International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 10 | Oct 2026 www.irjet.net p-ISSN: 2395-0072

NutriSense: A Multimodal Sensor Fusion Framework for Ingredient-Level Nutrition Tracking in Home Kitchens

B. Sri Hari[1], K. V. S. S. Dhanush[2], N. Nehalika[3], P. V. S. Rohith[4], Anitha K[5]

[12345]*Department of Artificial Intelligence and Machine Learning, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India*

Abstract - *Keeping a food diary is difficult during home cooking because ingredients are combined, portions are shared, and individual ingredients may no longer be available for separate measurement after preparation. NutriSense addresses this problem through a multimodal sensing framework that observes the cooking process rather than relying on manual dietary logging. The system combines an overhead ESP32-CAM with a YOLOv8-based detector for identifying ingredients and containers with a four-point load-cell platform connected through HX711 amplifiers for measuring changes in mass. A FastAPI processing engine aligns the visual and weight streams using a finite-state event model consisting of idle, motion, stable, and logged states. Detected mass changes are associated with ingredient identities and converted into estimated nutrient values using food-composition databases. A Llama-3-70B-based conversational module hosted through Groq is then used to provide personalised dietary guidance based on the logged intake and stored user profiles. A controlled salad-preparation session was used to demonstrate the complete workflow, including ingredient tracking, dish composition, serving allocation, nutrition estimation, and dashboard reporting. The prototype demonstrates the feasibility of combining computer vision and direct mass measurement for ingredient-level nutrition tracking during meal preparation.*

Keywords - sensor fusion, computer vision, load cell, YOLOv8, dietary assessment, Internet of Things, personalized nutrition, smart kitchen.

I. INTRODUCTION

Phones and watches now count our steps, sleep and heart rate with almost no effort on our part. Food is different. Most people still have to weigh, record, search and calculate by hand, and most give up within days [9]. Home cooking makes this worse. Ingredients are chopped and mixed in one vessel, then shared out among several people, so by the time the dish is done there is nothing left to weigh separately.

Existing approaches each leave part of the problem open. Image-based systems estimate portion size from a photo of the finished plate, but converting volume to mass depends on food density, camera angle and whatever the cookware or garnish happens to hide. The estimate also arrives only after the food is already made [2], [8], [19], [26]–[28]. Smartwatch-based tools raise awareness, but the user still has to remember or confirm what they ate [9]. IoT calorie trackers put sensors in the kitchen, yet they usually report what the food is or how much there is, rarely both [11], [29], [30]. Multi-agent systems built on large language models reason well once a meal has been logged, but they take the logged data as correct [3], [4], [7].

NutriSense addresses the gap between food recognition and quantity measurement by observing the cooking process rather than relying solely on the finished meal. The main contributions of the system are:

(1) a multimodal sensing architecture that combines an overhead camera for ingredient identification with a multi-point load-cell platform for measuring ingredient quantity;
(2) a finite-state event-processing mechanism that converts continuous visual and weight observations into discrete ingredient-transfer events;
(3) a running ingredient ledger that records the composition of a dish and supports subsequent serving-level allocation;
(4) integration of a nutrition lookup layer that converts measured ingredient quantities into estimated nutrient values;
(5) integration of a large-language-model-based conversational module that uses the logged dietary information and user profiles to provide personalised guidance; and
(6) a working prototype demonstrating the complete pipeline from ingredient preparation to serving-level nutrition reporting.

A salad-preparation prototype demonstrates the whole pipeline at Technology Readiness Level (TRL) 4–5.

The rest of the paper is organised as follows. Section II surveys related work, Section III presents the method and its mathematics, Section IV describes the implementation, Section V gives the prototype demonstration, Section VI covers limitations, and Section VII concludes.

II. RELATED WORK

Image-Based Recognition and Volume Estimation Much of the literature uses convolutional or transformer backbones to recognise food and estimate its volume from single or multiple images. Examples include multi-model classification and volume pipelines [2], transformer-based mass estimation [13], boundary-aware recognition

networks [19], VGG-based classifiers with Grad-CAM explanations [12], and volume estimation for powdered or irregular substances [8], [10]. Related vision work also reaches into food quality, for instance segmentation with out-of-distribution-aware classification for meat freshness [1]. Recent surveys [26], [27] and multiview geometric work [28] agree that recognition accuracy is now very high. Portion mass, however, is still inferred rather than measured, so density variation and occlusion keep causing errors.

