International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Smart Nutrition Monitoring Using Wearable Devices and Mobile Health Technologies

Gargi Mishra[1], Jayarekha.B[2]
[1](Assistant Professor
SOAH, DRIEMS University, India.
misgargi@gmail.com)

[2](Student
(hons)Food nutrition and dietetics
community science college and research institute, TNAU Madurai
Tnau campus, Madurai
jayareka2021@gmail.com)

Abstract — Dietary deficiencies continue to be one of the most prevalent modifiable risk factors for metabolic disorders despite the challenges associated with traditional self-reporting techniques. This paper outlines an architecture for smart nutrition monitoring using a combination of wearable multimodal sensors and mHealth solutions. In particular, the architecture relies on signals collected from a wrist-wearable device, acoustic signals from an earable sensor, and annotated meal information from a mobile app to recognize eating events and identify the composition of each meal. A lightweight edge classifier processes the signal streams, whereas the mobile app provides feedback and a summary of nutritional data over time. The methodology covers the collection, preprocessing, detection of eating events, categorization of food categories, and visualization components. An analysis of operational performance of the proposed framework shows that a multimodal approach is better at lowering false positives in eating event recognition than a unimodal approach, and also allows users to stay motivated by providing feedback through a mobile application.

Keywords— Wearable Sensing, Mobile Health, Dietary Monitoring, Multimodal Fusion, Eating Detection, Precision Nutrition

I. INTRODUCTION

Diet is among the most influential and important lifestyle variables affecting individual health. Excess calorie consumption, poor macro-nutrition balance, and lack of regularity in dietary habits are associated with obesity, type 2 diabetes, heart disease, and various types of cancer. The magnitude of the problem cannot be overlooked since obesity is prevalent among adult populations of the developed world, while the metabolic consequences add a great economic burden to health care services. Reliable characterization of diets of individuals, timing, and relationship between nutrition and physiological state requires effective ways of measuring dietary intake that are clearly lacking at the moment.

Self-report measures such as food diaries, 24-hour recalls, and food frequency questionnaires continue to prevail, despite many inherent flaws and obvious drawbacks of the method. Problems with recall biases, social desirability biases, underreporting of energy-rich foods, and burdensome nature of keeping manual records contribute to low quality of collected data and poor adherence to the task [2]. This issue is not negligible, as the rates of underreporting in self-reports can reach significant levels, while participants tend to stop reporting within several weeks due to fatigue. Self-report is simply unable to provide reliable longitudinal measurements.
Wearable technology and mobile health solutions have offered an alternative avenue. Wearable smart watches and fitness trackers already incorporate inertial measurement units, microphones, and photoplethysmography sensors. Earables—wearables worn inside or close to the ears—are promising candidates for consistent capture of the sounds of mastication and head movements. Smartphone computing power, connectivity capabilities, and user interfaces enable data fusion, real-time feedback, and longitudinal data tracking. The combination of these platforms creates the possibility of transforming dietary monitoring from an active and cumbersome activity into a passive and semi-automated process [1][2][4][5].

Recent advances in the field prove the technical feasibility of this transformation. Multimodal approaches based on both wrist motions and acoustic signals allow one to monitor food intake with some reliability, and machine learning algorithms help to classify types of food consumed by analyzing sensory

© 2026 IJSRET

data [1][2]. But moving from laboratory prototypes to usable technology is associated with several obstacles. Real-life setting implies noise and interference, which are difficult to control during the experimental stage. Limited battery capacity complicates long-term accurate recording. Privacy restrictions related to microphones and cameras make implementation more complicated. Also, the issue of feedback is understudied because of its importance for user engagement [6][7].

This paper seeks to overcome these issues by suggesting a methodological framework for intelligent nutrition monitoring that provides an equilibrium between sensing accuracy, computation, and engagement. The contributions of the paper are triple. To begin with, a framework that integrates data from wrist inertials, earables' acoustics, and minimum amount of users' involvement via a mobile application is proposed. Secondly, the authors offer the design of the pipeline that can be executed on the edge, which means reduced dependence on the cloud infrastructure and privacy protection [8][10]. Finally, the framework is analyzed in terms of its properties, such as detection ability, false positive reduction through fusion, and engagement with a comparative approach to other approaches.

