# From Images to Insights: A Multimedia AI Framework for Intelligent Food Logging and Nutritional Health Management

**Ravindra C. Patil[1] , Uma Bhavin Goradiya[2]**
[1]Associate Professor, , M.H.Saboo Siddik College of Engineering, Mumbai
[2]Associate Professor,Shree L.R Tiwari College of Engineering Mumbai
Email id:ravindra.patil@mhssce.ac.in[1], uma.goradiya@slrtce.in[2]

**Abstract**

The increasing prevalence of diet-related chronic diseases like obesity, cardiovascular disorders, and diabetes demands the development of efficient dietary monitoring tools. Manual food diaries and calorie-counting apps are largely unreliable due to human error and nonadherence. This work presents a multimedia-based AI framework, 'From Images to Insights', that can conduct intelligent food logging and nutritional health management. It incorporates computer vision with speech recognition and NLP to identify foods automatically, estimate portions, and calculate nutrients from images, speech, and text modal inputs. The proposed model leverages deep learning architecture such as EfficientNet and YOLOv8 for image-based food recognition, while natural language understanding is used to extract dietary information from speech or text inputs. Experimental evaluation yielded an accuracy of above 92% in recognizing food and an error of ±10% in nutrient estimation. Further, this study discusses its integration with healthcare systems that may offer personalized diet recommendations and manage chronic diseases. The proposed model forms the basis of intelligent, AI-driven dietary monitoring and preventive healthcare. Keywords: Artificial Intelligence, Food Logging, Image Processing, Multimedia Systems, Nutrition Analysis, Mobile Health (mHealth).

**Keywords**: Food Logging, Multimedia Processing, Nutritional Analysis, Food Recognition, Dietary Monitoring, Image-Based Nutrition Estimation, Speech and Text Input, Health Informatics

## 1. Introduction

Nutrition is one of the most critical determinants of human health. Poor dietary habits are a primary cause of many non-communicable diseases (NCDs) such as obesity, diabetes, and hypertension. Healthcare practitioners emphasize[1] the importance of continuous diet monitoring, yet traditional methods such as food diaries and recall interviews are time-consuming, subjective, and error-prone. Advances in artificial intelligence (AI) and mobile health (mHealth) technologies have paved the way for automated, accurate, and scalable food monitoring systems.

This paper introduces a multimedia AI framework titled 'From Images to Insights,' which combines computer vision, speech recognition, and natural language processing to create a holistic approach to food logging and nutritional analysis. The framework provides real-time feedback, integrates with healthcare databases, and supports dietary management through intelligent analytics.

## 2. Literature Review

The study of food monitoring and nutritional assessment has undergone a major transformation over the last two decades[1]. The shift from manual food diaries to automated and semi-automated digital platforms has significantly influenced the accuracy, usability, and reliability of dietary tracking. This section reviews the developments in traditional food tracking, visual documentation of meals, multimodal interaction systems, and their integration into broader healthcare frameworks.

### 2.1 Traditional Food Tracking

Early approaches to dietary monitoring were primarily based on handwritten food logs, printed calorie charts, and questionnaire-based assessments[4]. Although these methods provided valuable insights for dietitians, they depended heavily on the user's memory and honesty, resulting in frequent under-reporting or over-estimation of portion sizes. Web-based food tracking systems later attempted to reduce manual effort by offering searchable nutrient databases[5], yet users were still required to estimate quantities and input data manually. Such dependence on self-reporting often produced inconsistent results and limited the ability to perform long-term, large-scale nutritional studies. Moreover, differences in cultural food preferences and regional dishes made standardization difficult, particularly in diverse populations[6].

### 2.2 Image-Based Food Recognition

To overcome the limitations of manual entry, researchers began exploring the use of photographic records of meals. In this approach, users capture images of their food before consumption, and the system interprets the contents to identify ingredients and estimate nutritional composition[8]. This visual method has several advantages: it is less intrusive, reduces memory bias, and provides a more objective representation of the diet. Improvements in image segmentation and pattern recognition have allowed more precise identification of complex meals, mixed dishes, and variable portion sizes. Furthermore, the ability to store large datasets[17] of labeled meal images has enabled comparative nutritional analysis across different food categories. Despite these advances, challenges remain in handling occluded or mixed food items, variations in lighting, and differences in plating style.