TinyML and Edge Health Monitoring

Another line of work moves inference onto microcontrollers to protect privacy and cut latency. Examples include federated TinyML for medical IoT devices [5], TinyML surveys across healthcare, education and transport [6], [15], edge-centric health monitoring [16], unsupervised stress and sleep monitoring [21], assistive devices [25], and smart-city deployments [20]. They show that low-power inference at the sensor is practical, but none of them targets ingredient-level dietary composition.

LLM-Driven Personalised Nutrition

Several reviews look at how generative AI and language models can support dietary counselling [3], [4], [17], [22], and one group has proposed a closed-loop multi-agent architecture for meal-level nutrition management [7]. Multi-agent designs have also been explored in other health settings, such as mental healthcare [14]. These systems give context-aware advice, but they typically start from a meal record that someone typed in, not one that was sensed.

IoT Calorie and Nutrition Tracking

Real-time calorie trackers [11], nutrition monitoring platforms for preventive care [29], edge-based food recognition on low-resource devices [30], CNN-based recognition and tracking [23], and reviews of IoT monitoring for clinical populations [18], [24] show that connected sensing can scale. What they generally lack is a direct, ingredient-level mass measurement tied to visual identity.

Table I compares NutriSense with these approaches on four practical dimensions: how portion mass is obtained, the granularity of the record, the burden on the user, and the degree of personalisation.

TABLE I COMPARATIVE POSITIONING OF DIETARY-SENSING PARADIGMS

Approach	Mass Measurement	Granularity	Personalization
Image-only CV / volumetric [2], [8], [10], [19], [26]–[28]	Inferred from image volume	Meal / dish level	Limited
Wearable / smartwatch tracking [9]	Not measured (self-report)	Meal level	Limited
IoT calorie trackers [11], [23], [29], [30]	Sensor-dependent, often none	Meal level	Limited
LLM multi-agent nutrition [3], [4], [7], [17], [22]	Not sensed (user input)	Meal level	High (reasoning only)
TinyML / edge health [5], [6], [15], [16], [20], [21], [25]	Varies by application	Not food-specific	Low–Medium
NutriSense (proposed)	**Direct (load cell)**	**Ingredient + serving**	**High (LLM + profiles)**

III. METHODOLOGY AND MATHEMATICAL FORMULATION

NutriSense is a five-layer pipeline (Fig. 1): sensing, communication, processing engine, data, and application.

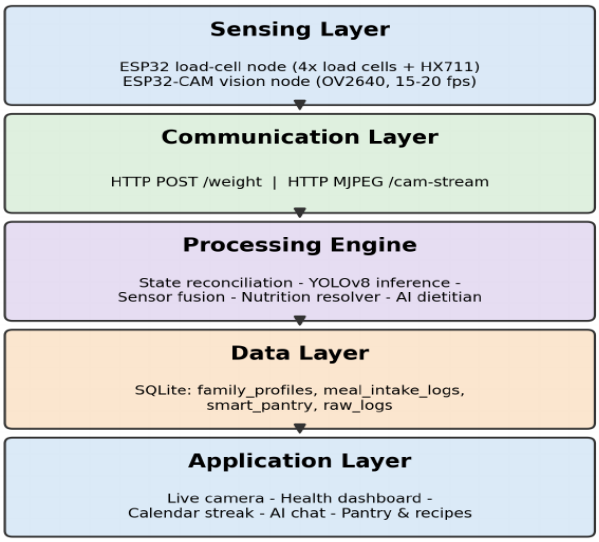


Fig. 1. Five-layer system architecture of NutriSense.

Conditioning the Weight Signal

Raw load-cell readings are noisy, so we smooth W(t) with a moving average over N samples:

$$W_{avg} = (W_1 + W_2 + \ldots + W_N) / N \quad (1)$$

We treat the platform as settled, and declare an ingredient-transfer event, once the variance within the window drops below a threshold σ_{th}^2:

$$Var(W) = (1/N) \Sigma_{i=1}^{N} (W_i - W_{avg})^2 < \sigma_{th}^2 \quad (2)$$

The processing engine follows this logic through four states: IDLE → MOTION → STABLE → LOGGED. The mass belonging to the k-th event is the signed difference between the stable readings before and after it:

$$\Delta W_k = W_{after} - W_{before} \quad (3)$$

Sensor Fusion

For each detected event, the vision component provides the most likely ingredient class C_k together with its detection confidence $p_{v,k}$. The weight-processing component determines whether a corresponding change in platform mass occurred and whether the measurements before and after the event satisfied the stability condition.

