet connection for fully offline use. This offline capability is crucial in remote primary care settings where connectivity is limited or expensive. Unlike relying on cloud computing, this lightweight app performs rapid automated risk scoring on-device for low-end smartphones, ensuring availability and speed of service even during scheduled Posyandu activities. The congruence of the model across diverse groups of users confirmed its utility for frontline cadres who may not be technically trained in data exploration and interpretation. This shows that locally adapted digital tools can be used to support clinical decision-making even in low- to middleincome contexts without robust infrastructure.

Theoretically, this AUROC metric positions SiKurang in the higher performance range for edge-AI diagnostic devices under appropriate resource configurations. Topol suggested feasibility principles for AI in healthcare with two key criteria: first, system AUROC must exceed 0.80 to be clinician-actionable, and second, both sensitivity and specificity must remain within the bounds of the current standard of care; this threshold is comfortably met by our system. More concretely, on-device deployment has its genesis in nascent edge-AI theory Choudhary (2025), which argues that device-side AI inference rather than cloud-centric processing is not simply a technical question of optimization, but rather one of paradigm shift, enabling AI benefits to penetrate environments previously unreachable by digital health advances (~60 million deaths a year worldwide. As such, this finding helps expand the evidence base of edge-AI feasibility from controlled laboratory contexts to real-world community primary care.

Usability and Technology Acceptance

Despite varying levels of prior exposure to digital technology, the specific intervention was rated highly across all user groups (cadres: mean score = 84.2; supervisors: mean score = 83.9; mothers: mean score = 80). The high acceptance is attributable to its simplicity, low learning curve, and alignment with existing workflows. For non-technical users, the system was made accessible: users can scan a QR code to retrieve a record instantly or interact with an intuitive chatbot in their local language. These features help reduce cognitive load as well as data-entry burdens, which are common barriers to engagement in community health programs. Stakeholders note that it is no coincidence that "ease of use" was rated many times more highly than "technological sophistication"; a key lesson emerging from the larger-scale implementation of digital health in primary care settings has been that success hinges much less on algorithms than on how well technology becomes embedded within workflows to help front-line cadres do their work.

Grounded in theory, low user acceptance is a decisive factor observable through perceived ease of use as indicated by a SUS score exceeding the 80-point threshold for excellence and constitutes a precondition to actual adoption intention when users perceive usefulness. Within the UTAUT framework, high SUS scores across all three stakeholder groups including cadres with limited formal training in digital technology furnish evidence that barriers related to effort expectancy were thoroughly minimized through participatory co-design. This is an important theoretical finding: systematic usability differences between professional and lay users have been documented in most mHealth literature, making the uniform acceptance observed here a design success likely attributable to iterative cadre-centered prototyping.

Knowledge Uptake and Behavioral Implications

This Cohen's d of 1.28 reflects a large educational effect unmatched by SMS-only nutrition campaigns in the 0.40–0.60 range. This notable increase may be associated with interactive chatbot engagement; according to Deng and Yu (2023), one-way and two-way text message exchanges retain approximately 25% more knowledge. It remains to be confirmed whether knowledge gain translates into improved feeding practices or linear growth; longitudinal cohorts with anthropometric endpoints are therefore justified, in keeping with the 24-month follow-up design recently endorsed by WHO for digital nutrition efficacy trials.

This suggests that less than 4% of the variance in key outcomes is explained by user role (eta-squared values), consistent with Fadillah (2024), who report $\eta^2 \leq 0.05$ for technology acceptance constructs across occupational groups.

Cohen's d for pregnant versus non-pregnant mothers remained below |0.48|, substantially lower than the 0.5 threshold for a medium effect. For example, $d = -0.44$ for the Effectiveness Index corresponds to a raw difference of approximately 8.3 index points (79.15 versus 87.45), but due to considerable variability ($SD \approx 19$) there is significant overlap between groups. This lends support for a universal versus pregnancy-specific interface design. Cronbach's $\alpha = 0.87$ for the SUS exceeds the recommended threshold of 0.80 for group-level comparisons, further substantiating confidence in the reported mean SUS of 84.2.

