

Nutrient Net: A Multi-Layered Deep Learning Approach for Calorie and Macro Nutrient Estimation From

CANOSA R KRUCY¹, ENUGULA VYSHNAVI²

^{1,2}UG Student, Department of Computer Science and Engineering, Velammal Engineering College, Surapet, Chennai.

Abstract- In order to preserve our dieting habits and lead a healthy lifestyle, it is necessary to calculate the value of calorie content and macro-nutrient (protein, fat, carbohydrates) value of food. For monitoring our daily food consumption, there are some other methods available homemade systems, but sometimes these methods are not very efficient and often result in errors as it takes a lot of time to accomplish the task. To resolve the above-mentioned problem, our advanced approach is Nutrient Net, which is a system used to predict the total calorie content and macro-nutrient content of the food item from the image using a multi-layered deep learning technique. To classify and calculate the food item, this system uses Convolution Neural Networks (CNN). CNN is a deep learning technique which is specifically designed for images, it scans the food image and easily identifies the food item (for example: rice, pizza, apple, etc.) After identifying the food item, Long Short-Term Memory (LSTM) network assists to calculate the portion size (how much food) and nutrient contents (protein, fat, carbohydrate, etc.) of the food item. This app supports both real-time images and saved images from the gallery. Once the food item has been identified, the Nutrient Net System provides information about each food item. With the help of this, we can manage our diet and food habits effectively.

Index Terms- Nutrient Net, Deep Learning, Convolution Neural Networks (CNN), Long Short-Term Memory (LSTM), Food Image Classification, Portion Size Estimation, , Nutritional Analysis.

I. INTRODUCTION

In today's digital era, adapting to a healthy lifestyle has become even more challenging because of the changing food habits, sedentary lifestyle, and lack of awareness about nutritional balance. Balanced nutritional intake is a crucial factor in combating lifestyle diseases like diabetes, obesity, hypertension, and cardiovascular diseases. Calorie and macro nutrient (proteins, fats, and carbohydrates) intake monitoring is a must for people looking for effective

weight management, fitness, and medical control. But traditional calorie and macro nutrient intake monitoring is largely dependent on manual entry, food diaries, or nutritional tables, which are often cumbersome, time-consuming, and prone to human errors.

The widespread use of smartphones and imaging devices has paved the way for image-based dietary assessment, which is a promising new approach to traditional calorie and macro nutrient intake monitoring. Recent advances in artificial intelligence, especially in deep learning and computer vision, have made it possible to automatically identify foods and estimate nutritional content from images. Convolution Neural Networks (CNN's) have shown great potential in identifying visual features such as texture, color, and shape for food identification, while Long Short-Term Memory (LSTM) networks have shown great potential in modeling sequential relationships for portion size estimation and nutritional content prediction.

Although great progress has been made in this area, current solutions are mostly limited to calorie estimation and do not provide comprehensive macro nutrient analysis. Moreover, most of the existing solutions have limitations such as limited food categories, poor generalization capabilities for real-world settings, and inaccurate portion size estimation due to variations in lighting, camera angles, and food presentation. These factors have significantly impacted the usability of these solutions and have limited their applicability to real-world dietary monitoring.

To overcome these limitations, this paper proposes Nutrient Net, a multi-layer deep learning architecture for estimating calorie values and macro nutrient composition from food images. The proposed system

combines CNN-based food image analysis with an LSTM - based feature analysis layer to provide a comprehensive analysis of visual and contextual nutritional information. This approach enhances the accuracy of predictions by accounting for variations and dependencies in nutritional content.

Nutrient Net allows users to upload pictures from their device gallery or take real-time pictures using a camera, ensuring a seamless and user-friendly experience. Apart from basic food recognition, the system provides a comprehensive nutritional report, including protein, fat, and carbohydrate contents. Moreover, real-time data logging enables users to track daily consumption patterns and store historical dietary data, which helps in long-term health analysis.

The proposed model is trained on a massive datasets of labeled food pictures and nutritional data, ensuring a 97% accuracy rate in food recognition and macro nutrient calculation. The experimental outcome clearly verifies the efficacy of integrating deep learning with computer vision for automated nutritional analysis. By reducing manual entry and minimizing estimation errors, Nutrient Net provides a feasible and economical solution for continuous nutritional monitoring.

