Cureus Journal of Computer Science

Open Access Original Article

Deep Learning-Based Nutritional Analysis of Food Images Using EfficientNetB4 and Multilayer Perceptron

Received 09/05/2025
Review began 09/11/2025
Review ended 09/21/2025
Published 09/29/2025

DOI:
https://doi.org/10.7759/s44389-025-09841-z

Anshul Diwakar [1] , Rashi Agarwal [1]

1. Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, IND

Corresponding author: Rashi Agarwal, rashi@hbtu.ac.in

Abstract

Nutritional estimation is the growing research area within the domain of computer vision and deep learning for a balanced and healthy diet. Innovations in computer vision recently have shown greater improvements in food classification and analysis of nutritional values. Convolutional neural networks (CNNs) have resulted in a greater efficiency in the extraction of visual characteristics of food images. This paper proposes a fine-tuned EfficientNetB4 model along with multilayer perceptron (MLP), which resulted in an accuracy of 85% for food image classification. MLP regressor takes the feature vectors extracted from CNN backbone EfficientNetB4 for predicting nutritional values. The model is trained on Food-101 dataset with nutritional values metadata sourced from various sources to estimate nutritional targets. However, nutritional value estimation still presents a challenge due to mixed dishes and indefinite portion sizes. The dataset is preprocessed by normalization and resizing the food images, while nutritional labels are scaled using StandardScaler(). The model is also evaluated on various metrics such as mean absolute error and mean squared error and resulted in 0.30 mean absolute error and 0.21 mean squared error for estimating the nutritional values. Thus EfficientNet, complemented with an MLP, offers a robust framework for tackling the challenges posed by large-scale and varied datasets like Food-101 in the context of nutritional analysis. This hybrid model leverages the strengths of CNNs for fine-grained image feature extraction and the ability of MLPs to handle the final classification tasks, thereby enhancing the accuracy and reliability of predictions in real-world applications.

Categories: Computer Vision, Machine Learning (ML), Deep Learning
Keywords: efficientnetb4, cnn, mlp, food-101 dataset, nutritional values

Introduction

In light of the present push for dietary health and calorie literacy, automated food recognition systems can greatly facilitate the task of determining the caloric content and nutritional value of food, or even set up a diet tracking system to assist patients or obese individuals in avoiding unhealthy food choices. The significance of nutritional information, finding credible sources, and interpreting it for individual requirements are key to responsible use in diet planning and health management. Integration of Convolutional Neural Networks (CNNs) into numerous areas has advanced machine learning and deep learning frameworks. CNNs are best at processing picture data but can also handle other jobs. Various study results suggest certain CNN integration methods. CNNs are integrated using visualization approaches to understand neural network behavior beyond performance measurements. Visualizing input and output pictures may reveal CNN performance, especially in image restoration. These solutions improve CNN architecture efficiency without affecting performance [1]. CNNs learn signal direction from many antenna arrays for direction-finding. This shows their flexibility beyond picture categorization. CNN integration with other deep learning architectures for difficult classification and regression problems has improved performance when CNNs and Recurrent Neural Networks are merged [2]. CNN complexity and resource intensity may be addressed by integrating quantization-based pooling algorithms inspired by the bag-of-features paradigm. This method trains lightweight CNN architectures to classify images at different sizes by lowering parameter count. This significantly improves CNN accessibility and efficiency [3]. CNN-based image denoising approaches have been developed for image processing. CNNs are better at handling noise-laden pictures than standard denoising algorithms, making them useful for image restoration [4]. Evolutionary methods optimize CNN structures for particular applications, balancing performance and computational economy. CNN is more applicable to many deployment situations because of its multi-objective methodology, which develops design solutions for distinct computational limitations [5]. CNNs and transfer learning are widely used in medical imaging, such as MR brain image categorization. In accuracy and adaptability-intensive sectors, CNNs bring great value by customizing pre-trained models [6]. Quantum CNNs are being studied for their ability to speed up processing while preserving accuracy; however, quantum computing technology is still developing [7]. CNN architecture EfficientNet strikes a good balance between performance and computational needs, making it a versatile choice for food categorization applications to boost computational speed and efficiency. CNNs' integration approaches increase picture classification, demonstrating their flexibility and adaptability across technology landscapes. A key design element of EfficientNet is its capacity to deliver state-of-the-art performance with fewer parameters and quicker inference times than standard models. For example, EfficientNet-B7 had 84.3% top 1 accuracy on ImageNet despite being 8.4 times smaller and 6.1 times quicker than the previous best convnet. Multiple empirical investigations across applications demonstrate the benefits of the EfficientNet design. Medical image analysis is successful using EfficientNet. EfficientNet models provide great accuracy in brain tumor classification tests, improved as model complexity grows, reaching 99.64% for the EfficientNet-B7 model. These high accuracy rates show the model's clinical value in medical diagnosis. EfficientNet-B4 also classifies paddy illnesses with over 96% accuracy in agriculture.