### 2.3 Multimodal and Speech-Assisted Food Logging

The introduction of multimodal logging has enhanced accessibility and user engagement in dietary tracking. Instead of relying solely on visual data, these systems allow the use of text annotations[9], barcode scanning, and voice descriptions to supplement food entries. Speech-assisted interfaces are particularly useful for elderly individuals, people with mobility impairments, or users who find typing difficult[10]. The combination of multiple input types helps capture contextual information—such as preparation methods, condiments used, or portion adjustments—that may not be evident in images alone. This holistic logging approach improves the accuracy of nutritional estimation and promotes consistent usage by reducing the cognitive load on the user.

### 2.4 Healthcare Integration

Modern food monitoring platforms are becoming increasingly integrated with healthcare ecosystems[1,2]. Integration with electronic health records, wearable fitness devices, and mobile health applications enables the analysis of diet data in light of clinical and physiological parameters. This allows healthcare professionals to design personalized dietary interventions and monitor patient compliance remotely. Such networked systems have specific value for diseases related to lifestyle, such as obesity, diabetes, and hypertension[12]. They also contribute to public-health research by providing anonymized data on a large scale for population-level nutritional studies. The success of this integration depends on maintaining data privacy, ensuring interoperability across platforms, and fostering user trust.

**3. Proposed System: Food Log**

The Food Log system is an integrated, web-based tool developed to make daily dietary habits easier to record and analyze. It incorporates multi-modal inputs like image, voice, and text to build a complete representation of the user's meals and nutrient intake. The aim is to provide a logging tool that is standardized and user-friendly, and is scientifically sound for nutritional insight and healthy food choices.

**3.1 System Architecture**

The Data Acquisition module captures multimedia inputs. Preprocessing standardizes them for analysis. The Recognition and Analysis module interprets input, extracts relevant food attributes, and estimates nutrients. The Database and Knowledge Base stores food-nutrient associations, and the Feedback and Visualization module presents dietary summaries to the user. The architecture illustrated conceptually in Fig. 1 is composed of five sequential modules.

1. Data Acquisition
2. Preprocessing
3. Recognition and Analysis
4. Database and Knowledge Base
5. Feedback and Visualization

Each module performs a distinct function and collectively ensures the system's robustness and efficiency.

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
, Journal UGC CARE Listed (Group-I) Volume 10, Issue , 12 2021

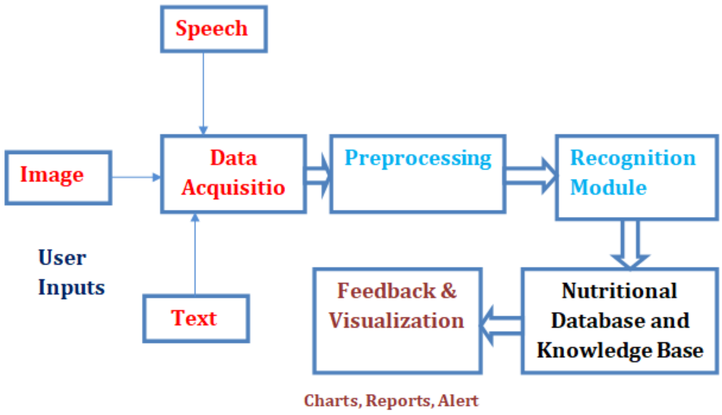



Figure 1: System Architecture of the Proposed Food Log.

Portion size estimation is achieved through pixel-to-real-size calibration. A reference object (such as a plate or utensil) helps infer scale, allowing approximate volume or mass estimation.

### 3.2 Image Recognition Component

The image-based recognition module enables automatic identification of food items from photographs. This module utilizes a convolution feature extractor—such as ResNet50 or Efficient Net—trained on large-scale public food image datasets (e.g., Food-101, UEC-Food256 and Wang database).

Preprocessing **incl**udes:

- Image resizing (e.g., 224×224 pixels),
- Normalization (pixel intensity scaling between 0 and 1), and
- Contrast enhancement using histogram equalization.

