

## CUSTOMIZED DIET ASSISTED SYSTEM BASED ON FOOD RECOGNITION USING DEEP LEARNING

Amsaleka. R<sup>1</sup>, Pavithra.V<sup>2</sup>, Poorani. S<sup>3</sup>, Sharmila. G<sup>4</sup>, Yogeshwari. R<sup>5</sup>

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering, Vivekanandha College of Technology for Women, Tiruchengode, Tamilnadu

<sup>2</sup>UG Scholar, Department of Computer Science and Engineering, Vivekanandha College of Technology for Women, Tiruchengode, Tamilnadu

DOI: <https://www.doi.org/10.58257/IJPREMS31236>

### ABSTRACT

Today, people across the universe are becoming more sensitive to their diet. Unbalanced diets can cause many problems, like weight gain, obesity, sugar, etc. So different systems were developed analyze food images to calculate calories, nutrition levels, etc. Food is one of the most important requirements for every living being on earth. Human beings require their food to be fresh, pure, and of standard quality. The standards imposed and automation carried out in the food processing industry take care of food quality. Food is very important for a successful healthy diet, and measuring calories and nutrition in daily food is one of the most challenging methods. Smartphones play a vital role in today's technological world using this technique, which enhances the issue of dietary intake. In this methodology, a food image recognition system for measuring calorie and nutrition values has been developed. The user has to take a picture of the food image; this system will classify the image to detect the type of food and portion size, and the recognition information will estimate the number of calories in the food. In the proposed food area, size and volume are used to calculate the calories and nutrition in an accurate way.

**KEYWORDS:** Food recognition, dietary consumption, machine learning, deep learning, nutrition values.

### 1. INTRODUCTION

Good eating is essential to human health. Natural products have been commonly used as foodstuffs and can also be manufactured to fulfil market demand. Food attributes such as type, structure, nutrients, and process types (natural products and refined foods) are concerned with balanced diet issues. It is a fact that individuals have different eating patterns in other areas. Knowing the characteristics of foods (type, composition, nutrients, process types, etc.) and their consistency and protection for customers worldwide is essential for the inspection. A realistic demand in everyday life is the swift, precise, and automatic determination of food attributes. Modern techniques have been commonly used to detect food characteristics, including electronic noses, computer vision, spectroscopy, spectral imaging, and so on. A large amount of digital data relating to food properties can be collected through such methods. The data analysis of these methods is important because the enormous data volume includes a lot of repetitive and irrelevant material. It is an urgent and vital challenge to deal with such a vast volume of data and extract useful features from the acquired data, and the complexity of bringing these methods into real-world use. Commonly, dietary intake measurement methods can be classified into traditional and electronic approaches. The use of traditional methods has been well known for decades, whether in hospitals or through research studies. Electronic methods have started to appear recently due to the widespread use of technology globally. A food nutrition and energy intake recognition system for medical purposes is proposed. This system is built on food image processing and shape recognition in addition to nutritional fact tables. This is a measurement method that estimates the number of calories from a food's image by extracting the volume of the food inside the image by using the thumb as a reference. The application is designed to aid dieticians in the treatment of obese or overweight people, although normal people can also control their daily eating more closely without worrying about overeating and weight gain. In this measurement system, the goal is to develop and implement an instrument that measures daily food intake using mobile devices with a built-in camera to capture a photo of the food intake before and after eating in order to estimate the number of consumed calories. Figure 1 shows the food image datasets.