In this research, we make use of Food-101, a large dataset of food photos used for computer vision and machine learning problems relating to food image detection and categorization. A total of 101 food categories and 101,000 photos make up this collection. The dataset was generated as part of a broader multimodal dataset project that integrates textual information with visuals to improve recipe recognition [8]. The dataset uses visual and textual data to assist simple categorization to advanced recipe recognition applications [8]. The Food-101 dataset has been widely used in research to enhance food picture categorization using deep feature extraction and CNN model fine-tuning. Using deep features from pre-trained models like AlexNet and VGGl6 for feature extraction and concatenation improves classification performance. The Food-101 dataset is essential in computer vision for food picture analysis, allowing researchers to build and test new food identification methods [8,9]. Due to the large degree of visual variation in food and its intrinsic unpredictability, food picture recognition is a fine-grained categorization problem. The high degree of similarity among subcategories and the large amount of variety across classes makes food image identification more difficult than conventional image recognition. The rise of nutrition-related illnesses throughout the globe has brought more attention to the significance of maintaining good eating habits. Some cancers, food intolerance responses, obesity, and malnutrition may be prevented with a healthy diet. Several methods exist that allow us to manually record our food intake and identify foods before they are eaten.

A key deep neural network design used in this paper is the multi-layer perceptron (MLP), with several hidden layers for estimating nutritional targets. Using the models' design to extract features useful for food item classification and nutritional content analysis is the key to applying them to the photos of food from 101 distinct categories, with 1,000 photos per category, making up the Food-101 dataset. Dietary monitoring and health apps are finding more and more uses for this large dataset, which offers a solid foundation for training models to detect and evaluate nutritional values. The use of transfer learning, which involves pretreating models on big picture datasets and then fine-tuning them on Food-101, greatly improves the performance of these models. The scalability of EfficientNet can be especially useful in this case because it lets the model be fine-tuned to accommodate different sized and complicated food images without sacrificing accuracy alongside MLP for nutritional values estimation. MLPs are ideal for applications that need complicated data patterns because their design captures complex connections [10]. Feedforward artificial neural networks like MLPs send data from input nodes to output nodes via hidden layers [11]. MLPs' accuracy in complicated mappings makes them popular in classification issues. Sigmoid, hyperbolic tangent, and other activation functions inject non-linearity into the model and impact decision boundaries, affecting network performance [12].

Literature review

CNNs are highly effective in recognizing patterns and classifying images in various fields. It is possible that the nutritional information extracted from food photographs is comparable to the complex patterns extracted from biological datasets using CNNs in pharmacogenomics [13]. Also, deep networks can learn nuances in data, which helps with food variety identification and composition analysis. EfficientNet improves the scalability of CNNs for high-dimensional datasets such as Food-101 by achieving a balance between accuracy and efficiency [14]. To facilitate feature extraction from complicated and dynamic data, such as food images, these models can be fine-tuned for specific purposes, such as nutritional analysis, after being pre-trained on big datasets. A CNN architecture fine-tuned model achieved 79.86% accuracy on the Food-101 dataset [9]. Due to its excellent food identification algorithm benchmark, the dataset is essential for developing nutritional monitoring apps. These algorithms are needed to monitor food consumption and help diabetics estimate carbohydrate intake using food recognition systems [15]. Researchers have also considered adding diffusion model-generated synthetic food visuals to Food-101 as real-world food picture classification applications face data shortages and class imbalance.

Using CNN designs such as SqueezeNet and VGG-16 represents a prominent technique. The study found that SqueezeNet - known for its small size - was able to get a 77.20% accuracy with fewer parameters than competing models, making it a viable choice for real-world deployment. As a result of being able to extract more complicated features from food photos, the VGG-16 network, which now has higher depth [16], significantly improved accuracy to 85.07%. Similar to the previous study, this one used CNNs to successfully categorize a dataset of 16,643 food photos, reaching a remarkable accuracy of 92.86%. This shows that CNNs can handle food category data distributions that are quite non-linear [17].