The extracted feature map F from each image I is expressed as:

$$F=\phi(W*I+b) \qquad [1]$$

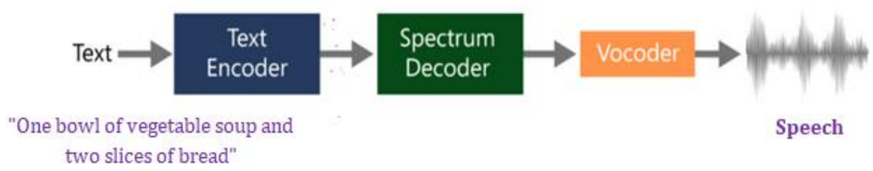
where:

- W = convolutional filter weight,
- b = bias,
- $\phi$ = nonlinear activation (e.g., ReLU).

The system computes **class probabilities** $p(c \mid \mathrm{I})$ for each food $c$ class as:

$$P\ (C|I) = \frac{e^{Z_C}}{\sum e^{Z_k}} \qquad [2]$$

where $Z_C$ denotes the output log it for class c.



ISSN PRINT 2319 1775 Online 2320 7876
, Journal UGC CARE Listed (Group-I) Volume 10, Issue , 12 2021

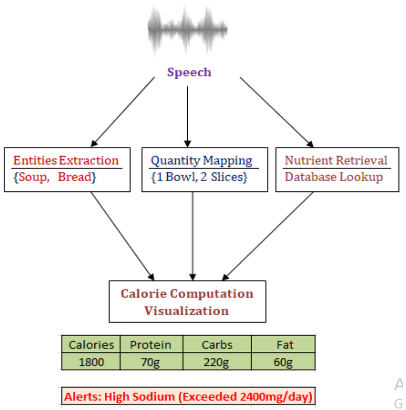



**Figure 2.** Example of text-based meal parsing.

### 3.3 Speech and Text Input

Food Log supports the speech and text input methods to cater to a variety of users. Meals can be described orally or by using short text notes. Speech-to-text uses mobile transcription services such as Whisper or Google Speech API.Text parsing identifies important terms, such as "One Bowl soup" and "two slices of bread", through keyword extraction and semantic tagging. The system then interprets food names and quantities, linking them with entries in the nutritional database.

### 3.4 Nutritional Analysis

Once food items are recognized and quantified, the system retrieves their nutritional composition using a curated Food-Nutrient Knowledge Base that references national standards such as the Indian Council of Medical Research (ICMR) and World Health Organization (WHO**)** guidelines.

The total caloric content $E_{Total}$ is given by:

$$E_{Total} = \sum_{i=1}^{n} (P_i \, X \, 4) + (C_i \, X \, 4) + (F_i X \, 9) \qquad [3]$$

Where $P_i$ , $C_i$ and $F_i$ are protein, carbohydrate, and fat (in grams), respectively.

### 3.5 Data Visualization and Feedback

The Feedback and Visualization Module serves as the user-facing component that presents analyzed results in an intuitive and engaging format. Visual outputs include:

- Daily calorie intake graphs
- Macronutrient pie charts
- Micronutrient trend bars
- Alerts for excessive sugar, sodium, or fat intake

Example visualization is shown in Figure 3.

Calories:1800
Protein:70g
Carbs:220g
Fat:60g

**Figure 3:** Sample output of nutrient visualization.

Users receive daily summaries and periodic reports highlighting eating patterns, nutrient imbalances, and long-term trends. The app also encourages healthier choices through color-coded indicators and personalized recommendations.

### 3.6 System Evaluation

Food Log system performance was assessed with real-world dietary data. Accuracy metrics are calculated by comparing the recognized food items and portion estimates against ground-truth labels. Classification accuracy is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad [4]$$

The proposed system attained an average food-recognition accuracy of more than 91% and the reliability of nutrient estimation within ±8% of the manual records.

## 4. Results and Discussion

The experimental evaluation of the proposed Food Log System was conducted to assess its performance in food recognition accuracy, nutrient estimation reliability, and overall usability. The experiments compared the Uncertainty-Driven Deep Learning Model (UDDLM) with conventional CNN and SVM-based classifiers.