Fig 1: Food image datasets

## 2. RELATED WORKS

Goh, Alex M., et al. [1] proposed a method of food image classification based on inception-v3 transfer learning in food images, and the method was compared with other methods. In particular, the method of food image classification based on migration learning can achieve higher accuracy. Moreover, the neural network model based on transfer learning performs better in food image classification on the Food 101 database than the model based on the original DCNN. Fine-tuning Inception-v3 transfer learning can effectively improve the accuracy of food image classification, so the model can provide an effective and rigorous computer-assisted diagnostic when food image data is insufficient. If the network selection for transfer learning is inappropriate, the problem of negative transfer may occur, which will lead to a decrease in accuracy and an increase in training time. Therefore, how to select the appropriate network for food image tasks is the next research direction. For the problem of the large gap between the specificity and sensitivity of deep learning models in this study, we found that ensemble learning has a good performance in medical images from many pieces of literature; their sensitivity and specificity are both high. Saha, Dhritiman, et al. [2] implemented the system for extracting useful information from the high-dimensional hyperspectral data containing redundant information, which is a challenging task. Hence, to make online hyperspectral imaging inspection a reality, emerging and efficient algorithms are needed. In this context, machine Learning algorithms can play an effective role in the analysis of hyperspectral images with high accuracy. Besides, advanced machine learning algorithms like deep learning have found their potential application in hyperspectral image analysis of agricultural products. Since deep learning involves automatic feature learning during the training stage, it has more potential for real-time applications than other traditional machine learning algorithms. The scope of lifelong machine learning should be explored further, and its application should be extended to other agricultural crops for quality monitoring. The selection of effective wavelengths from the hyperspectral data is of paramount importance since it greatly reduces the computational load and time, which enhances the scope for real-time applications. Due to the feature-learning nature of deep learning, it is one of the most promising and powerful techniques for real-time applications. However, the field of deep learning is relatively new and needs further research for its full utilisation. Lo, Frank Po Wen, et al. [3] investigated the underlying algorithms and mathematical models used in the field of dietary assessment, especially in food recognition and volume estimation. After a comprehensive review of several state-of-the-art food recognition systems, recent research has found that they are focused on exploring the potential of assessing dietary intake based on a deep learning approach. Furthermore, the state-of-the-art approaches to food volume estimation are summarised and discussed in this study. An extensive comparison has also been presented to highlight the main advantages and challenges of different approaches. Overall, there is currently a growing potential for integrating different approaches to improve the overall accuracy of food volume estimation. And provided an overview of computing algorithms, mathematical models, and methodologies used in the field of image-based dietary assessment. It also provides a comprehensive comparison of the state-of-the-art approaches in food recognition and volume/weight estimation in terms of their processing speed, model accuracy, efficiency, and constraints. It will be followed by a discussion on the deep learning method and its efficacy in dietary assessment. After a comprehensive exploration, we found that integrated dietary assessment systems combining with different approaches could be the potential solution to tackling the challenges in accurate dietary intake assessment. Tahir, Ghalib Ahmed, et al. [4] implemented deep learning networks to extract features, Relief F for feature ranking and selection, and ARCIKELM for classification. For feature extraction, this paper takes into account the excellent generalization ability of deep model features. It has evaluated three state-of-the-art deep networks and found that