The two observations are combined to determine whether an ingredient-transfer event should be recorded. In the current prototype, the visual and weight observations are given equal importance:

$$F_k = w_v p_{v,k} + w_m p_{m,k} \quad (4)$$

where $w_v + w_m = 1$, and $p_{m,k}$ represents the confidence associated with the weight-domain observation. The prototype uses equal weighting as a simple fusion heuristic rather than as an empirically optimized configuration. An event is recorded when the combined evidence satisfies the selected event threshold.

This fusion mechanism is intended to reduce cases where either sensing modality produces an observation without sufficient evidence from the other modality. For example, a visual detection without a corresponding mass change can be treated differently from an observation in which both visual and weight changes occur simultaneously.

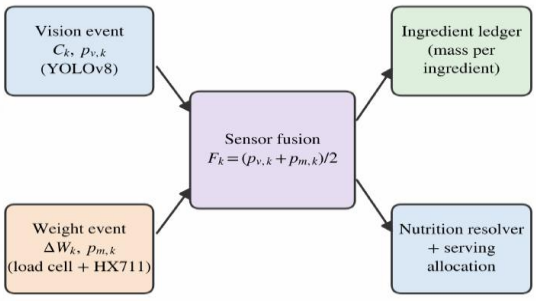


Fig. 2. Fusion of a visual identity event and a weight-change event into a single ingredient-portion record.

Nutrition and Serving Calculations

Once an event is accepted, its mass is converted to nutrient content i using the per-100 g value $n_{i,Ck}$ from the ICMR-NIN Indian Food Composition Tables or USDA FoodData Central for class C_k:

$$N_{i,k} = (|\Delta W_k| / 100) \times n_{i,Ck} \quad (5)$$

Nutrient totals for the dish are the sum over all K accepted events:

$$N_{i,total} = N_{i,1} + N_{i,2} + \ldots + N_{i,K} \quad (6)$$

At serving time, the drop in finished-dish weight when a portion is removed for family member j gives that person's share s_j of the whole dish, and from it their nutrient intake:

$$N_{person} = N_{meal} \times W_{person} / W_{meal} \quad (7)$$

Here N_{meal} is the total from Eq. (6) and $s_j = W_{person} / W_{meal}$. Table II lists the variables and their typical values in the current prototype.

TABLE II VARIABLE DEFINITIONS USED IN THE METHODOLOGY

Symbol	Description	Typical Value / Unit
W(t)	Raw load-cell reading at time t	grams
N, Δt	Filter window size and HX711 sample interval	8–12 samples, ≈100 ms
σ_{th}	Stability (variance) threshold for state STABLE	≈0.5 g^2
ΔW_k	Mass associated with event k	grams
C_k, $p_{v,k}$	Detected ingredient class and vision confidence	YOLOv8 softmax score
$p_{m,k}$	Weight-domain confidence for event k	0–1
w_v, w_m	Fusion weights for vision and mass confidence	sum to 1 (0.5 each)
$n_{i,Ck}$	Nutrient i per 100 g of ingredient class C_k	ICMR-NIN / USDA lookup
K	Number of accepted ingredient-transfer events	per cooking session
s_j, M	Serving share of member j; number of family members	0–1; integer

IV. SYSTEM IMPLEMENTATION

HARDWARE

The sensing layer has two microcontroller nodes. The first is an ESP32-WROOM-32 board that reads four 3 kg, 4-wire full-bridge load cells through four HX711 24-bit amplifiers, one under each corner of the platform, and reports the combined mass to the backend. The second is an AI-Thinker ESP32-CAM with an OV2640 sensor, mounted overhead and streaming a top-down view of the workspace at 15–20 fps. Both run from a regulated 5 V / 2 A USB supply and talk over Wi-Fi.