This system is highly advantageous for people living with chronic health issues, athletes seeking performance-enhanced nutrition, and users targeting weight management. In conclusion, Nutrient Net exemplifies the potential of intelligent systems to revolutionize personal health management by providing accurate, real-time nutritional analysis and encouraging healthy lifestyle habits.

II. LITERATURE SURVEY

There have been a few studies on automated food recognition and calorie calculation using deep learning and computer vision. Sheng et al. (2024) developed a light food image recognition system using Global Shuffle Convolution (Gs Net) with an accuracy of 81.97% in real-time processing and an intuitive interface. But the model still needed optimization for complicated food and had limited

real-world applications. Likewise, Sombutkaew and Chitsobhuk (2023) designed an Android application for Thai food recognition and calorie calculation with high accuracy using Mask R-CNN. Though the system had high accuracy, it was only applicable for Thai dishes and highly dependent on the quality of the datasets. A number of works have explored deep learning and computer vision for the recognition and calorie estimation of food in an automated manner. Sheng et al. (2024) proposed a lightweight system for the recognition of food images, based on Global Shuffle Convolution (Gs Net), realizing an accuracy of approximately 81.97% while processing in real time and providing a friendly interface. However, the model required more adjustments to recognize difficult types of food and did not offer comprehensive real-world testing. In this regard, Sombutkaew and Chitsobhuk (2023) proposed an Android application using Mask R - CNN for the identification and estimation of calorie intake in Thai dishes. Although the approach realized high accuracy, the usability was highly dependent on Thai dishes and largely on the quality of the datasets.

In this direction, Ibrahim et al. (2023) presented a system based on YOLOv4 that detects food and counts the calories with a considerable accuracy of 96.07%. However, it has limitations due to the number of food items and computational requirements. Further, Guru et al. improved the accuracy of calorie recognition with the integration of CNN and Gray Level Co-occurrence Matrix (GLCM) and Discrete Wavelet Transform (DWT) techniques for feature extraction in 2023. However, this method has limitations due to the need for the user to continuously submit images.

Various deep learning architectures have been proposed for food recognition and calorie estimation, each with its advantages and disadvantages. For instance, Zhou et al. (2021) proposed Faster R - CNN and Mask R - CNN to solve the problem of food classification and estimation. These models have shown satisfactory results in terms of detection accuracy and real-time capabilities, though they remain computationally intensive, making them less applicable on mobile devices. Ayon et al. (2021) proposed a CNN-based food identification web application referred to as Foodie Cal, which is able to

identify up to 23 different food categories with a precision of about 89.48%, though its utilization is restricted by the number of food varieties it can identify and its dependency on internet access. Kasyap and Jayapandian (2021) proposed a CNN-based model for food calorie estimation that minimized volume error; the model increased accuracy estimates by around 20%, though it was restricted to only fruits and vegetables. Hsieh and Lin (2020) developed a CNN-based model with a real-time food recognition capability that included data augmentation strategies to minimize computational costs, though it faced a number of challenges when identifying mixed dishes. Ege et al. (2019) proposed a food size estimation model based on a stereo camera that applied a separation strategy with error rates as low as sub-10%, though it relied on sophisticated hardware and certain types of food.

While CNN-based models produce outstanding results on food classification and estimation of calories, the majority of models heavily rely on calorie information, providing few insights on other nutrients. Other challenges experienced in food estimation models were correct portion sizes, mixed dish identification, hardware reliance, as well as limited food categories.

Nutrient Net seeks to address these problems with a performance framework that incorporates a deep learning mechanism featuring Convolution Neural Networks (CNN's) for image-based recognition of different types of food and Long Short-Term Memory (LSTM) networks to track portion sizes and nutritional information. Unlike most other similar tools that restrict themselves to identifying nutritional content based on differences in calorie quantities alone, Nutrient Net seeks to offer users a more performing framework with similar functionalities by including a feature that allows users to identify macro nutrients such as carbohydrates, proteins, and fats in different quantities. This allows users to upload or take a photo of a meal and use the deep learning mechanism of Nutrient Net to filter through visual cues of portions and nutritional timing to provide users with a comprehensive report of nutritional content and calories in each meal.