Inception-Resnet-V2 has superior performance as compared to others. However, features extracted from deep learning models have a very high dimension and increase classification time. The framework has used the Relief F method to determine the optimal length and found that the Relief F method has reduced the accumulative learning time of the proposed classifier for all the datasets by 52.14%. For addressing the challenges of data incremental learning and class incremental learning, the framework used the novel adaptive reduced class incremental kernel extreme learning machine. It dynamically increases hidden neurons and output neurons. The decreased plasticity of previous neurons reduces catastrophic forgetting. Menichetti, Giulia, et al. [5] introduce a machine learning algorithm that accurately predicts the degree of processing for any food, indicating that over 73% of the U.S. food supply is ultra-processed. We show that the increased reliance of an individual's diet on ultra-processed food correlates with a higher risk of metabolic syndrome, diabetes, angina, elevated blood pressure, and biological age and reduces the bioavailability of vitamins. Finally, we find that replacing foods with less processed alternatives can significantly reduce the health implications of ultra-processed foods, suggesting that access to information on the degree of processing, currently unavailable to consumers, could improve population health. And introduce FoodProX, a machine learning classifier trained to predict the degree of processing of any food. Importantly, FoodProX allows us to define a continuous index that captures the degree of processing of any food and can help quantify the overall diet quality of individuals, unveiling statistical correlations between the degree of processing characterizing individual diets and multiple disease phenotypes. He, Jiangpeng, et al. [6] proposed a multi-task framework for food classification and food portion size estimation by using L2-norm-based soft parameter sharing. We also investigated cross-domain feature adaptation together with different normalization techniques to further reduce portion estimation error. Our method is evaluated on a real-life eating occasion food image dataset with ground-truth category and portion size provided by registered dietitians. Our best result achieved 88.67% classification accuracy, with mean absolute errors of 56.82 Kcal for all food and 50.86 Kcal for correctly classified food for portion size estimation, surpassing the baseline results of 86.08% and 62.27 Kcal, respectively. And introduce a food image dataset collected from a nutrition study where the ground-truth food portion is provided by registered dietitians. The multi-task learning uses L2-norm-based soft parameter sharing to train the classification and regression tasks simultaneously. And also propose the use of cross-domain feature adaptation together with normalization to further improve the performance of food portion size estimation. The results outperform the baseline methods for both classification accuracy and mean absolute error for portion estimation, which shows great potential for advancing the field of image-based dietary assessment. Jiang, Landu, et al. [7] explored the food recognition and dietary assessment problem by leveraging deep learning techniques. In particular, to have a better understanding of object detection and nutrition analysis, we apply a state-of-the-art Faster R-CNN model to generate ROIs and use a deep neural network to extract the feature map for food item recognition. And analyze the nutrition of the detected food and summarize the report of the meal based on modern technology-based dietary assessment tools. We conduct extensive experiments to evaluate the efficiency and effectiveness of our system. Results show that our proposed solution achieved comparable performance and has great potential to promote healthy diets and feasible advice. And also generate a new type of dataset about food items based on FOOD101 with bounding. The model is evaluated using different evaluation metrics. The experimental results show that our system is able to recognize the food items accurately and generate the dietary assessment report efficiently, which will benefit the users with a clear insight into healthy diets and guide their daily recipes to improve body health and wellness. Oliveira Chaves et al. [8] summarize the latest information on the use of different ML algorithms to evaluate food intake. It can serve as a guide for health professionals who want to work in the area of AI. It is concluded from the results found that, currently, there is a great and growing interest in the use of ML algorithms in the area of nutrition, mainly due to a significant increase in publications in recent years. In addition, it is also noted that the supervised learning algorithms, more precisely those based on decision trees, were the most used. The more frequent use of DT is possibly because they are fast to apply, simple to understand, and whose results are easy to interpret and explain. In addition, there was a change in the use of computational tools for statistical analysis, with a tendency to use other software, such as R, because it is more complete, instead of the classic statistical programmers. Regarding the assessment of food intake, we observed that the FFQ was the most commonly used method since it allows a long-term evaluation and is simple, fast, and easy to administer and process. However, even understanding the importance of investigating food intake in each population and how the use of ML algorithms can be interesting, there was little diversity of countries involved in the studies analyzed. Manoharan, Samuel, et al. [9] implemented the recommendation system for dietary food based on the disorders, health conditions, and other features of the patients. The system is developed using the K-clique and the gated recurrent network. Based on the dataset collected over the internet. The features of the products and the patients are sorted out, pre-processed, encoded, and segregated based on their similarities and used for training the model. The model is trained quickly using the gated recurrent network. The developed design is trained, tested, and cross-validated. The outcomes of the training, testing, and cross-validating

