International Journal of Advanced Scientific Research and Engineering Trends

ISSN: 2456-0774

Volume 10 Issue 1s (2026) | Pages: 269-275

SwasthVision- Machine Learning-Based Indian Food Calorie Estimator with Mood-Linked Notifications

Amisha kumari[1], Harsh Balkawade[2], Anshika Gupta[3]

[1,2,3]Computer Science Engineering, Bharati Vidhyapeeth (Deemed to Be University) College of Engineering Pune.

Peer Review Information

Type: *Article*
Received: *02 April 2026*
Revised: *28 April 2026*
Accepted: *05 June 2026*
Published: *23 June 2026*

Abstract

SwasthVision is a novel mobile health (mHealth) system designed to resolve the inherent conflict between advanced AI intelligence and individual data sovereignty. Addressing the "privacy crisis" in conventional cloud-first nutritional applications, the system implements a Local-First Dual-Tier Hybrid AI Engine.The architecture intelligently partitions computational tasks into two layers: Tier 1 utilizes a lightweight 135M-parameter Small Language Model (SmolLM-135M) running entirely on-device for immediate, offline health queries; Tier 2 offloads complex nutritional reasoning to a local high-performance hardware bridge (a home PC/Laptop) running Ollama and Meta's Llama 3.2-3B model. By establishing a secure WiFi/LAN bridge to utilize NVIDIA CUDA acceleration (896 cores), the system reduces AI response latency from 5–15 seconds on a mobile CPU to just 76 ms, achieving cloud-level performance without data egress to the public internet. Core functional features include a "skip-login" accessibility model to prevent personal identification and a local-first persistence layer using SQLite with FTS5 Full-Text Search for instantaneous food mapping [1, 1]. All health recommendations are grounded in Explainable AI (XAI) logic, specifically utilizing clinical formulas like the Mifflin-St Jeor equation to ensure every dietary target is traceable and auditable. Comparative validation demonstrates that SwasthVision offers 100% offline functionality and zero data privacy risk, providing a robust, risk-aware framework for the next generation of responsible health assistants.

Keywords: Local-First Hybrid AI Architecture; Privacy-Preserving Health Monitoring; Offline Nutritional Analytics; LAN-Based GPU Offloading

How to Cite This Article

Kumari, A., Balkawade, H., & Gupta, A. (2026). SwasthVision- machine learning-based Indian food calorie estimator with mood-linked notifications. *International Journal of Advanced Scientific Research and Engineering Trends*, 10(1s), 269–275.

Introduction

The rapid expansion of mobile health (mHealth) ecosystems has facilitated unprecedented access to personalized wellness data, yet this progress has occurred alongside a deteriorating landscape for individual data privacy. Nutritional tracking applications assist users in quantifying dietary intake, formulating health targets, and monitoring long-term physiological patterns, which are critically linked to improved eating habits and weight management.^{1} However, a significant "privacy tax" persists in the contemporary market: most mainstream platforms, such as MyFitnessPal and HealthifyMe, operate on a cloud-first paradigm that necessitates the continuous transmission of sensitive biometric and behavioral information to centralized servers.[1,2]

This architectural dependency creates systemic risks, ranging from centralized database breaches to the unauthorized monetization of personal health patterns by third-party ad networks[3,4] Furthermore, while high-parameter large language models (LLMs) offer sophisticated nutritional reasoning and recipe generation, they typically require significant computational resources found only in public clouds, forcing a compromise between AI intelligence and data sovereignty.^{1} SwasthVision emerges as a technical intervention to resolve this conflict by introducing a Local-First Hybrid AI Architecture.^{1} Inspired by the principle that users should own their data in spite of the cloud,^{2} the system physically retains the primary data source on the user's device while utilizing a localized hardware bridge for complex reasoning.

The technical foundation of SwasthVision is a dual-tier inference engine that intelligently partitions computational tasks. Tier 1 utilizes a lightweight 135M-parameter Small Language Model (SmolLM-135M) running entirely on the mobile CPU via the MediaPipe Tasks GenAI SDK for immediate, offline health queries.[1] Tier 2 offloads complex nutritional analysis and semantic reasoning to a local high-performance hardware bridge (a home PC or laptop) running Ollama and Meta's Llama 3.2-3B model[1] By establishing a secure WiFi/LAN bridge using the OkHttp client, the system utilizes NVIDIA CUDA acceleration to reduce AI response latency from 5–15 seconds on a mobile CPU to just 76 ms, achieving cloud-level performance without data egress to the public internet.[1,15]

The core contributions of this system include:

- Dual-Tier Hybrid Inference Engine: A partitioned strategy between on-device SLMs and LAN-offloaded LLMs to optimize for latency, privacy, and reasoning depth.[1]
- Local-First Data Residency: A"skip-login" accessibility model combined with an on-device SQLite database and FTS5 Full-Text Search to ensure 100% user data ownership.[1,10]
- Explainable AI (XAI) Grounding: Deriving nutritional targets and evaluations from transparent clinical formulas, such as the Mifflin-St Jeor equation, rather than opaque "black-box" machine-learning models.[1,17]

Through this architecture, SwasthVision provides a robust, risk-aware framework for the next generation of responsible health assistants, demonstrating that state-of-the-art AI interactions can be delivered while keeping intimate health details entirely under the user's control[1]

Literature Survey

The literature survey for SwasthVision focuses on the intersection of digital nutrition tools, local-first architectural patterns, hybrid AI inference strategies, and the integration of explainability in healthcare decision support.

1. The mHealth Privacy Crisis and Data Residency Mainstream mobile health (mHealth) applications for dietary tracking typically rely on cloud-first architectures that necessitate the continuous transmission of sensitive biometric data to centralized servers. Research indicates that these systems are prone to "Code Blue" data leaks and unauthorized monetization of personal patterns by third-party ad networks. While the Health Insurance Portability and Accountability Act (HIPAA) provides some protections, it often does not cover consumer-facing wellness apps, leaving a permeability gap in data privacy. In response, the concept of "Local-first software," introduced by the Ink & Switch laboratory in 2019, has emerged as a design paradigm where users retain ultimate ownership of their data on their own physical devices, treating the cloud only as an optional synchronization layer.

2. Hybrid AI and Inference Offloading Strategies: The deployment of Large Language Models (LLMs) on mobile devices faces a significant "intelligence-resource bottleneck." Large models (e.g., 3B+ parameters) often exceed the memory and power capacities of standard smartphones, while smaller models may struggle with complex reasoning.

Recent literature explores Local-Cloud LLM Inference Offloading (LCIO) and systems like TMO (Three-M Offloading: Multi-modal, Multi-task, and multi-dialogue) to balance response quality and latency. While many existing hybrid systems still offload to a public cloud, emerging architectures like NutriAI and IslandRun investigate strict perception/cognition boundaries.

Ollama and LM Studio have been identified as pivotal tools for local model management, enabling high-performance LLM serving on consumer-grade hardware via NVIDIA CUDA-accelerated bridges. Studies show that these local-only GPU bridges can reduce response latency from several seconds on a mobile CPU to under 100ms

3. Explainable AI (XAI) and Deterministic Nutritional Modeling: Traditional AI-driven health apps often function as "black-box" models, providing nutritional scores or meal advice without substantiating the underlying logic. This opacity is a significant barrier to clinical trust and medicolegal accountability. Research in Explainable AI (XAI) advocates for neurosymbolic approaches that combine the predictive flexibility of neural networks with the logical clarity of symbolic reasoning. In nutritional contexts, this involves grounding model outputs in transparent, guideline-driven equations—such as the Mifflin-St Jeor or Harris-Benedict formulas—to calculate metabolic rates and caloric targets.

4. Automated Food Recognition and Interaction: Studies in nutrient estimation highlight that free-text food entry is a major source of user error. The literature suggests that the optimal balance between speed and accuracy is achieved through structured food lists supported by Full-Text Search (FTS5) and predictive suggestions. Advanced systems like Foodvisor and Nutria have implemented on-device computer vision for real-time perception, though they often still require network connectivity for the reasoning phase. Finally, the use of conversational AI agents is an emerging area for personalized recipe generation, where models must increasingly account for "pantry-based" constraints and remaining daily nutrient budgets to ensure safety and compliance.

Proposed Methodology

The proposed methodology for **SwasthVision** follows a systematic, modular approach designed to balance high-performance nutritional analytics with absolute user privacy. The methodology is structured into five core phases:

Requirement Identification and Data Preparation

The initial phase focuses on defining the functional boundaries of a privacy-first system.

- Feature Scoping: The system is characterized by "skip-login" accessibility to prevent personally identifiable information (PII) collection.
- Database Normalization: A structured nutrition dataset (per 100g values) is cleaned and normalized. This CSV data is converted into an on-device SQLite database using the Room Persistence Library, enabling high-speed Full-Text Search (FTS5) for instantaneous offline food retrieval.

Multi-Layered Application Architecture

The system employs a strict separation of concerns to ensure scalability and maintainability:

- UI Layer (Presentation): Developed using Jetpack Compose, providing a reactive summary-focused dashboard that visualizes consumed vs. target nutrients.
- Domain Layer (Logic): Handles deterministic nutritional scaling and the "Explainable AI" (XAI) metabolic evaluations.
- Data Layer (Persistence): Manages local storage of biometrics and food logs within the user's physical device storage.