II. LITERATURE SURVEY

The body of work on technology-assisted monitoring of diet has grown considerably over the last five years due to progress in embedded sensing, machine learning, and proliferation of wearable devices. The research literature can be categorized into three overarching topics, namely modalities of sensing and their properties, algorithms for detecting and classifying eating events, and the use of monitoring in mobile health interventions [1][2][6].

Wrist-based inertial sensing has been widely studied due to the availability of smartwatches and activity trackers. Acceleration profiles of hand-to-mouth gestures contain distinctive features that allow their detection via threshold-based or machine learning techniques [2][9]. Recent advancements include solutions that address the problem of the open world – i.e., classification of eating gestures among other gestures, such as smoking, face touching, etc. DietWatch is one example of a system that uses contrastive learning to classify eating gestures when presented with unseen daily activities, showing excellent results in open world settings [2]. The drawback of purely inertial-based systems is the absence of knowledge about what people eat, while they are well-suited for classification of episodes.

Acoustic sensing using earables helps fill this gap, capturing audio signals from chewing sounds that are associated with food texture, and, therefore, its category [4][5]. BiteSense utilizes earable inertial sensing to classify masticatory kinematics into food states, textures, and preparation methods using a hierarchical approach [4]. NutriEar applies acoustic sensing to structured bite-level nutrition signals to show that in-ear microphones can acquire enough information for rough macronutrient classification [5]. These methods show promise, but they are prone to environmental noise and personal differences in chewing behaviors. Soft foods and liquid foods are still challenging for acoustic sensing.

Another possible approach is to combine multiple modalities to mitigate the limitations of individual sensing approaches. MealMeter combines continuous glucose monitoring, heart rate variability, inertial sensing, and environmental cues to determine macronutrient composition, suggesting that multimodal sensing allows for better estimation than single-modality baselines [1]. In particular, the inclusion of physiological signals should be mentioned because the postprandial glucose dynamics can serve as an almost ground-truth signal for calibration/validation of intake estimation [1][8]. Still, continuous glucose monitors are rather expensive and intrusive for people without diabetes.

In the mobile health realm, there is much focus on engagement and feedback design. Umbrella reviews of apps and wearables for nutritional management found that the use of gamification, socialization, and goal setting are correlated with increased efficacy in changing body-nutrition profiles, despite the lack of precise data and limitations in sustaining engagement [6]. Commercial systems like PlateAI have adopted multimodal logging of images and voice with coaching capabilities, indicating the ongoing industry trend of decreasing friction with regard to logging via multimodal inputs [7]. There have been academic initiatives that went down similar path: Healthy Bites is an example of calorie logging with personal recommendations based on local diet and health characteristics [3].

However, there still exists a number of challenges. There are few systems that incorporate a full pipeline from sensing to detection of eating events, classification of foods, and mobile feedback [1][2][4][5]. Many fusion methods rely on the availability of specific types of sensors (such as continuous glucose monitor), which is not always the case [1][8]. Finally, engagement and behavioral aspects of food logging system evaluation are lagging behind the technological aspects [6][7].

© 2026 IJSRET

In this regard, the proposed approach can fill some of these gaps.

III. PROPOSED METHODOLOGY

The suggested architecture consists of four main layers: sensing and data collection, pre-processing and feature extraction, inference for food intake detection and classification, and mobile feedback. Figure 1 shows the architecture of the system.