demonstrate that the proposed system has better precision and accuracy than the other ML and DL procedures such as the MLP, RNN, and Logistic Regression, Navies' Bayes, respectively. The better precision and accuracy observed for the developed system experimentally are compared with machine learning techniques such as logistic regression and Nave Bayes and other deep learning classifiers such as the MLP and RNN to demonstrate the proficiency of the K-Clique deep learning classifier-based recommendation system (K-DLRS). Shen, Zhidong, et al. [10] present a novel system that tells us information about the type of food we eat and its attributes. This system takes the image of the food from the user, and after correct classification, the system will tell us about the attributes of the food. A dataset that consists of a common meal of Food-101 and our subcontinental food has been used in our system. And we have fine-tuned the Inception V-3 and V-4 models to recognize the food items and also proposed a method to measure the attributes of the food using the attribute estimation model. The results are enhanced via data augmentation, multi-cropping, and similar techniques. The proposed method for classification as well as for the extraction of attributes achieves a considerably high accuracy of 85%. In order to better nurture the basic awareness of calorie calculations among users, it can be extended by introducing new calories to measure nutritional characteristics and combining it with deep learning techniques. Data sets and features have a great impact on detection results. Existing data sets are not sufficient and contain limited parameters such as different lighting conditions, camera angles, different backgrounds, etc. In future research, better review techniques should be used to review various types of data sets. In addition, the system and application are optimized architecturally, and a database for storing calculated values, food labels, and other parameters is combined with a faster lookup technique to process the image.

### 3. BACKGROUND OF THE WORK

Automatic food identification and calorie estimation have become important issues in the last few years because of the negative impact of obesity on our health. Obesity may cause cardiovascular diseases, diabetes mellitus type 2, obstructive sleep apnea, cancer, osteoarthritis, asthma, etc. Researchers said that junk foods and processed foods are responsible for increasing childhood obesity. Eating extra calories can harm the healthy production and functioning of the synapses in our brain. Fried chicken, pizza, burgers, etc. are favorite fast foods for both children and adults. People often buy these high-calorie foods to control their appetite, especially when they are busy and unable to take their meal in time. Today's people are more conscious of their health issues and try to maintain a healthy diet. Due to the availability of smart phones, computer-aided object recognition techniques have become more popular for dietary assessment. Although the identification of food and the estimation of its calorie content are very challenging tasks, many effective steps have already been taken in this regard. And proposed an easy but more effective calorie measurement technique that helps people identify the amount of junk food and snacks they can intake, as well as to decide whether the food is harmful or not good for their health. A semi-automatic system was proposed that measures the caloric content of food intake. System utilizes nutrition tables for better results. A manual approach is needed to prepare the diet chart for every food. Experts are needed to prepare the diet table for particular patients (diabetic patients). In the existing system, images are taken by the user with a mobile device, followed by a preprocessing step. Then, at the segmentation step, each image will be analyzed to extract various segments of the food portion. It is known that without a good image segmentation mechanism, it is not possible to process the image appropriately. That is, jointly used color and texture segmentation tools. Food identification is a challenging challenge since food products are presented. Sometimes, they are different within the same group. A sort of issue with categorizing fine-grained pictures is the identification of food pairwise local characteristics that take advantage of eight specific food ingredients' positional relationships. The proposed multi-food image recognition system that detects first foods recognizes color, texture, gradient, and SIFT extracted by several detectors using multiple kernel learning regions (MKL). The web application estimates nutritional balance as a food recognition device with food recording. The food is divided into 300 blocks, and five classes are further classified, such as staple, main dish, side dish, fruit, and non-food, based on each block's extract colour and DCT coefficients Food identification and quantity estimation are part of the TADA dietary evaluation system [35]. Still, there are certain limitations on placing food on white dishes and taking food pictures for food quantity calculation with a checkerboard. Image recognition processes are performed on servers in all the systems mentioned above, preventing operations from occurring interactively because of delays in contact. On the other side, for example, a device can identify food products in real-time on the customer side, which does not require contact with external computer resources and allows customers to use it interactively.

### 4. PROPOSED WORK

Commonly, dietary intake measurement methods can be classified into traditional and electronic approaches. The use of traditional methods has been well known for decades, whether in hospitals or through research studies. Electronic methods have started to appear recently due to the widespread use of technology globally. A food nutrition and energy intake recognition system for medical purposes is proposed. This system is built on food image processing and shape

recognition in addition to nutritional fact tables. This is a measurement method that estimates the number of calories from a food's image by extracting the volume of the food inside the image by using the thumb as a reference. The application is designed to aid dieticians in the treatment of obese or overweight people, although normal people can also control their daily eating more closely without worrying about overeating and weight gain. The initial effort has focused on identifying food items in an image by using shape recognition, image processing, and classification by using CNN, and as a result, we reach a reasonable measurement error for this method. In this measurement system, the goal is to develop and implement an instrument that measures daily food intake using mobile devices with a built-in camera to capture a photo of the food intake before and after eating in order to estimate the number of consumed calories. Propose the system to recommend nutrition based on the user's BMI calculation, and need to take the image of the food and extract the features from the image. Fig. 2 shows the proposed framework for a food recognition system.