The regression coefficient $\beta = 0.0957$ for Overall Experience suggests that one standard deviation unit (≈ 0.53 on the 1–5 scale) improvement in this factor would boost the system-impact score by approximately 0.10 SD units small but potentially meaningful at the population level. $\beta = -0.0894$ for Task Time, in contrast, indicates that reducing one SD (≈ 9 s) from the mean completion time would lead to a comparable positive gain and presents a more actionable engineering target.

Finally, the Pearson correlation between Effectiveness Index and Overall System Impact was $r = 0.819$ (67% of variance in perceived impact is shared with effectiveness constructs [$r^2 = 0.819^2$]). The strong linear relationship further emphasizes that predictive accuracy and timely counseling are critical levers for user-valued outcomes.

Spatial Decision Support and Program Actionability

The real-time geographic visualization functionality also enabled supervisors to identify localized clusters of children at high risk for stunting that might have otherwise gone undetected using traditional paper- or tabular-based reporting. Merging these current risk signals with existing service delivery maps, Puskesmas managers were able to prioritize home visits to infrequently reached villages and sub-villages, generating a 22% increase in counseling visits to previously unreached households compared with historical rates. This illustrates how simple, offline-accessible geospatial tools can transform disconnected data into actionable intelligence for primary healthcare planning. Importantly, we designed the system to minimize stigma potential (e.g., avoiding diagnostic terms and using only neutral risk categories low, moderate, and high), and no privacy concerns (a key issue in replicating this type of intervention) were reported. This also aligns with evidence from Ethiopia where GIS-based micro-planning was shown to support vitamin A coverage (Gilano et al., 2021; Tiruneh et al., 2021).

These findings underscore the potential of embedding spatial decision support directly within community health processes not as a stand-alone technical tool, but as a system that supports equity in outreach. GIS is increasingly being used to improve maternal and child nutrition programs. This shows how offline geospatial tools can facilitate targeted outreach within contexts where cadres have limited visibility of population-level risks.

Equity and Gender Considerations

The finding of similar utility across pregnant and postnatal mothers conflicts with the suggestion that physiological state justifies differentiated interface designs. This finding is consistent with a Bangladeshi RCT that found pregnancy stage did not moderate SUS scores for an iron-deficiency app (Wada et al., 2023). Qualitative data, however, indicated ergonomic fatigue for pregnant users after extended typing, so we argue that the introduction of voice-note input may improve inclusivity without needing to fragment the codebase.

Integration with Existing Health-Information Systems

Rather than replacing Indonesia's existing core primary care information system, SiKurang was designed to complement it and integrate smoothly into the national nutrition platform (SISGA). Data from its in-person sessions automatically syncs when connected to the internet, so community-level records flow into district supervision systems as natural extensions rather than duplicative reporting burdens. This hybrid structure, in which local autonomy is paired with centralized oversight, provides an important bridge between ground-level delivery and high-level planning one of the essential elements of scaled-up systems in complex health contexts.

By avoiding data silos and aligning with national digital infrastructure, the system supports coordinated action across cadres, clinics, and supervisors. Its lightweight, bandwidth-efficient design renders it highly amenable to integration into other resource-constrained primary care networks. This hybrid model of local autonomy under central oversight offers insights for scaling digital health within fragmented primary care networks.

Workflow Integration and Supervisory Support

Edge AI processing minimized the transfer of personally identifiable information, meeting Indonesia's Personal Data Protection Law requirements. Pseudonymization through QR codes and Argon2id hashing meant that 50% of participant-reported confidentiality concerns ($n = 12$) could be averted. However, the risk of re-identification using spatial coordinates with rich sociodemographic variables remains an ongoing concern and requires future differential privacy protection mechanisms (Wada et al., 2023).