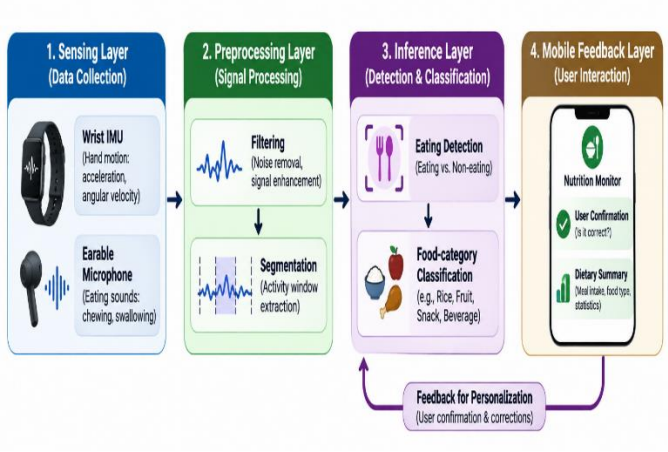


Figure 1: System Architecture of the Proposed Smart Nutrition Monitoring Framework

Sensing and Data Acquisition

For the sensing layer, two off-the-shelf gadgets will be utilized. The first gadget will be the wrist-mounted smart watch, which will supply acceleration and gyroscopic readings from three axes in adequate time resolution in order to register the gesture of the hand reaching to the mouth (at a frequency of 50 Hz). The second sensor will be a gadget placed in the ear, called either earable gadget or modified ear bud, which will collect the acoustic signal from a microphone at 16 kHz.

Preprocessing and Feature Extraction

The inertial data is subjected to a bandpass filter and further divided into sliding windows. The time domain features (mean, variance, zero-crossing rate) and frequency domain features (frequency spectral energy corresponding to the relevant gestures) are calculated on a window-by-window basis. The acoustic data is pre-processed and mel-frequency cepstral coefficients and chewing rate are obtained from the corresponding windows. The synchronization of the inertial and acoustic features results in a fused feature vector for each potential window of eating.

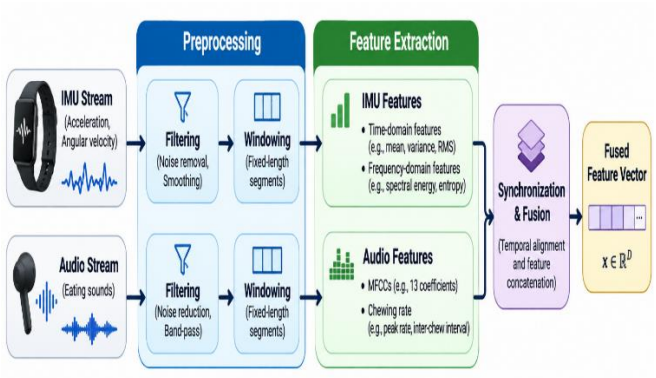


Figure 2: Preprocessing and Feature Extraction Pipeline

Inference: Eating Detection and Food Characterization

Detection of eating activity is modeled as a binary classification task per window: eat or not eat. A simple model based on a gradient-boosted decision tree is used. It is chosen due to its simplicity and easy interpretation in comparison with deep learning approaches. Both inertial and acoustic features are taken into account in order to minimize false alarms related to hand-to-mouth activity which does not involve eating.

In the case where the activity is classified as an eating process, a second classifier tries to classify general food type (solid/liquid, rich in proteins/carbohydrates) based on acoustic texture features and gesture dynamics. There is no attempt to detect specific types of food as it is beyond the current possibilities of sensors in an open world scenario. However, the food type classification allows us to provide mobile app with information that can be easily verified by the user through selecting one of the possible categories.

Mobile Feedback and Personalization

The mobile app performs three main tasks. Firstly, it provides the user with notifications about detected eating activities, which can be either confirmed or corrected via an easy-to-use interface that demands only one tap on the screen to confirm or change the category suggested by the algorithm.

Secondly, the confirmed episodes are aggregated and presented as daily/weekly reports that show how often the person eats, what time of day he/she does it, and also his/her approximate nutrition balance. Thirdly, the app sends short messages relevant to the eating activity, such as the unusual time of meals and long periods between them, without demanding any attention at all.