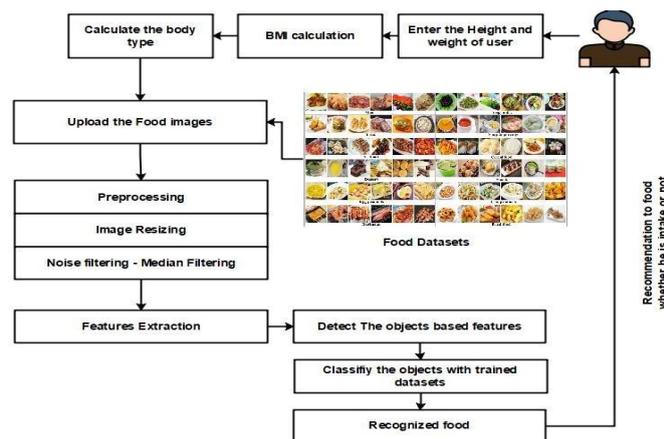


Fig 2: Proposed Work

## 5. CONVOLUTIONAL NEURAL NETWORK

In a CNN, the input is a tensor with shape (number of images) x (image height) x (image width) x (input channels). After passing through a convolutional layer, the image becomes abstracted to a feature map, with shape (number of images) x (feature map height) x (feature map width) x (feature map channels). A convolutional layer within a neural network should have the following attributes:

- Convolutional kernels are defined by a width and height (hyperparameters).
- The number of input channels and output channels (hyper-parameter).
- The depth of the convolution filter (the input channels) must equal the number of channels (depth) of the input feature.

Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus. Each convolutional neuron processes data only in its receptive field. Although fully connected feed-forward neural networks can be used to learn features and classify data, this architecture is impractical for images. It would require a very high number of neurons, even in a shallow architecture, due to the very large input sizes associated with images, where each pixel is a relevant variable. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10,000 weights for each neuron in the second layer. Instead, convolution reduces the number of free parameters, allowing the network to be deeper. For example, regardless of image size, tiling 5 x 5 regions, each with the same shared weights, requires only 25 learnable parameters. Using regularized weights over fewer parameters avoids the vanishing gradient and exploding gradient problems seen during back propagation in traditional neural networks.

**Pooling layers-** Convolutional networks may also include nearby or worldwide pooling layers to streamline the underlying computation. Pooling layers lessens the size of the statistics by combining the outputs of neuron clusters at one layer right into an unmarried neuron in the subsequent layer. Local pooling combines small clusters, commonly 2 by 2. Global pooling acts on all the neurons in the convolutional layer. There are two common varieties of pooling: max and common. Max pooling uses the highest cost of each cluster of neurons at the previous layer, while common pooling alternatively makes use of the average cost.

**Fully connected layers-** Fully linked layers connect every neuron in one layer to each neuron in some other layer. It is similar to a conventional multi-layer perceptron neural community (MLP). The flattened matrix is going through a totally linked layer to categorise the pictures.

**Receptive field-** In neural networks, every neuron receives input from a small number of locations within the

preceding layer. In a fully related layer, every neuron gets input from each neuron in the previous layer. In a convolutional layer, each neuron gets input from only a confined vicinity of the previous layer, referred to as the neuron's receptive area. Typically, the vicinity is a square (e.g., five by way of five neurons). (So, in a completely linked layer, the receptive subject is the complete previous layer.) Thus, in every convolutional layer, every neuron takes input from a larger number of pixels in the input image than in previous layers. This is due to making use of the convolution time and again, which takes into consideration the size of a pixel and its surrounding pixels.