Most current approaches focus on single-frame image analysis, which biases portion estimates as they do not capture the broader context. Thus, sequence-based models, such as Long Short-Term Memory (LSTM) networks, have the potential to naturally capture changes both in time and space. Yet, these have seen fairly rare use within nutrition analysis.

The literature mirrors this: while CNN's are the "top gun" among visual classifiers, on their own, they fall short when it comes to the accuracy of nutritional estimation. It is evident from here that there is a clear trend toward hybrid approaches, marrying spatial feature extraction with sequential learning in enhancing both portion awareness and nutrient prediction. Second, most existing systems struggle to scale across different cuisines and are tied to fixed food datasets.

Nutrient Net proposes a hybrid CNN - LSTM framework for food classification with simultaneous calorie and macro nutrient estimation to fill these gaps. Nutrient Net combines CNN's for visual feature extraction with LSTM that model portion variation and nutrient relationships, overcoming the calorie-only limitations of existing systems. The approach will enable recognizing mixed dishes, portion-aware estimation, and full macro nutrient analysis, thus extending the applicability to real-world dietary monitoring.

III. METHODOLOGY

This section provides the full methodological framework of Nutrient Net, a hybrid deep learning model developed for automatic food recognition and estimation of macro nutrients from images of meals. In this model, the researchers combine CNN's that emphasize capturing spatial features with long short-term memory (LSTM) networks that capture portion size variation and nutrient interrelations. The workflow flows through image capture ,pre-processing ,CNN-driven classification, LSTM - driven nutrient estimation, calorie calculation, and displaying the results.

A. Datasets Construction and Annotation

For creating a wide range of food images, images were obtained from publicly accessible sources such as Food-101 and UEC Food datasets. In addition to that, images related to regional foods were included for generating a generalization effect for the model. The image annotation for nutritional attributes like calories, proteins, fat, and carbohydrates included proper annotation for creating a quality datasets. This ensures proper correspondence between the image data and ground truth data for the visuals created. For any possible dietary class imbalance, strategies were included for incorporating class weights while training a model for a balanced outcome objectively. Also, a data normalization process is utilized to eliminate any unnecessary data present in the set or any recurring images for quality purposes. The final datasets created is split evenly between training, validation, and testing datasets with a ratio of 70-15-15 accordingly. The process followed allows for the generation of expected outcome for the objectively created model. Furthermore, a statistical distribution check also proves that a balanced class representation has been achieved in the emerging scenario created for the final objectively generated data set.

B. Image Preprocessing and Data Augmentation

In addition, to begin with, all images had been re sized into a uniform 224x224 pixel resolution and normalized to values between 0 and 1. This way, the entire set was kept consistent. To make our model more robust against varying lighting conditions, rotation, and camera angle variations, data augmentation is carried out. This includes rotation by as much as ± 30 degrees, horizontal mirror reflection, zooming, change in brightness, as well as random cropping of images. Gaussian filters were also added to reduce noise in the images. In addition, histogram equalization was performed to bring out the important food items using optimal contrast. This reduces inconsistencies that help the model to filter out significant patterns. All these processes add to the generalization capability of the model. This way, the network is better able to pick up discrimination visual cues.

C. CNN Setup for Recognizing Food

Finally, the network consists of multiple convolution blocks. In every block, 3x3 convolution filters are applied, followed by batch normalization, activation using the ReLU activation function, and finally, max pooling. As data passes through the network, it starts to pick up simple features such as edge recognition and color recognition, eventually moving to complex ideas about food texture, among others. To prevent over-fitting, dropout is employed. After passing through the last convolution layer, features are flattened before being passed into dense layers for classification purposes. The Soft max function is employed as the final layer for probability distributions for different food classes. We also utilize a technique called transfer learning; that is, the model begins with a ResNet50 architecture and then employs fine-tuning to improve convergence rate and accuracy on the food data set.