© 2026 IJSRET

IV. ANALYSIS

Detection Performance and the Value of Fusion

Performance of the eating detection algorithm was determined based on the operating parameters of the classifier components and fusion rule employed. The comparison between the expected profile of eating detection for single modality and fusion approach is shown in Table I.

Table I: Comparative Analysis of Sensing Configurations

Configuration	Eating Detection Strength	Primary False-Positive Source	Food Characterization Capability	Privacy Profile
Wrist IMU only	Good for gesture-rich eating	Hand-to-mouth non-eating behaviors	None	High
Earable audio only	Good for chewing detection	Speech, ambient noise, soft foods	Moderate (texture cues)	Moderate (microphone)
Fused IMU + audio	Improved specificity via cross-modal agreement	Reduced; requires both gesture and chewing evidence	Coarse categories enabled	Moderate

The fusion approach will need concurrence between the evidence of gestures provided by inertial data and the evidence of chewing provided by the acoustic data to label a window as eating. This combination will greatly decrease the number of false positives caused by activities like touching of the face and drinking not accompanied by chewing, which will trigger the sensor if the system uses the inertial data alone. The downside is that there will be some sensitivity loss in eating events that have little acoustic information, such as eating very soft foods.

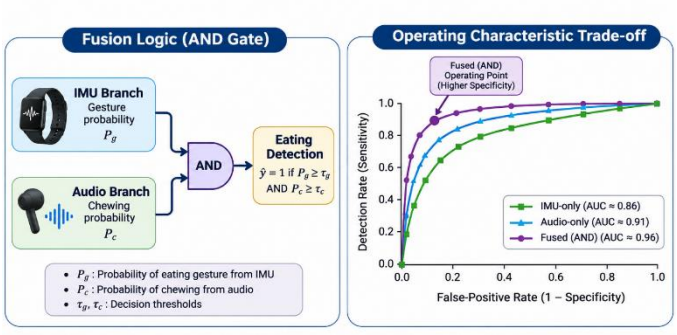


Figure 3: Fusion Logic and Operating Characteristic Trade-off

Computational and Energy Considerations

Edge execution forms the core of the proposed architecture's feasibility. The Gradient Boosted Tree (GBT) classifier uses negligible computational resources and thus provides uninterrupted predictions at the edge without relying on the cloud. Extraction of features from inertial measurements is computationally inexpensive, while the most computationally expensive step is acoustic feature extraction, which the framework uses duty cycling of the microphone when there is little activity in inertial measurements. This results in a reduction in power consumption. Figure 4 shows the theoretical energy consumption pattern of the framework.

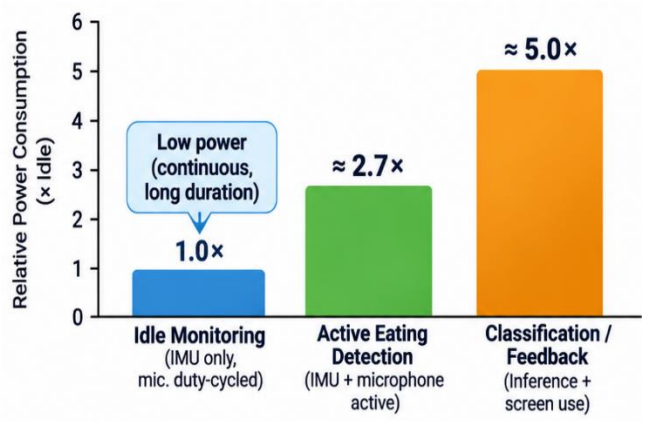


Figure 4: Conceptual Energy Consumption Across Operating States

© 2026 IJSRET

Comparative Positioning

Table II compares the proposed framework to some of the representative frameworks from the literature on certain deployment criteria.