Communication and Processing

The load-cell node sends weight samples to a FastAPI backend by HTTP POST (/api/v1/hardware/weight). The camera node serves an HTTP MJPEG stream (/api/v1/hardware/cam-stream) that the backend reads for inference. On the server, the processing engine runs the state logic from Section III-A, applies YOLOv8 to the incoming frames, performs the fusion step in Eq. (4), and calls a nutrition resolver that queries the ICMR-NIN and USDA tables. A conversational dietitian module, running Llama-3-70B on Groq LPU infrastructure, combines the member's stored profile with the day's intake to give guidance.

Data Layer

A SQLite database holds four tables: family_profiles (each member's goals and clinical persona), meal_intake_logs (a date-partitioned record of nutrients consumed), smart_pantry (ingredient stock and shelf life), and raw_logs (weight samples and detection events, kept for auditing).

Application Layer

A browser dashboard shows the live camera feed with YOLO detection overlays, a health summary of calories, protein, fiber and other biomarkers, a seven-day streak calendar, a chat window for the AI dietitian, and a pantry module that suggests recipes from what is in stock.

V. RESULTS AND EVALUATION

The integrated prototype was evaluated through a controlled salad-preparation session. The objective of the evaluation was to demonstrate the complete system workflow rather than to perform a large-scale benchmark of the individual machine-learning components.

During the session, the overhead camera was used to identify ingredients involved in the preparation while the load-cell platform captured corresponding changes in mass. The event-processing layer combined the visual and weight observations and generated ingredient-transfer records.

The resulting records were accumulated in an ingredient ledger representing the composition of the prepared dish. The nutrition-processing component then associated the measured ingredient quantities with corresponding entries in the selected food-composition databases and generated estimated nutrient values.

At the serving stage, the change in the finished-dish weight was used to estimate the portion allocated to an individual family profile. The resulting portion information was combined with the estimated dish-level nutrient composition to generate a per-member nutrition summary. The conversational module subsequently used this information to provide personalised dietary guidance through the application dashboard.

The prototype therefore demonstrated the intended end-to-end workflow:

ingredient identification → mass measurement → event fusion → ingredient ledger → nutrition estimation → serving allocation → personalised guidance.

The evaluation was intentionally limited to a controlled prototype session. Quantitative benchmarking across a larger collection of ingredients, preparation conditions, camera viewpoints, lighting conditions, occlusions, portion sizes, and simultaneous object movements has not yet been performed. Such evaluation is necessary before assessing the robustness and generalisation of the system in ordinary home-kitchen environments.

VI. DISCUSSION AND LIMITATIONS

The current prototype should therefore be considered a proof-of-concept system rather than a clinically validated nutrition-monitoring platform. Several limitations remain. Hands and objects may temporarily block the camera during ingredient transfer, while simultaneous movement of multiple bowls can cause visual and weight events to overlap. Small mass changes may also become difficult to distinguish from sensor noise. Conversely, moving a container without transferring food can produce a weight event without a corresponding ingredient-transfer event.

The nutrition estimates also depend on the selected food-composition database. The current approach uses reference values expressed per unit mass and does not explicitly model changes caused by cooking, such as moisture loss, oil absorption, or changes in edible portion. The resulting values should therefore be interpreted as estimates rather than clinical measurements.

The system has also only been demonstrated in a controlled preparation session. Further evaluation with diverse ingredients, multiple users, different kitchen environments, and repeated preparation sessions is required to establish reliability and generalisation. Privacy considerations associated with continuous kitchen video

must also be addressed before deployment across multiple households.

On readiness, we first identified and framed the problem (TRL 1–2), then prototyped the sensing and software components separately (TRL 3–4). The integrated system has now been shown in a controlled single-session setting (TRL 4–5). Reaching TRL 6–7 will require testing across varied foods, multiple users and ordinary, uncontrolled home cooking. The work also fits three UN Sustainable Development Goals: Goal 3 (easier dietary monitoring), Goal 9 (bringing computer vision, embedded sensing and fusion into kitchen infrastructure) and Goal 12 (greater awareness of food quantities and less avoidable waste).

VII. CONCLUSION

NutriSense shows that ingredient-level nutrition tracking can be largely passive if we watch the cooking instead of the finished plate. Our central insight is simple: a camera tells us what moved, and a scale tells us how much. Fusing the two supports a full chain from ingredient tracking to dish composition, serving measurement and individual nutrition reports, as described in Sections III and IV. The controlled salad-preparation session demonstrates the feasibility of this architecture and interaction concept as a prototype; broader quantitative evaluation remains future work.