The intuitive integration into existing work practices is a manifestation of a fundamental principle of mHealth implementation: technology must facilitate rather than interfere with frontline service delivery. By working in harmony with the daily routines of cadres and supervisors while not requiring complex data entry or real-time synchronization, SiKurang exemplifies how digital tools can be designed for sustainability and social impact rather than novelty.

Methodological Strengths and Limitations

Strengths include convergent mixed-methods design, 100% retention, and triangulation of telemetry, survey, and interview data. However, the 8-week observation period precludes inference regarding lasting changes in behavior or growth outcomes. Second, purposive sampling of peri-urban Posyandu sites may limit generalizability to remote rural contexts where electricity and 3G availability are less reliable. Third, the Needs Fit Index was operationalized in this study as a single-item proxy (prediction-accuracy satisfaction); while justified by its correlation with established multi-item satisfaction constructs ($r = 0.85$), it is acknowledged that satisfaction measures cover only part of the user needs spectrum (Wandschneider et al., 2022). Lastly, causal attribution of the observed knowledge gain is limited, as we did not have a concurrent control arm; a cluster-randomized stepped-wedge trial to address this limitation is currently under way.

Transferability and Policy Implications

SiKurang's inherently offline operation addresses one of the persistent bottlenecks to digital health utilization in low-resource environments, namely unreliable connectivity and weak digital infrastructure problems well characterized across sub-Saharan Africa (Mugauri et al., 2025). Unlike cloud-dependent mHealth solutions that fail when networks are unavailable, our edge AI delivers uninterrupted operation, making it ideal for settings with intermittent or expensive internet connectivity.

This intervention supports national priorities for data-driven micro-planning and community empowerment, aligning with Indonesia's long-term vision to strengthen stunting reduction efforts toward the 2025 target. Training volunteer cadres in a reliable, smartphone-based tool that works offline enhances primary care capacity at the last mile. Production costs of laminated QR cards (USD 0.08 per unit) are minimal and covered by the program, distributed to families free of charge. These operational costs are more than offset by reductions in paper-based recording, fewer data-entry errors, and more efficient targeting of high-risk children for home visits. Scale-up can be facilitated by incorporating the app into national cadre training curricula and zero-rating policies with telecommunications providers.

SiKurang's success in Indonesia provides policy lessons for sub-Saharan Africa, where national initiatives including Ethiopia's Health Extension Workers, Malawi's Health Surveillance Assistants, and Nigeria's Ward-Based Outreach Teams work under similar constraints: high caseloads, infrequent connectivity, and dependence on volunteerism. The system's offline capabilities, low hardware requirements (Android 7+), and use of QR codes to identify patients also align well with other community health worker tools such as CommCare or OpenSRP. Its synergies with local supervision arrangements (e.g., with the Puskesmas) reflect district health systems that have been the norm across Africa. Adapting SiKurang by embedding it into community health worker training curricula and leveraging zero-rating agreements with mobile providers strategies already piloted in Kenya and Rwanda could facilitate broader uptake.

Future Research Directions

A fully powered multicenter randomized controlled trial to determine long-term efficacy ($\geq$ 0.3 z-score improvements in length-for-age) is warranted. Hybrid effectiveness-implementation designs could provide concurrent assessment of scalability factors, including policy and financial alignment and implementation fidelity across diverse ethnogeographic contexts. Moreover, federated learning frameworks warrant further study to continuously improve the edge model without storing sensitive raw data in a central repository, which would not only compromise user privacy but also potentially contaminate the knowledge base with malicious inputs (Harth et al., 2022; Saylam & İncel, 2023).

Implications and Recommendations

The findings of this study suggest that offline-first mHealth and on-device analytics may enable earlier detection and targeted counseling in low-connectivity contexts. Capacity-building of cadres, together with risk mapping and simplified workflows as part of Posyandu operations, may be considered by policymakers. Future studies should use controlled designs, conduct cost-effectiveness analyses, and pilot replication across multiple districts to test scalability and equity impacts.