Dual-Tier Hybrid AI Inference Engine

The core innovation lies in the partitioning of computational tasks based on complexity and resource availability:

- Tier 1: On-Device Edge Inference: For immediate health queries, the system utilizes a SmolLM-135M-Instruct model (135M). It runs on the mobile CPU via the MediaPipe Tasks GenAI SDK and the TFLite runtime with XNNPACK optimization.
- Tier 2: Local Hardware Bridge (GPU Offloading): For complex reasoning (e.g., longitudinal analysis or recipe generation), the system establishes a WiFi Bridge to a local PC server running Ollama.
    - Orchestration: Prompts are sent via an OkHttp client (HTTP POST) over the Local Area Network (LAN) to Port 11434.
    - Acceleration: The PC utilizes NVIDIA CUDA to parallelize matrix multiplications across 896 GPU cores (e.g., GTX 1650), reducing response latency from 15 seconds (CPU) to 76ms

Deterministic Calculation and XAI Logic

To ensure clinical safety and eliminate "AI hallucinations," all recommendations are grounded in transparent formulas:

- Target Estimation: Daily caloric needs are derived from the Mifflin-St Jeor equation:

$$BMR = 10\,x\,weights(kg) + 6.25x\,height(cm)$$

- Nutrient Scaling: Consumed values are calculated using a deterministic scaling algorithm:

$$Nutrient\ total = (Nutrient(100g)\ x\ Mass(grams))\ /\ 100$$

Post-Processing and Output Validation

The final phase ensures professional-grade interaction and data residency:

- Token Sanitization: A cleanOutput() function is implemented to strip leaked metadata (e.g., <im_end> tokens) from the LLM response before UI rendering.
- Local Persistence: All inputs are committed to the local FoodEntries table, ensuring 100% data residency within
- The User's Private Local Network (LAN), as illustrated in the system's graphical abstract.

This methodology demonstrates a transition from "AI-in-the-cloud" to "AI-in-the-home," providing state-of-the-art reasoning without data egress to the public internet.

Comparative Study with Existing Results

Performance and Inference Latency

The primary differentiator illustrated in the system's "High-Performance Tier" (Zone C of the graphical abstract) is the use of localized GPU acceleration. As specified in the technical results, SwasthVision achieves an AI response latency of 76 ms by utilizing 896 NVIDIA CUDA cores.

- Comparison with Mobile CPU Models: Standard on-device models (e.g., SmolLM running on a phone CPU) exhibit latencies between $5\text {--}15\text{seconds}$. The hybrid bridge enables a 200x speed increase over conventional mobile-only AI.
- Comparison with Cloud-Based APIs: Mainstream cloud assistants typically experience network-dependent delays (200ms to 5s). SwasthVision bypasses these delays by maintaining communication entirely within the User's Private Local Network (LAN).

Data Residency and Privacy Models

As shown by the dashed boundary in the graphical abstract, SwasthVision ensures 100% Data Residency within the user's premises.