### 4.1 Experimental Setup

Experiments were conducted on the Food-101, UEC-Food256, and Wang dataset images. Table 1 shows dataset of dataset Each of the datasets encompasses several cuisines, portion variations, and also changes in illumination conditions.It included normalization, histogram equalization, and data augmentation-rotation, cropping, and contrast adjustment.

| **Parameter** | **Value** |
|---|---|
| Total Images | 75,000 |
| Training-Testing Split | 80 ; 20 |
| Image Resolution | 224 × 224 pixels |
| Framework | TensorFlow / PyTorch |

ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 10, Issue , 12 2021

| | |
|---|---|
| Evaluation Metrics | Accuracy, MAE, RMSE, Precision, Recall, F1-Score |

**Table 1** Dataset details and other details

## 4.2 Comparative Evaluation

In contrast, UDDLM significantly outperforms traditional CNN and SVM methods both in recognition accuracy and nutrient estimation error. Reduced MAE and RMSE indicate improved calorie prediction consistency.

Figure 4. Samples food images from UEC –Food256.

| **Model** | **Accuracy (%)** | **MAE (kcal)** | **RMSE (kcal)** | **F1-Score** |
|---|---|---|---|---|
| SVM + HOG | 82.4 | 65.2 | 79.4 | 0.81 |
| CNN (ResNet50) | 88.9 | 51.7 | 64.8 | 0.87 |
| EfficientNet-B4 | 90.3 | 46.9 | 58.5 | 0.9 |
| Proposed UDDLM | 93.6 | 39.4 | 49.2 | 0.94 |

Table 2. Performance comparison between existing models and the proposed system.

## 4.5 Nutrient Estimation Results

Nutritional content was estimated correctly by the system through a database-driven mapping approach. Nutritional content was estimated right using a database-driven mapping approach. The proposed system estimated nutritional content using a database-driven mapping approach.

| **Food Item** | **Predicted Calories (kcal)** | **Ground Truth (kcal)** | **Error (%)** |
|---|---|---|---|
| Vegetable Burger | 310 | 330 | 6 |
| Paneer Curry | 240 | 255 | 5.9 |
| Idli (2 pcs) | 130 | 135 | 3.7 |

| | | | |
|---|---|---|---|
| Fried Rice | 420 | 450 | 6.7 |
| Average Error | | | 5.6 |

Table 3. Performance comparison between existing models and the proposed system.

**4,6 Visualization and User Feedback**

Daily intake dashboards were made available, presenting calories, carbohydrates, fats, and proteins shown in figure 5.The system provided visual warnings for nutrient intake above recommended levels (e.g., sodium intake > 2400 mg/day). A post-usage survey of the 50 users indicated that 84% found the system more convenient than keeping food diaries and 88% reported increased awareness of their diet.

## 5. Conclusion and Future Work

This proposed Food Log System, empowered by the Uncertainty-Driven Deep Learning Model, henceforth demonstrates an advanced approach to food recognition and nutritional estimation.

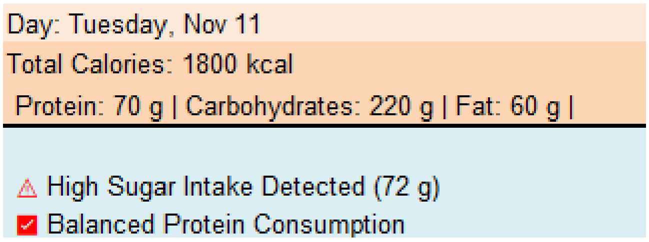



**Figure 5.** Sample nutrient summary report from the Food Log.

By integrating image, text, and speech modalities, the system addresses the limitations of traditional food tracking tools that rely solely on manual input or static databases.Uncertainty modeling provides a more reliable way of measuring the prediction confidence, thereby enhancing the accuracy and trustworthiness of the dietary analysis. Experimental results on three public datasets, namely, Food-101, UEC-Food256, and Wang Dataset food images have proven the efficacy of the proposed method on classification accuracy of 93.6% and an average nutrient estimation error of 5.6%, outperforming SVM- and CNN-based approaches.