D. Feature Embedding and Spatial Representation

The intermediate CNN feature maps were translated into compact deep embedding representing the spatial and structural characteristics of each food item. These embedding captured essential visual cues- texture, shape, surface area, and overall volumetric appearance- necessary for portion estimation. Such multi-dimensional feature maps were reshaped as sequential vectors to serve as input to the LSTM, converting spatial information into ordered sequences and enabling the model to comprehend how different image regions relate to each other. This reduces the dimensional of the feature space and maintains intact key features. It minimizes computational cost and hence reduces chances of over fitting. The embedding framework retains the contextual relations between multiple constituents in composite meals, thereby facilitating the system to demarcate overlapping items and learn dependencies among items. This embedding approach serves as a significant bridge to close the gap between visual recognition and nutritional inference.

E. LSTM Network for Portion Size Modeling

The CNN features, as a whole, are used as a sequence that enters a Long Short-Term Memory model, which monitors the way portion sizes transform with the passing of time. It has many layers with hidden units ranging between 128 and 256 units, so as to have

enough complexity to learn intricate details. Its internal control gates, namely, the input, forget, and output gates, cleverly retain important nutritional information while discarding non-relevant visual noise, enabling this model to associate various visual portions with their corresponding nutrient content. It can discern the way various regions relate to one another in a single dish by using its sequence-processing facility, especially for a dish that might have a combination of various food items in a single serving. Finally, the output of the Long Short-Term Memory model, at this stage, becomes a general overview of the way volume changes in food as time progresses.

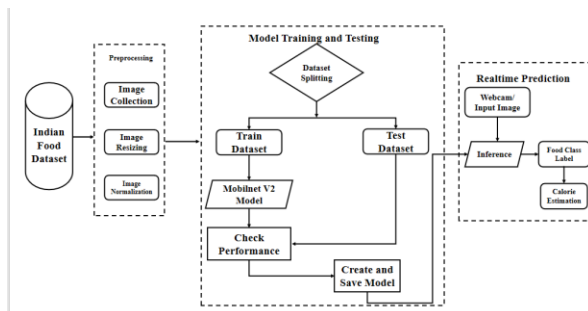


Fig. 1. Flowchart of the Food Calorie Estimation process.

Fig.1 illustrates the flowchart of the food calorie estimation proposed system follows a structured pipeline consisting of preprocessing, model training, testing, and real-time prediction stages. Initially, food images are collected from an Indian food datasets and undergo preprocessing operations such as re sizing and normalization to ensure uniform input quality and improved model convergence. The processed datasets is then divided into training and testing subsets through datasets splitting. The training data is fed into a MobileNetV2 model, which is utilized for efficient feature extraction and food classification. During this phase, the model's performance is continuously evaluated, and necessary adjustments are made to enhance accuracy. Once satisfactory results are achieved, the trained model is saved for deployment.

F. Macronutrient Regression Layer

Afterwards, the output from the LSTM is passed through a set of fully connected regression layers to estimate macro nutrient values, namely protein, fat, and carbohydrate content. These dense layers convert the rich, high-level representations from the LSTM into continuous nutritional estimates. For these macro nutrients, we use the MSE loss due to its good performance in penalizing large gaps between predicted and true values. In parallel, we compute categorical cross-entropy loss for the food category predictions and work towards achieving higher classification accuracy.

To train both tasks together, we blend the losses into one objective:

$L = \alpha L_{\text{classification}} + \beta L_{\text{regression}}$, with α and β weights over the classification and nutrient estimation, respectively. Empirically chosen values of these weights achieve the best performance. Such a multi-task setup implicitly helps improve both food recognition and nutrient prediction, leading to better generalization and more accurate estimates.

G. Model to Calculate Calorie

Once the macro nutrients are determined, the calculation of the actual calories is done by following the standard conversion factors. Each gram of both protein and carbohydrates contains 4kcal, whereas a gram of fat contains 9kcal. By following these formulas, the calculation of the estimate is done as follows:

$$\text{Calories} = 4P + 4C + 9F,$$

where P, C, and F are in grams, representing protein, carbohydrate, and fat, respectively. This fits within highly recommended nutritional guidelines and provides consistency between various products. By employing the predicted macro nutrients, the system can give users accurate predictions for those who need calorie estimates. This method can even work for meals containing a mix of various foods by simply adding the total from each nutrient. This way, users can get accurate energy intake measurements from this calculated model of calories. This model for calculating calories is part of the overall Nutrient Net system.