**Weights** -Each neuron in a neural network computes an output cost by making use of a selected feature to enter values coming from the receptive field within the previous layer. The function that is carried out on the input values is decided by means of a vector of weights and a bias (generally real numbers). Learning includes iteratively adjusting these biases and weights. The vector of weights and the bias are referred to as filters and constitute unique features of the input (e.g., a selected shape). A distinguishing characteristic of CNNs is that many neurons can have the same clear-out. This reduces the memory footprint because a single bias and a single vector of weights are used across all receptive fields sharing that filter, instead of every receptive field having its own bias and vector weighting. First, ITER max, ERR min, BATCH Straining, and, size batch are initialised which are nothing but learning rate, number of maximum iteration, minimum error, training batches, and, batch size respectively;

```
According to n11 and n5,5; compute 2, n3, and n44, k1, k2 Randomly generate the weights of the CNN; cnnModel = InitCNNModel(, n1-5-5); iter = 0; err = +inf; while err > ERRmin and iter < ITERmax do err = 0; for batch = 1 to BATCHStraining do [J(), J(J(theta))] cnnModel_train (TrainingDatset, TrainingLabelsets), and theta should be updated; err = err + mean(J(theta)); end for err = err/BATCHStraining; iter++;end while ave parameters theta of the CNN
```

The pooling process is similar to the convolution process in that it involves a sliding window similar to a filter, but the calculation is simpler. Mean pooling uses the average value in an image area as the pooled value of the area. This approach preserves the background of the image well. Max pooling takes the maximum value of the image area as the pooled value of the area and preserves the texture of the image well. The function of the fully connected layer is to integrate the multiple image maps obtained after the image is passed through multiple convolution layers and pooling layers to obtain the high-layer semantic features of the image for subsequent image classification.

## 6. PERFORMANCE METRICS

The performance of the system can be analysed using KAGGLE food datasets. Different performance measures such as accuracy, sensitivity, specificity, and error rate can be derived for analysing the performance of the system.

- True positive (TP): number of true positives (perfect positive prediction)
- False positive (FP): number of false positives; imperfect positive prediction
- True negative (TN): number of true negatives (perfect negative prediction)
- False negative (FN): number of true negatives; imperfect negative prediction

### Error rate

The error rate (ERR) is computed as the fraction of the total number of imperfect predictions to the total number of test data. The finest possible error rate is 0.0, whereas the very worst is 1.0. The minimization of this error rate will be the prime objective of any classifier.

$$ERR = \frac{FP + FN}{TP + TN + FN + FP}$$

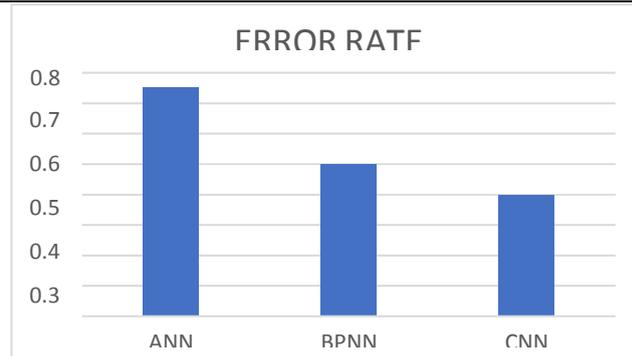


Fig 3: ERROR Rate

ALGORITHM	ERROR RATE
ANN	0.75
BPNN	0.5
CNN	0.4

From the above graph (Fig. 3), the proposed system improved efficiency over the existing system. And recommend the foods based on user information.

## 7. CONCLUSION

People across the universe are becoming more attentive to their health. They are adopting various ways to keep themselves fit. One way is to measure the calorie and nutrition levels of the meal. This method has given a brief review of different calorie and nutrition measurement systems. After discussing various systems, it is found that there is scope for another system that can be developed in order to help patients and dieticians. A system is proposed that uses segmentation and classification using a convolutional neural network to measure the calorie and nutrition levels of the meal. The system is cost-effective and simple. The practical results of the system might boast research in the field of food processing. In the implementation of food recognition system based on image processing the comparative study of various software schemes is done. We proposed a measurement method that estimates the number of calories from a food's image by measuring the area of the food portions from the image and using nutritional facts tables to measure the amount of calorie and nutrition in the food. And calorie is shown in final results with approximate value.