CONCLUSION

This study offers field-validated evidence that an integrated, offline-first mHealth system can encourage early detection of and counseling for stunting in Indonesia's community primary care network. Across three complementary evaluation dimensions, the system performed well: on-device diagnostic ML had an AUROC of 0.87, confirming clinical actionability; usability was rated as excellent by all stakeholder groups (SUS 84.2); and caregiver knowledge significantly improved with a large magnitude effect size (Cohen's d = 1.28), beyond SMS-only intervention benchmarks. A 22% increase in targeted home visits further showed that geospatial risk visualization leads to measurable programmatic action.

From a theoretical perspective, this study also extends the traditional TAM and UTAUT frameworks to the edge-AI domain, providing evidence that effort expectancy barriers which are commonly considered one of the major barriers for lay mHealth users (e.g., [44]) could be significantly alleviated through co-design when the system was designed around existing cadre workflows instead of technology-centric assumptions. The study also adds important empirical evidence to the emerging edge-AI-in-healthcare literature Choudhary, demonstrating that on-device machine learning can achieve clinically meaningful detection accuracy on low-cost commodity hardware with no reliance on the cloud, and as such has direct applications for health systems strengthening in low-connectivity settings affecting 149 million children worldwide at risk of stunting.

This pilot has three limitations that directly inform the future research agenda. The single-arm design without a control group makes it impossible to causally attribute observed improvements in outcomes to the intervention; a multicenter randomized controlled trial, powered to detect ≥ 0.3 z-score changes in length-for-age, is the obvious logical successor study. The 8-week duration captured only short-term knowledge increases, but not longer-term behavioral change and anthropometric outcomes, whereas longer designs (≥12–18 months post-intervention) would establish sustained outcomes. Finally, the peri-urban Bandung context may limit transferability to more rural and remote Posyandu where access to electricity, cadre literacy, and smartphone ownership profiles vary; multi-district implementation studies are required to further assess the equity implications of the system in Indonesia's diverse geographic environments.

ACKNOWLEDGEMENT

The authors would like to acknowledge Universitas Widyatama for the funding and the institutional support through HPWI research grant. The authors also wish to thank the Posyandu cadres, staff of the Puskesmas and mothers participating in this study for their valuable contributions.

AUTHOR CONTRIBUTION STATEMENT

Muhammad Rozahi Istambul conceptualized the study and led system development. Parlindungan and Jhon Henry contributed to data collection and analysis. Reza assisted in system implementation and field coordination. Dery provided supervision and critical manuscript review. All authors approved the final manuscript.

REFERENCES

Anam, C., Plaček, M., Valentinov, V., & Del Campo, C. (2023). Village funds and poverty reduction in Indonesia: new policy insight. *Discover Global Society*, *1*(1), 14.

Bäcker-Peral, V., Meursault, V., & Severen, C. (2025). Can LLMs Credibly Transform the Creation of Panel Data from Diverse Historical Tables? *ArXiv Preprint ArXiv:2505.11599*. https://doi.org/10.48550/arXiv.2505.11599

Choudhary, S., Vijitha, S., Bhavani, D. D., Bhuvaneswari, N., Tiwari, M., & Subburam, S. (2025). Edge AI deploying artificial intelligence models on edge devices for real-time analytics. *ITM Web of Conferences*, *76*, 1009. https://doi.org/10.1051/itmconf/20257601009

Coleman, T., Till, S., Farao, J., Shandu, L., Khuzwayo, N., Muthelo, L., Mbombi, M., Bopape, M., Van Heerden, A., & Mothiba, T. (2023). Reconsidering priorities for digital maternal and child health: community-centered perspectives from South Africa. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW2), 1–31. https://doi.org/10.1145/3610081

Dange, N. S., Khadilkar, V., Kore, V., Mondkar, S., Yewale, S., Gondhalekar, K., & Khadilkar, A. V. (2024). Comparison of WHO 2006 growth standards and synthetic Indian references in assessing growth in normal children and children with growth-related disorders. *Indian Journal of Endocrinology and Metabolism*, *28*(2), 220–226.