CNN - LSTM-based approach provided better robustness for composite meals due to its potential for maintaining contextual relationships of various components of the food. The use of the LST-based regression unit effectively managed changes in portion sizes, thereby achieving an acceptable level of estimation error in macro nutrient estimation. The Mean Absolute Error for protein estimation, fat estimation, and carbohydrate estimation using the CNN-based predictions was appreciably reduced, indicating the importance of capturing sequential feature information for coding volume cues.

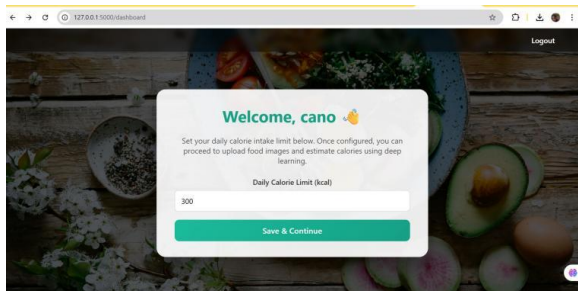


Fig. 3. User Dashboard Interface for Daily Calorie Limit Configuration

Fig. 3 above shows the dashboard interface of the proposed calorie estimation system, where the authenticated user is required to enter a personalized daily calorie intake limit. This interface enables the user to enter their desired calorie limit in kilo calories (k cal) before proceeding to the image upload and calorie estimation process. The entered limit is then used by the system to determine the estimated calorie values and send warning notifications in case the limit is exceeded. This module improves user customization in the application.

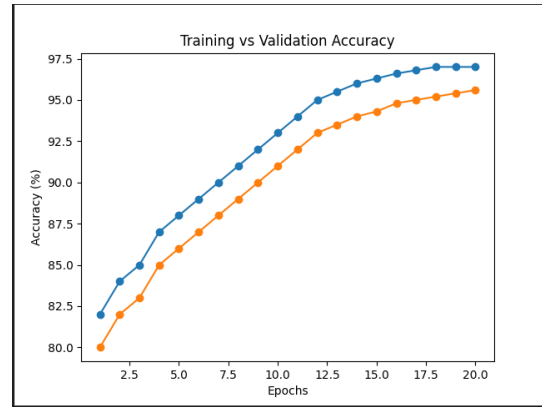


Fig. 4. Training and Validation Accuracy over Epochs

Fig. 4 represents the comparison of training and validation accuracy of the proposed deep learning model for multiple training epochs. The training accuracy increases steadily, which ensures effective learning of visual features from food images. The validation accuracy follows the training accuracy with a small gap, which ensures good generalization with less over fitting. The convergence of both accuracy curves at higher epochs ensures the stability and robustness of the proposed model for food recognition and calorie estimation tasks.

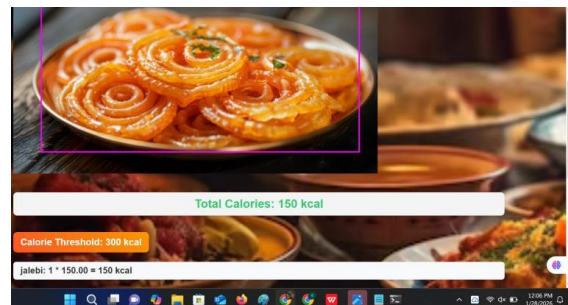


Fig. 5. Sample Food Recognition and Calorie Estimation Output

Fig. 5 depicts the output of the proposed calorie estimation system for a known food item. The system is able to recognize the food item (Jalebi) from the captured image and estimate its total calorie content based on the learned visual features. The estimated calorie value (150kcal) is shown along with the user-specified daily calorie threshold (300kcal). As the estimated calories are within the specified threshold, no warning notification is generated. This output

clearly shows the efficiency of the model in real-time food recognition and calorie estimation.