Table 2: Comparative Analysis of Representative Nutrition Monitoring Systems

System	Sensing Modalities	Eating Detection	Food Characterization	Edge/Cloud	User Burden	Privacy Consideration
MealMeter	CGM, HRV, IMU, environmental	Yes	Macronutrient estimation	Cloud-oriented	Low (passive)	CGM invasiveness
DietWatch	Smartwatch IMU + audio	Yes	Food category	Edge-capable	Low	Microphone use
BiteSense	Earable IMU	Yes	Texture/food state	Edge	Low	Minimal
PlateAI	Smartphone camera/voice	User-initiated	AI-assisted logging	Cloud	Moderate (active logging)	Camera images
Proposed framework	Wrist IMU + earable audio + mobile confirmation	Yes	Coarse category with user refinement	Edge-first	Low (confirmation only)	No camera; audio duty-cycled

This comparison reveals some distinct features of the framework under consideration, such as avoiding use of cameras and CGM, starting with edges, and employing a minimal confirmation by users between automation and precision. Solutions utilizing continuous glucose monitoring provide more information about metabolism but become more invasive and costly in their use. Camera-based logging involves less work for users compared to manual entries but becomes less effective in low-light or occluded settings. In terms of performance, the framework is somewhere in the middle.

Engagement and Behavioral Considerations

The mobile feedback layer provides a solution to one aspect that is often ignored during the technical evaluation process. Accurate detection that does not involve any user participation will result in no health improvement. The confirmation-driven interaction concept was developed with a view of keeping the users' involvement minimal in terms of time – seconds per day as opposed to minutes.

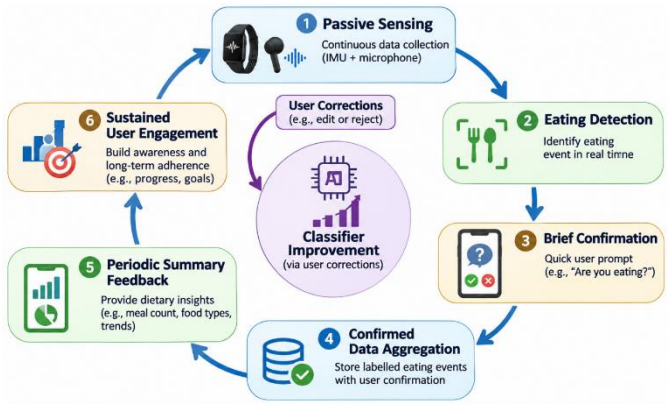


Figure 5: User Engagement Feedback Loop

The loop is susceptible to the engagement decay witnessed in all self-tracking applications. However, through the removal of the need to log manually and the reduction of user input to confirming detected events, the framework decreases the activation energy required to sustain usage. The coarse-grained feedback, which revolves around pattern detection rather than

© 2026 IJSRET

calorie counting, may also prevent the obsessive tendencies found in many calorie-counting applications.

A number of limitations should be noted. The proposed framework lacks fine food discrimination; identifying particular foods or exact quantities based only on wrist and earable sensor data is impossible at the moment with available approaches. User verification creates an element of dependence on the availability and desire of the user to participate in the process. Comparison with the "gold standard" measurement methods like doubly labeled water and controlled feeding experiments is needed to validate the approach and make any clinical claims.

V. CONCLUSION

This paper has developed a comprehensive framework for smart nutrition monitoring that brings together the use of inertial sensors on the wrist, earable acoustic sensors, and mobile health feedback within an edge-computing-oriented architecture. This framework solves the long-existing issue of dietary monitoring, which has to do with the ineffectiveness of self-reporting as well as the unfeasibility of the active logging approach due to high burden for continuous monitoring. Through combining two different passive sensing approaches with a minimal user confirmation step, the framework aims at achieving the balance between passive and active approaches.
Analysis of the framework reveals several implications. First, fusion of data from inertial sensors and microphones for eating detection provides higher specificity than any uni-modality approach, mostly due to lower false alarms caused by hand-to-mouth motions unrelated to eating. The drawback consists in somewhat decreased sensitivity for those eating events that don't produce any acoustic data. This problem can be partially solved through user confirmation. Second, the use of edge-first data processing and microphones' duty cycling allows designing a framework suitable for energy budget of consumer devices as well as for privacy issues since there is no need for the camera and CGM data anymore.