We plan three directions for future work. The first is smarter fusion through better object tracking, occlusion handling and confidence-based event validation, together with systematic quantitative benchmarking. The second is deeper personalisation using longitudinal diet history and trend modelling. The third is tighter smart-kitchen integration, with automatic pantry updates, ingredient tracking and recipe suggestions. All three are needed before the system can handle the variety of everyday home-cooked meals.

ACKNOWLEDGMENT

The authors thank Ms. Anitha K, Assistant Professor, Department of Artificial Intelligence and Machine Learning, Sir M. Visvesvaraya Institute of Technology, Bengaluru, for her guidance throughout the design and evaluation of this project.

REFERENCES

1) H. A. Bramantyo, M. A. Faridi, R. Chen, C. Harris, and Y. Sun, “Deep Learning-Based Meat Freshness Detection With Segmentation and OOD-Aware Classification,” IEEE Access, vol. 14, pp. 34939–34952, 2026.

2) N. Karthikeyan, B. Natarajan, S. R. Mahapatro, and D. M. Hussain, “Enhancing dietary management and nutritional analysis through deep learning: A multi-model approach for food classification and volume estimation,” Scientific Reports, vol. 16, no. 1, Art. no. 2875, 2026.

3) L. Abdur Rahman et al., “Generative AI in Precision Nutrition: A Review of Current Developments and Future Directions,” Nutrients, vol. 18, no. 6, Art. no. 938, Mar. 2026.

4) M. Guney-Coskun and M. Basaranoglu, “The Future of Artificial Intelligence-driven Personalized Nutrition in Gastroenterology and Hepatology: Emerging Trends and Perspectives,” Journal of Translational Gastroenterology, vol. 4, no. 1, pp. 79–85, Jan. 2026.

5) K. Ibrar, P. Fusco, G. P. Rimoli, F. Palmieri, and M. Ficco, “Cross-Modal Federated TinyML for MCU-based Internet of Medical Things,” IEEE Journal of Biomedical and Health Informatics, pp. 1–14, 2026.

6) M. Terziyska et al., “A Review of the Transition from Industry 4.0 to Industry 5.0: Unlocking the Potential of TinyML in Industrial IoT Systems,” Sci, vol. 8, no. 1, Art. no. 10, Jan. 2026.

7) M. Xu, “A Closed-Loop Multi-Agent System Driven by LLMs for Meal-Level Personalized Nutrition Management,” in Proc. 2025 5th International Conference on Robotics, Automation, and Artificial Intelligence (RAAI), 2025, pp. 515–520.

8) J. Šulc, V. Reljić, V. Jurošević, L. Krstanović, B. Banjac, and Ž. Santoši, “The Use of Computer Vision Methodologies to Estimate the Volume of Powdered Substance Shapes,” Applied Sciences, vol. 16, no. 4, Art. no. 2053, Feb. 2026.

9) F. Ahmed, B. M. Almutairi, and S. S. Alharbi, “Impact of smartwatch data analytics on dietary consumption awareness among college students in Ar Rass, Al Qassim Region,” Nutrition & Metabolism, vol. 23, Art. no. 24, 2026.

10) A. P. Alcontin et al., “SCALEeat: Vision-Guided Food Scale for Automated Macronutrient Estimation,” Engineering Proceedings, vol. 134, no. 1, Art. no. 83, Apr. 2026.

11) A. Kumar, S. Tarique, P. N, Y. P, T. V. S., and C. R. T., “IoTCaloTrack: An IoT-Enabled Real-Time Calorie Tracking System,” International Journal of Advanced Research in Computer and Communication Engineering, vol. 15, no. 4, Apr. 2026, doi: 10.17148/IJARCCE.2026.154159.

12) D. L. Dharani et al., "Real-Time Food Classification Using VGG19 With Grad-CAM Visualization for Enhanced Interpretability," Journal of Theoretical and Applied Information Technology, vol. 104, no. 3, pp. 282–295, Feb. 2026.

13) S. Besrour, G. Rouhafzay, and J. Jbilou, "A Three-Stage Transformer-Based Approach for Food Mass Estimation," Engineering Proceedings, vol. 118, no. 1, p. 36, Nov. 2025.