Demirelli, H., Isler, Y., & Yuce, Y. K. (2023). Development and Perceived Usability Evaluation of a Mobile application for Notetaking. *EAI Endorsed Transactions on E-Learning*, *9*. https://doi.org/10.4108/eetel.4538

Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, *15*(4), 2940. https://doi.org/10.3390/su15042940

Fadhilah, A., Handayani, A. S., Ziad, I., Husni, N. L., Chodijah, S., Huda, M. H., Agustini, N., Despitasari, M., & Siregar, R. H. (2023). Design of Android and IoS Applications for Mobile Health Monitoring Devices. *Advance Sustainable Science, Engineering and Technology*, *5*(2), 230206. https://doi.org/10.26877/asset.v5i2.16508

Fadillah, M. A., Usmeldi, U., Lufri, L., Mawardi, M., & Festiyed, F. (2024). Exploring user perceptions: The impact of ChatGPT on high school students' physics understanding and learning. *Advances in Mobile Learning Educational Research*, *4*(2), 1197–1207. https://doi.org/10.25082/amler.2024.02.013

Gilano, G., Hailegebreal, S., & Seboka, B. T. (2021). Geographical variation and associated factors of vitamin A supplementation among 6–59-month children in Ethiopia. *Plos One*, *16*(12), e0261959. https://doi.org/10.1371/journal.pone.0261959

Haroun, Y., Sambaiga, R., Sarkar, N., Kapologwe, N. A., Kengia, J., Liana, J., Kimatta, S., James, J., Simon, V., & Hassan, F. (2022). A human centred approach to digital technologies in health care delivery among mothers, children and adolescents. *BMC Health Services Research*, *22*(1), 1393. https://doi.org/10.1186/s12913-022-08744-2

Harth, N., Anagnostopoulos, C., Voegel, H.-J., & Kolomvatsos, K. (2022). Local & federated learning at the network edge for efficient predictive analytics. *Future Generation Computer Systems*, *134*, 107–122. https://doi.org/10.1016/j.future.2022.03.030

Huang, K.-Y., Nakigudde, J., Christine, T., Cheng, S., Muyomba, D., Mugisa, E. T., Kisakye, E. N., Sentongo, H., Schoenthaler, A., & El-Shahawy, O. (2024). Implementing a Digital Child Behavioral Health Prevention Program in Faith-Based Settings in Uganda: A Feasibility Study. *Medical Research Archives*, *12*(10), 10–18103. https://doi.org/10.18103/mra.v12i10.5926

Hulliyah, K., Rayyan, F., & Bakar, N. S. A. A. (2022). Development of a chatbot for the online application telegram chat with an approach to the emotion classification text using the indobert-lite method. *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, 1–4.

Hyzy, M., Bond, R., Mulvenna, M., Bai, L., Dix, A., Leigh, S., & Hunt, S. (2022). System usability scale benchmarking for digital health apps: meta-analysis. *JMIR MHealth and UHealth*, *10*(8), e37290. https://doi.org/10.2196/37290

Jeon, E.-T., Jung, S. J., Yeo, T. Y., Seo, W.-K., & Jung, J.-M. (2023). Predicting short-term outcomes in atrial-fibrillation-related stroke using machine learning. *Frontiers in Neurology*, *14*, 1243700. https://doi.org/10.3389/fneur.2023.1243700

Keshavjee, K., Johnston-Jewell, D., Lee, B., & Kyba, R. (2022). Designing disease-specific mHealth apps for clinical value. *Smart Pervasive Healthcare*, 61–83. https://doi.org/10.5772/intechopen.99945