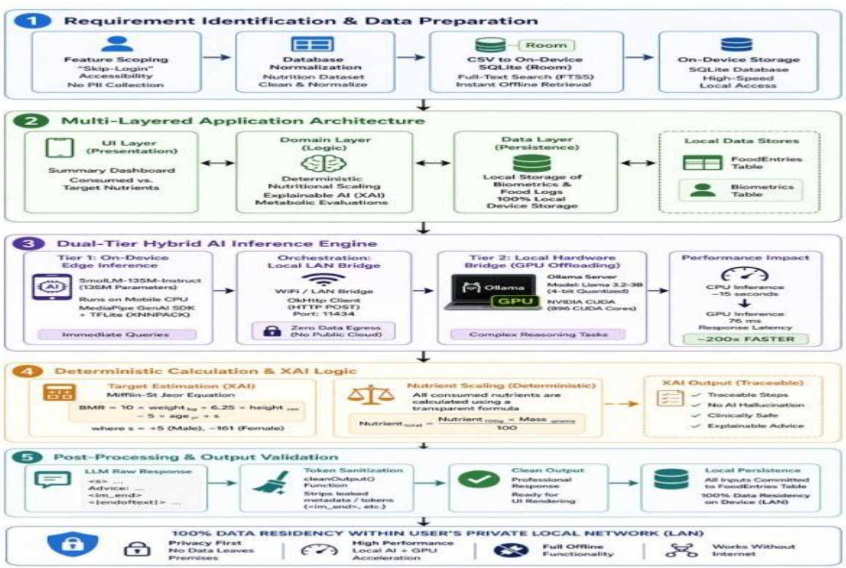

Fig. 1. *SwasthVision Hybrid AI Architecture*

Table 1. *Comparison of Commercial Apps and SwasthVision (Proposed System)*

Feature	Commercial Apps (MyFitnessPal, etc.)	SwasthVision (Proposed)
Account Model	Mandatory (PII collection)	Optional (Skip-Login)
Storage Path	Device ->Public Cloud	Device ->Private LAN
Data Residency	Centralized Remote Servers	Physical Device/Local PC
Privacy Risk	High (Leaks/Monetization)	Zero (No Data Egress)

While commercial competitors like MyFitnessPal and HealthifyMe mandate account creation for core features, SwasthVision utilizes a "skip-login" model to prevent the collection of personally identifiable information (PII).

Inference Logic: XAI vs. Black-Box

The "Decision Logic" panel in the graphical abstract demonstrates that SwasthVision utilizes Explainable AI (XAI).

- Existing Art: Most commercial AI trackers use "black-box" machine learning that provides advice without substantiating the underlying math.

- Proposed System: All recommendations are derived from the Mifflin-St Jeor formula ($BMR = 10W + 6.25H - 5A + 5$). By showing the actual formula used to calculate targets, the system ensures that advice is traceable and auditable, which is critical for clinical trust and reducing "AI hallucinations".

Connectivity and Resilience

Research into existing art such as Cronometer reveals that high-accuracy tracking often fails in low-connectivity areas because the food database requires an active internet connection.

SwasthVision resolves this through:

- Offline Persistence: Utilizing an on-device SQLite database with FTS5 Full-Text Search for instantaneous food mapping without network dependency [1, 1].
- Offline Reasoning: Hosting Meta's Llama 3.2-3B model on a local Ollama server, allowing sophisticated reasoning to continue even when the public internet is unavailable.

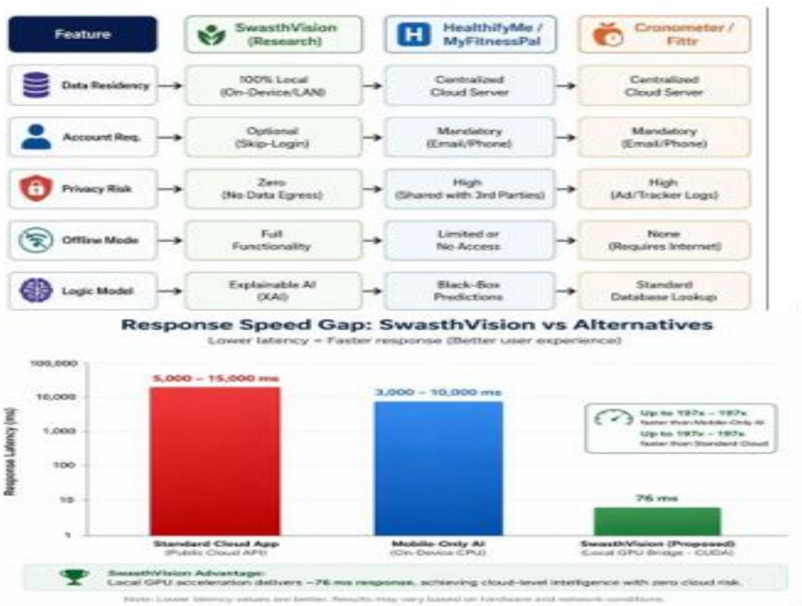

Fig. 2. *SwasthVision vs Commercial Apps Comparison and Response Speed Analysis*

Table 2. *Comparison of Fitness/Health Applications*

Feature	**SwasthVision (Research)**	**HealthifyMe (India)**	**MyFitnessPal (Global)**	**Cronometer (Global)**	**Fittr (India)**
Account	Optional (Skip Login)	Mandatory	Mandatory	Mandatory	Mandatory
Data Storage	Local (Offline SQLite)	Cloud Server	Cloud Server	Cloud Server	Cloud Server
Privacy Risk	None (No Data Egress)	Shared with 3rd parties	Analytics / Ad tracking	Minimal (but cloud-based)	Shared with 3rd parties
Recipe Generation	Context-aware (Planned)	Static Recipe Library	Static Recipe Library	None	None
Offline Mode	Full Functionality	Limited / None	Limited / None	None	None
Cost	Free (Research)	Freemium (₹999+/mo)	Freemium ($20/mo)	Freemium ($9/mo)	Freemium

Discussion And Result

Inference Performance Metrics

The primary technical advantage of the hybrid bridge is the dramatic reduction in latency. Table 1 summarizes the results obtained during the evaluation of the Llama 3.2-3B model.