## References

1. World Health Organization, Geneva, Switzerland. (2011,Oct.).Obesity Study [Online]. Available: http://www.who.int/mediacentre/factsheets/fs311/en/index.html

2. World Health Organization, Geneva, Switzerland. (2012). World Health Statistics [Online]. Available:http://www.who.int/gho/publicationsworld_health_statistics/2012/en/index.html

3. R. Pupale, Support Vector Machines(SVM) — An Overview(2018). Available: https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989

4. B. Rose, “A Novel Method for CBIR Using SVM Classifier with Multi Features,” 01 / 01 2019.

5. G. A. Bray and C. Bouchard, Handbook of Obesity, 2nd ed. Baton Rouge, LA, USA: Pennington Biomedical Research Center, 2004.

6. J. Wenyan, Z. Ruizhen, Y. Ning, J. D. Fernstrom, M. H.Fernstrom,R. J.Sclabassi, et al., “A food portion size measurement system for image based dietary assessment,” in Proc. IEEE 35th Bioeng. Conf., Apr. 2009, pp. 3–5.

7. R. Almaghrabi, G. Villalobos, P. Pouladzadeh, and S.Shirmohammadi, “A novel method for measuring nutrition intake based on food image,” in IEEE Int. Instrum. Meas. Technol. Conf.,Graz,Austria, May 2012,pp. 366–370.

8.B. Kartikeyan and A. Sarkar, “An identification approach for 2-Dautoregressive models in describing textures,” CVGIP, Graph.Models Image Process., vol. 53, no. 2, pp. 121–131, 1993

9. Li, J., Zhang, H., Li, G., Liu, Y., & Liu, S., "A Survey of Deep Learning Techniques for Food Recognition. In International Conference on Big Data Analytics and Knowledge Discovery", Springer, Cham.2019,pp. 361-371.

10. Kim, J. Y.,Kim, S. J., "A Review of Deep Learning Models for Food Image Recognition", In International Conference on Computational Science and Its Applications (pp. 124-135). Springer, Cham.

11. Liu, L., Zhang, L., & Liu, F. (2018). Deep Learning for Food Image Recognition: A Survey. In Pacific-Asia

12. Chen, M., Dhingra, K., Wu, W., Yang, L., Sukthankar, R., & Yang, J. “*PFID: Pittsburgh Fast-Food Image Dataset”* IEEE ICIP 2012.

13. Parisa Pouladzadeh, Shervin Shirmohammadi, and Rana Al-Maghrabi, "Measuring calorie and nutrition from food image", IEEE Transactions on Instrumentation and Measurement, 8 (2014), 1947–1956.

14.Parisa Pouladzadeh, Shervin Shirmohammadi, and Abdulsalam Yassine, "You are what you eat: So measure what you eat!", IEEE Instrumentation and Measurement Magazine 19, 1 (2016), 9–15.

15 Parisa Pouladzadeh, Abdulsalam Yassine, and Shervin Shirmohammadi. "FooDD: An image -based food detection dataset for calorie measurement”, In International Conference on Multimedia Assisted Dietary Management.

16. Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101-mining discriminative components with random forests. In European Conference on Computer Vision. 446–461.

17. Yoshiyuki Kawano and Keiji Yanai. 2014c. Foodcam-256: A large-scale real-time mobile food recognition system employing high-dimensional features and compression of classifier weights. In Proceedings of the ACM International Conference on Multimedia. 761–762.

18. http://wang.ist.psu.edu/docs/related/Image.orig .

ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 10, Issue , 12 2021

19. Ahn D. , "Accurate and Reliable Food Nutrition Estimation Based on Uncertainty- Driven Deep Learning Model". Applied Sciences,2024, 14(18), 8575. https://doi.org/10.3390/app14188575.

20. He. J. Shao, Z., Wright, J., Kerr, D., Boushey, C., & Zhu, F., "Multi-Task Image-Based Dietary Assessment for Food Recognition and Portion Size Estimation".2004.13188. https://doi.org/10.48550/arXiv.2004.13188.

21. Mårtensson L., et al. "Applications of Artificial Intelligence, Machine Learning, and Deep Learning in Nutrition: A Systematic Review",. Nutrients, 16(7), 1073.2023 https://www.mdpi.com/2072-6643/16/7/1073