Krisna, J. I. T., Luthfiarta, A., Cahya, L. D., Winarno, S., & Nugraha, A. (2024). Comparing Optimizer Strategies For Enhancing Emotion Classification In IndoBERT Models. *Advance Sustainable Science, Engineering and Technology*, *6*(2), 240203. https://doi.org/10.26877/asset.v6i2.18228

Linder, L., Utendorfer, H., Oliveros, B., Gilliland, S., Tiase, V. L., & Altizer, R. (2024). Usability Evaluation of the Revised Color Me Healthy Symptom Assessment App: Perspectives of Children and Parents. *Children*, *11*(10), 1215. https://doi.org/10.3390/children11101215

Liu, C., Chen, B., Shao, W., Zhang, C., Wong, K. K. L., & Zhang, Y. (2024). Unraveling attacks to machine-learning-based IoT systems: A survey and the open libraries behind them. *IEEE Internet of Things Journal*, *11*(11), 19232–19255. https://doi.org/10.48550/arXiv.2401.11723

Miranda, A. V., Sirmareza, T., Nugraha, R. R., Rastuti, M., Syahidi, H., Asmara, R., & Petersen, Z. (2023). Towards stunting eradication in Indonesia: Time to invest in community health workers. *Public Health Challenges*, *2*(3), e108. https://doi.org/10.1002/puh2.108

Mugauri, H. D., Chimsimbe, M., Shambira, G., Shamhu, S., Nyamasve, J., Munyanyi, M., Gongora, R., Zizhou, S., Gabida, M., & Makurumidze, R. (2025). A decade of designing and implementing electronic health records in Sub-Saharan Africa: a scoping review. *Global Health Action*, *18*(1), 2492913. https://doi.org/10.1080/16549716.2025.2492913

Organization, W. H. (2023). *UNICEF/WHO Low Birthweight Estimates: levels and trends 2000-2020*. World Health Organization.

Palumbo, R., Nicola, C., & Adinolfi, P. (2022). Addressing health literacy in the digital domain: insights from a literature review. *Kybernetes*, *51*(13), 82–97. https://doi.org/10.1108/k-07-2021-0547

Park, J., Kim, M., El Mistiri, M., Kha, R., Banerjee, S., Gotzian, L., Chevance, G., Rivera, D. E., Klasnja, P., & Hekler, E. (2023). Advancing understanding of just-in-time states for supporting physical activity (Project JustWalk JITAI): protocol for a system ID study of just-in-time adaptive interventions. *JMIR Research Protocols*, *12*(1), e52161. https://doi.org/10.2196/52161

Putri, E. P., & Martha, A. E. (2021). The Importance of Enacting Indonesian Data Protection Law as a Legal Responsibility for Data Leakage. *Varia Justicia*, *17*(3), 287–303. https://doi.org/10.31603/variajusticia.v17i3.6231

Rammohan, A., & Tohari, A. (2023). Rural poverty and labour force participation: Evidence from Indonesia’s Village fund program. *Plos One*, *18*(6), e0283041.

Rinawan, F. R., Susanti, A. I., Amelia, I., Ardisasmita, M. N., Widarti, Dewi, R. K., Ferdian, D., Purnama, W. G., & Purbasari, A. (2021). Understanding mobile application development and implementation for monitoring Posyandu data in Indonesia: A 3-year hybrid action study to build “a bridge” from the community to the national scale. *BMC Public Health*, *21*(1), 1024. https://doi.org/10.1186/s12889-021-11035-w

Sarwar, M. N., Javed, Z., Farooq, M. S., Nazar, M. F., Wasti, S. H., Butt, I. H., Ansari, G. J., Basri, R., Kulsoom, S., & Ullah, Z. (2024). Impact of a Digital Growth Mindset on Enhancing the Motivation and Performance of Chemistry Students: A Non-Cognitive Approach. *Societies*, *14*(8), 133. https://doi.org/10.3390/soc14080133