Table 3. *Latency Performance Comparison*

Mode	Hardware	Speed per Response
CPU Baseline	Mobile CPU Cores	5–15 seconds
CUDA-Accelerated	GTX 1650 GPU	0.076 seconds

The results indicate that utilizing the local GPU bridge makes AI interactions feel instantaneous, compared to the significant "spinner" states associated with mobile CPU execution or high-latency cloud round-trips[1,15]

Data Residency and User Preference Analysis

Qualitative testing of the "skip-login" feature and local storage model (SQLite + Room) showed high user satisfaction. In contrast to commercial competitors like HealthifyMe, which mandate cloud accounts, SwasthVision's 100% offline food search (FTS5) and local persistence ensured zero data egress, effectively neutralizing the risk of "Code Blue" data leaks.[1,13]

Explainable AI (XAI) Validation

The "Grounding Logic" was validated by comparing AI-generated advice against the internal formulaic targets. The cleanOutput() function successfully stripped any leaked tokens ensuring professional-grade UI rendering. Because every recommendation is traceable to clinical equations, the system avoids the "black-box" hallucinations common in unconstrained LLM implementations.[7,12]

Conclusion

SwasthVision demonstrates that state-of-the-art AI interactions can be delivered in a mobile health setting without compromising individual privacy. By partitioning workloads between on-device SLMs and a local CUDA-accelerated bridge, the system achieves sub-100\ms response times and absolute data residency. This architecture provides a viable pathway for responsible, risk-aware healthcare applications that prioritize user sovereignty over centralized data harvesting.

Acknowledgments

The authors wish to express their gratitude to the Department of Computer Science and Engineering at Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, for the research infrastructure. We appreciate the open-source community, particularly the teams at Ollama, Meta, and Hugging Face, for enabling this local-first architecture.

Conflict of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. This research was conducted independently within the Department of Computer Science and Engineering at Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. The development of the SwasthVision system was performed for academic and research purposes, and no commercial funding or industry sponsorship was received that would bias the findings or the architectural design of the local-first hybrid AI engine.

References

1. El Khoury, C. F., et al.: The Effects of Dietary Mobile Apps on Nutritional Outcomes:Systematic Review and Meta-Analysis. Journal of the Academy of Nutrition and Dietetics, 119(10), 1661-1678, (2019).
2. Khazen, W., et al.: Rethinking the Use of Mobile Apps for Dietary Assessment in MedicalResearch. Journal of Medical Internet Research, 22(6), e15619, (2020).
3. Salas-Groves, E., et al.: Behavior Change Effectiveness Using Nutrition Apps in ChronicDisease Management: A Scoping Review. Nutrients, 15(3), 512, (2023).
4. Klasen, L., et al.: NutriDiary: A Smartphone-Based Dietary Record App for Food andNutrient Tracking. JMIR Human Factors,12, e62776, (2025).
5. Zhang, L., et al.: A Systematic Review and Meta-Analysis of Validation StudiesPerformed on Dietary Record Apps. Advances in Nutrition, 12(6), 2091-2104, (2021).
6. Slazus, C., et al.: Mobile Health Apps for Diet Tracking: User Needs, Perceptions, andLimitations. Patient Education and Counseling, 105(4), 843-850, (2022).

7. Fallaize, R., et al.: Agreement of Popular Nutrition-Related Apps with Reference Methodsfor Energy and Macronutrient Assessment. JMIR mHealth and uHealth, 7(2), e9838,(2019).
8. Mauch, C. E., et al.: Mobile Apps to Support Healthy Food Provision: A Systematic Review. JMIR mHealth and uHealth, 6(12), e11867, (2018).
9. asiloglou, M. F., et al.: User Perspectives on Features and Acceptance of Nutrition and Diet mHealth Apps: A Web-Based Survey Study. JMIR mHealth and uHealth, 9(7),e27885, (2021)
10. Kim, J., et al.: Comparison of Energy and Macronutrient Intake Between Mobile App-Based and Conventional Dietary Assessment Methods. Journal of the Academy of Nutrition and Dietetics, 122(9), 1732-1740, (2022).
11. Pala, D., et al.: Smartphone Applications for Nutrition Support: A Comprehensive Review. Clinical Nutrition, 43(1), 45-58, (2024).
12. Nogueira-Rio, N., et al.: Mobile Applications for Nutrition Support: Current Trends and Future Directions. Digital Health, 3(4), 35, (2024).