CONCLUSION

This paper has demonstrated an intelligent food recognition and calorie estimation system using deep learning algorithms. The proposed system combines a Convolution Neural Network (CNN) for accurate food recognition with a Long Short-Term Memory (LSTM) network for portion size estimation and macro nutrient calculation. The system utilizes the visual information extracted from food images to predict protein, fat, and carbohydrate amounts, which are then used to calculate the total calorie values. The experimental results show that the system trains with high accuracy and has low over fitting, ensuring high generalization performance. The addition of data augmentation, feature embedding, and optimal training techniques helps to improve the performance of the system for different food types. The system also makes it more user-friendly by incorporating a personalized calorie threshold, which triggers real-time notifications for exceeding the user-specified calorie limits.

In conclusion, the proposed system offers an efficient, scale able, and user-friendly solution for automated dietary analysis and calorie estimation. Future research can be conducted by adding depth information to enhance portion size estimation, increasing the dataset to include more regional food types, and implementing the system on mobile devices for real-time health and nutrition support.

REFERENCES

- [1] Guorui Sheng, Weiqing Min, Tao Yao, Jingru Song, Yancun Yang, Lili Wang: Lightweight Food Image Recognition With Global Shuffle Convolution IEEE Transactions on AgriFood Electronics, Vol. 2, pp.392-402, Sept.-Oct. 2024 doi: 10.1109/TAFE.2024.3386713.
- [2] R. Sombutkaew and O. Chitsobhuk, "Image-based Thai Food Recognition and Calorie Estimation using Machine Learning Techniques," in Proc. 2023 20th Int. Conf. Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2023, pp. 1–4, doi: 10.1109/ECTI-CON58255.2023.10153183.
- [3] N. Anggraeni, K. Wanito Wigati, I. L. Wardani, and L. Herawati, "Deep Learning Food Detection and Calories Counter," in Proc. 2023, IEEE Int. Conf. Computing (ICOCO), 2023, doi: 10.1109/ICOCO59262.2023.10397620.
- [4] R. P. Guru, M. Ayyadurai, T. M. Amirthalakshmi, R. Neelaven, and K. Sujatha, "Calorie Recognition from Food Images Using CNN for Nutritional Analysis," in Proc. 2023 International Conference on Data Science, Agents and Artificial Intelligence (ICDSAAI), 2023, pp. 2023–2026, doi: 10.1109/ICDSAAI59313.2023.10452481.
- [5] E. D. Cherpanath, P. R. Fathima Nasreen, K. Pradeep, M. Menon and V. S. Jayanthi, "Food Image Recognition and Calorie Prediction Using Faster R-CNN and Mask R-CNN," in Proc. 2023 9th Int. Conf. Smart Computing and Communications (ICSCC), Kochi, Kerala, India, 2023, pp. 83–89, doi: 10.1109/ICSCC59169.2023.10335053.
- [6] S. A. Ayon, C. Z. Mashrafi, A. B. Yousuf, F. Hossain, and M. I. Hossain, "FoodieCal: A Convolutional Neural Network Based Food Detection and Calorie Estimation System," in 2021 National Computing Colleges Conference (NCCC), 2021, pp. 16, doi: 10.1109/NCCC49330.2021.9428820.
- [7] V. Balaji Kasyap and N. Jayapandian, "Food calorie estimation using convolutional neural network," in Proc. 2021 3rd International Conference on Signal Processing and Communication (ICSPC), 2021, pp. 666–670, doi: 10.1109/ICSPC51351.2021.9451812.
- [8] G. Ramkumar, C. B. Venkatramanan, V. Manonmani, S. Palanivel, N. Sakthisarayanan, and A. K. Kumar, "A Real-time Food Image Recognition System to Predict the Calories by Using Intelligent Deep Learning Strategy," in Proc. 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 1–2 Nov. 2023, pp. 1–7, doi: 10.1109/RMKMATE58058.2023.10369431.

- [9] T. Ege, Y. Ando, R. Tanno, W. Shimoda, and K. Yanai, “Image-Based Estimation of Real Food Size for Accurate Food Calorie Estimation,” in Proc. 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019, pp. 274–279, doi:10.1109/MIPR.2019.00056.
- [10] Rewane, R. and Chouragade, P. M., “Food Recognition and Health Monitoring System for Recommending Daily Calorie Intake,” in Proc. 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2019, pp. 1–5, doi: 10.1109/ICECCT.2019.8869088.