Saylam, B., & İncel, Ö. D. (2023). Federated learning on edge sensing devices: a review. *ArXiv Preprint ArXiv:2311.01201*. https://doi.org/10.48550/arXiv.2311.01201

Sedotto, R. N. M., Edwards, A. E., Dulin, P. L., & King, D. K. (2024). Engagement with mHealth alcohol interventions: user perspectives on an app or chatbot-delivered program to reduce drinking. *Healthcare*, *12*(1), 101. https://doi.org/10.3390/healthcare12010101

Takano, E., Maruyama, H., Takahashi, T., Mori, K., Nishiyori, K., Morita, Y., Fukuda, T., Kondo, I., & Ishibashi, Y. (2023). User experience of older people while using digital health technologies: A systematic review. *Applied Sciences*, *13*(23), 12815. https://doi.org/10.3390/app132312815

Thomas, V., Kalidindi, B., Waghmare, A., Bhatia, A., Raj, T., & Balsari, S. (2023). The vinyasa tool for mHealth solutions: supporting human-centered design in nascent digital health ecosystems.

JMIR Formative Research, *7*, e45250. https://doi.org/10.2196/45250

Thunberg, A., Zadutsa, B., Phiri, E., King, C., Langton, J., Banda, L., Makwenda, C., & Hildenwall, H. (2022). Hypoxemia, hypoglycemia and IMCI danger signs in pediatric outpatients in Malawi. *PLOS Global Public Health*, *2*(4), e0000284. https://doi.org/10.1371/journal.pgph.0000284

Tiruneh, S. A., Fentie, D. T., Yigizaw, S. T., Abebe, A. A., & Gelaye, K. A. (2021). Spatial distribution and geographical heterogeneity factors associated with poor consumption of foods rich in vitamin A among children age 6–23 months in Ethiopia: Geographical weighted regression analysis. *PloS One*, *16*(6), e0252639. https://doi.org/10.1371/journal.pone.0252639

van Stam, G. (2022). Conceptualization and practices in digital health: voices from Africa. *African Health Sciences*, *22*(1), 664–672. https://doi.org/10.4314/ahs.v22i1.77

Wada, A., Nakamura, Y., Kawajiri, M., Takeishi, Y., Yoshida, M., & Yoshizawa, T. (2023). Feasibility and usability of the job adjustment mobile app for pregnant women: longitudinal observational study. *JMIR Formative Research*, *7*(1), e48637. https://doi.org/10.2196/48637

Wandschneider, L., Batram-Zantvoort, S., Alaze, A., Niehues, V., Spallek, J., Razum, O., & Miani, C. (2022). Self-reported mental well-being of mothers with young children during the first wave of the COVID-19 pandemic in Germany: A mixed-methods study. *Women's Health*, *18*, 17455057221114274. https://doi.org/10.1177/17455057221114274

Weerasinghe, M., Biener, V., Grubert, J., Quigley, A., Toniolo, A., Pucihar, K. Č., & Kljun, M. (2022). Vocabulary: Learning vocabulary in ar supported by keyword visualisations. *IEEE Transactions on Visualization and Computer Graphics*, *28*(11), 3748–3758. https://doi.org/10.1109/TVCG.2022.3203116

Yun, H., Choi, J., & Park, J. H. (2021). Prediction of critical care outcome for adult patients presenting to emergency department using initial triage information: an XGBoost algorithm analysis. *JMIR Medical Informatics*, *9*(9), e30770. https://doi.org/10.2196/30770

Zobair, K. M., Sanzogni, L., Houghton, L., Sandhu, K., & Islam, M. J. (2021). Health seekers' acceptance and adoption determinants of telemedicine in emerging economies. *Australasian Journal of Information Systems*. https://doi.org/10.3127/ajis.v25i0.3071