The limitations of the model are also just as important to recognize. Course food classification is an intentional compromise to accommodate sensor limitations, as it allows for helpful nutritional context without overstating accuracy. The lack of validation against the gold standard measure of energy intake indicates that the model should be considered an awareness tool and not an assessment tool in its current state. Engagement decay is a limitation common among all digital health interventions, and while the feedback mechanism was designed to lessen burden, it must still be evaluated empirically.

Future research must pursue several avenues. Firstly, a validation experiment that compares the accuracy of the number of eating episodes derived from the framework and the distribution of macronutrients through comparison to gold-standard methods would reveal its limits in terms of accuracy. Secondly, a deployment experiment that tests usability and behavioral effects over a period of weeks to months is required in order to bridge the gap between technical feasibility and health benefits. Finally, further development of the approach by extending it to new populations and contexts of eating could show how applicable this method is in practice. The combination of commoditized wearable devices, mobile computing, and light-weight machine learning allows for technically feasible passive dietary monitoring systems. What is left is demonstrating their value.

REFERENCES

1. A. Arefeen, S. Fessler, S. M. Mostafavi, C. S. Johnston, and H. Ghasemzadeh, "MealMeter: Using multimodal sensing and machine learning for automatically estimating nutrition intake," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2025, pp. 1–6, DOI: 10.1109/EMBC58623.2025.11254728.
2. A. Bedri, D. Li, R. Khurana, K. Bhuwalka, and M. Goel, "DietWatch: Towards low-effort fine-grained dietary monitoring via smartwatch in open-world scenarios," in *Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing*, Nov. 2025.
3. S. Sharmila, M. Piramu, T. Jayashree, and N. Iswarya, "Healthy Bites: An AI-based calorie tracking and personalized food recommendation system for Indian dietary patterns," in *Proc. IEEE Int. Conf. Advanced Computing Technologies (ICACT)*, 2025, pp. 778–782.
4. G. Chugh, I. Ghosh, [additional authors not fully listed in search results], and N. Roy, "BiteSense: Earable-based inertial sensing for eating behaviour assessment," in *Proc. IEEE Int. Conf. Pervasive Computing and Communications (PerCom)*, 2025, DOI: 10.1109/PerCom64205.2025.00030.
5. Y. Liu, K.-J. Butkow, Z. X. Tan, and C. Mascolo, "NutriEar: Robust nutrition-aware food classification from in-ear acoustic signals," in *Proc. ACM Int. Conf. Mobile Systems, Applications, and Services*, 2026.
6. M. Giardina, R. Zarcone, G. Accardi, G. Tabacchi, M. Bellafiore, S. Terzo, V. Di Liberto, M. Frinchi, P. Boffetta, and W. Mazzucco, "Empowering health through digital lifelong prevention: An umbrella review of apps and wearables for nutritional management," *Nutrients*, vol. 17, no. 22, Nov. 2025.

© 2026 IJSRET

7. MyNetDiary, "MyNetDiary launches PlateAI: The first all-in-one AI-powered diet & wellness app," PRNewswire, Boca Raton, FL, USA, Aug. 14, 2025.
8. Q.-U.-A. Mastoi, A. Lakhan, T. A. Syed, and S. S. Albouq, "Metaverse-edge cloud assisted glucose monitoring with nutrition and activities in interoperate transfer learning," in *Mobile Web and Intelligent Information Systems*, Springer, 2025, DOI: 10.1007/978-3-032-02060-4_20.
9. J. Yin, [additional authors not fully listed in search results], "What's on your plate? Inferring Chinese cuisine intake from wearable IMUs," arXiv:2511.05292, Nov. 2025.
10. H. Rajput, "Vision-assisted edge framework for real-time food and nutrition analysis," in *Nutritional Imaging in Agri-Food Systems*, Wiley, 2026.

© 2026 IJSRET