14) K. Saleem et al., "Multi-Agent-Based Cognitive Intelligence in Non-Linear Mental Healthcare-Based Situations," IEEE Access, 2025, doi: 10.1109/ACCESS.2025.3544096.

15) C. Yahyati et al., "A Systematic Review of State-of-the-Art TinyML Applications in Healthcare, Education, and Transportation," IEEE Access, vol. 13, pp. 204513–204562, 2025.

16) B. Swathi, S. Veerabomma, M. Archana, D. Bhadru, N. L. Somu, and M. Bhavsingh, "Edge-Centric IoT Health Monitoring: Optimizing Real-Time Responsiveness, Data Privacy, and Energy Efficiency," in Proc. 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), 2025, pp. 354–361.

17) A. Phalle and D. Gokhale, "Navigating Next-Gen Nutrition Care Using Artificial Intelligence-Assisted Dietary Assessment Tools—A Scoping Review of Potential Applications," Frontiers in Nutrition, vol. 12, Art. no. 1518466, 2025.

18) A. Rouhafzay, G. Rouhafzay, and J. Jbilou, "Image-Based Food Monitoring and Dietary Management for Patients Living with Diabetes: A Scoping Review of Calorie Counting Applications," Frontiers in Nutrition, vol. 12, Art. no. 1501946, 2025.

19) S. Liang and Y. Gu, "A Coarse-to-Fine Feature Aggregation Neural Network with a Boundary-Aware Module for Accurate Food Recognition," Foods, vol. 14, no. 3, p. 383, 2025.

20) A. Trigkas, D. Piromalis, and P. Papageorgas, "Edge Intelligence in Urban Landscapes: Reviewing TinyML Applications for Connected and Sustainable Smart Cities," Electronics, vol. 14, no. 14, Art. no. 2890, 2025.

21) M. Sudha, "An Unsupervised TinyML Approach with Efficient Edge AI for Effective Stress and Sleep Monitoring," in Proc. IEEE International Conference on Health Informatics, Dec. 2025, pp. 214–219.

22) K. Ngo, S. Mekhail, V. Chan, et al., "The Use of Artificial Intelligence (AI) to Support Dietetic Practice Across Primary Care: A Scoping Review of the Literature," Nutrients, vol. 17, no. 22, Art. no. 3515, Nov. 2025.

23) P. Indhumathi, J. T, R. K, and S. B. S, "Food Recognition System and Nutrition Tracking Using Convolutional Neural Network Model," in Proc. IEEE APCIT, 2024, pp. 1–6.

24) M. Sguanci et al., "The Internet of Things in the Nutritional Management of Patients with Chronic Neurological Cognitive Impairment: A Scoping Review," Healthcare, vol. 13, no. 1, Art. no. 23, 2025.

25) S. Ponnada et al., "Expanding Applications of TinyML in Versatile Assistive Devices: From Navigation Assistance to Health Monitoring System Using Optimized NASNet-XGBoost Transfer Learning," IEEE Access, vol. 12, pp. 168328–168338, 2024.

26) G. A. Tahir and C. K. Loo, "A Comprehensive Survey of Image-Based Food Recognition and Volume Estimation Methods for Dietary Assessment," Healthcare, vol. 9, no. 12, Art. no. 1676, Dec. 2021.

27) L. M. Amugongo, A. Kriebitz, A. Boch, and C. Lütge, "Mobile Computer Vision-Based Applications for Food Recognition and Volume and Calorific Estimation: A Systematic Review," Healthcare, vol. 11, no. 1, Art. no. 59, 2023.

28) F. P.-W. Lo, Y. Sun, J. Qiu, and B. Lo, "Food Volume Estimation Based on Deep Learning View Synthesis from a Single Depth Map," Nutrients, vol. 10, no. 12, Art. no. 2005, Dec. 2018.

29) P. Sundaravadivel, K. Kesavan, L. Kesavan, S. P. Mohanty, and E. Kougianos, "Smart-Log: A Deep-Learning Based Automated Nutrition Monitoring System in the IoT," IEEE Transactions on Consumer Electronics, vol. 64, no. 3, pp. 390–398, Aug. 2018.

30) B. Javadi, A. Khan, N. Jayasuriya, and Y. Guo, "Smart Nutrition Monitoring System Using Serverless Edge Computing," IEEE Transactions on Consumer Electronics, vol. 70, no. 3, pp. 6363–6375, 2024.