## 8. FUTURE ENHANCEMENT

There is still a lot of potential for future work in food classification, especially with the advancements in technology and the increasing demand for personalised nutrition. There is a need for more transparent and accurate food labelling, especially with the increasing number of processed and packaged foods. Food classification can help improve food labelling by providing consumers with more detailed information about the nutritional value and ingredients of the foods they eat.

## 9. REFERENCES

- [1] Goh, Alex M., and Xiaoyu L. Yann, "Food Image Classification Using Neural Networks," *Int. J. of Electronics Engineering and Applications* 9.3 (2021): 12–22.
- [2] Saha, Dhritiman, and Annamalai "Machine learning techniques for analysis of hyperspectral images to determine the quality of food products: A review." *Current Research in Food Science* 4 (2021): 28–44
- [3] Lo, Frank Po Wen, et al., "Image-based food classification and volume estimation for dietary assessment: A review," *IEEE Journal of Biomedical and Health Informatics* 24.7 (2020): 1926–1939.
- [4] Tahir, Ghalib Ahmed, and Chu Kiong "An open-ended continual learning for food recognition using class incremental extreme learning machines." *IEEE Access* 8 (2020): 82328-82346.
- [5] Menichetti, Giulia, et al., "Machine learning prediction of food processing." *Med Rxiv* (2022): 2021-05.
- [6] He, Jiangpeng, et al., "Multi-task image-based dietary assessment for food recognition and portion size," 2020 *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2020.
- [7] Jiang, Landu, et al., "DeepFood: food image analysis and dietary assessment via deep learning," *IEEE Access* 8 (2020): 47477–47489.
- [8] Oliveira Chaves, Larissa, et al., "Applicability of machine learning techniques in food intake assessment: A systematic review," *Critical Reviews in Food Science and Nutrition* (2021): 1–18.

- 
- [9] Manoharan, "Patient diet recommendation system using K-Clique and deep learning classifiers." *Journal of Artificial Intelligence* 2.02 (2020): 121–130
- [10] Shen, Zhidong, et al., "Machine learning-based approach to food recognition and nutrition estimation." *Procedia Computer Science* 174 (2020): 448–453.
- [11] Yang, M. Chen, D. Pomerleau, and R. Sukthankar, "Food recognition using statistics of pairwise local features," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2249–2256.
- [12] M. Farinella, M. Moltisanti, and S. Battiato, "Classifying food images represented as bags of textons," in *IEEE International Conference on Image Processing*, 2014, pp. 5212–5216.
- [13] Bossard, M. Guillaumin, and L. Van Gool, "Food-101-mining discriminative components with random forests," in *European Conference on Computer Vision*, 2014, pp. 446–461.
- [14] Zhou and Y. Lin, "Fine-grained image classification by exploring bipartite-graph labels," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1124–1133.
- [15] Mao, Runyu, et al., "Visual-aware hierarchy-based food recognition," *International Conference on Pattern Recognition*. Springer, Cham.
- [16] Salvador, N. Hynes, Y. Aytar, J. Marin, F. Ofli, I. Weber, and A. Torralba, "Learning cross-modal embeddings for cooking recipes and food images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3020–3028.
- [17] Aguilar, B. Remeseiro, M. Bolanos, and P. Radeva, "Grab, pay, and eat: Semantic food detection for smart restaurants," *IEEE Transactions on Multimedia*, vol. 20, no. 12, pp. 3266–3275, 2018.
- [18] Bettadapura, E. Thomaz, A. Parnami, G. D. Abowd, and I. Essa, "Leveraging context to support automated food recognition in restaurants," in *IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 580–587.
- [19] G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [20] Martinel, G. L. Foresti, and C. Micheloni, "Wide-slice residual networks for food recognition," in *IEEE Winter Conference on Applications of Computer Vision*, 2018, pp. 567–576.