Furthermore, when it comes to food identification and recognition tasks, CNNs outperform more conventional methods such as support-vector-machine based approaches that rely on handmade characteristics. Due to the inherent visual complexity of food photos, a study found that CNNs improved accuracy and made good use of color as an important element in the extraction process. In order to detect and identify food photos, Subhi and Md. Ali [18] debuted a newly created database for native Malaysian cuisines, which include 5,800 food photos divided into 11 categories. In their groundbreaking work, Mezgec and Seljak [19] introduced a CNN called NutriNet that could detect and recognize pictures of food and drink. With 225,953 photos over 520 classes, this architecture was taught to identify foods and drinks with an accuracy of 86.72%. Additionally, the architecture had a 55% success rate when tested on images that people with Parkinson's disease had taken themselves using their smartphones. The model was utilized by a mobile app to evaluate the diets of individuals with Parkinson's disease. Recently, novel approaches integrating YOLO-Siam networks were created by making use of deep learning architectures as Tiny-YOLO. These approaches demonstrated potential for widespread implementation in real-world settings due to their ability to eliminate manual marking while still providing general recognition accuracy [20].

Food-based dietary recommendations need well-designed food composition databases. These databases must be complete, current, and properly characterize food products to avoid mistakes in nutrient intake assessments and nutrient source identification [21]. The ONQI provides science-based nutritional advice. This index scores food micronutrient and macronutrient qualities to assist customers pick meals that meet dietary standards [22]. The Nutrition Labelling and Education Act requires product labels to provide nutrition information to improve dietary choices. Mandatory labelling seeks to give useful information, but its limits and how well it transmits nutritional information must be considered [23]. Effective nutrition labelling, particularly on restaurant menus, promotes healthy eating. Detailed nutritional information impacts customer decisions and the perceived value of better meals, demonstrating a preference for healthy foods even at a greater cost [24]. Nutrition and exercise guidelines from the American Cancer Society promote cancer prevention and wellness. This handbook emphasizes the significance of community support for healthy decisions [25]. Nutrition is important in illness treatment. In therapeutic settings, nutrition treatment and meal planning are essential to controlling different health issues and improving patient outcomes [26]. Table *1* shows existing deep learning models and results with various datasets for classifying foods.

Authors	Food dataset used	Model	Techniques	Accuracy
Sengur et al. [9]	Food-101	AlexNet and VGG16	End-to-end CNN retraining	79.86%
Islam et al. [17]	Food 11 Dataset	Inception V3	Custom CNN	92.86%
Mezgec and Seljak [19]	UNIMIB2016	NutriNet	Custom CNN	86.72%
Pouladzadeh and Shirmohammadi [27]	FoodDD Dataset	Pre-trained CNN	Submodular Optimization Method	94.11%
Yanai and Kawano [28]	UEC Food 100	CaffeeNet	Fine-Tuning Deep CNN	Top5 - 94.85%
Pandey et al. [29]	Food-101	Ensemble	Ensemble with AlexNet, GoogleNet, ResNet	Top10 - 95.95%
Liu et al. [30]	UEC Food 256	GoogleNet Inception	Custom CNN DeepFood	Top5 - 95.2%

TABLE 1: Existing deep learning models and results with various datasets

CNN, Convolutional Neural Network

Materials And Methods

In the following section, we propose a framework that can be used to model food image dataset multi-classification problems using EfficientNetB4 and nutritional values estimation through MLP.

Dataset and preprocessing

The Food-101 [31] dataset comprises 101 food items and each Food-101 category has 1,000 photos, giving valuable data for food identification algorithm training and testing. Initially, the original dataset is divided into two parts: 75% for training and 25% for testing. Furthermore, training data is split again in a 9:1 ratio. Following this, the images needed for training, validation, and testing are 68175, 7575, and 25250, respectively.

FIGURE 1: Sample food images of Food-101 dataset

The Food-101 dataset [31] was used for this research. The Food-101 dataset is used in accordance with its intended purpose for academic research and non-commercial use, as provided by the original authors. All rights belong to the dataset authors.

The images of the Food-101 dataset are downsized to a specific dimension 380 × 380 pixels that is compatible with the EfficientNetB4 model's input layer and given to preprocessing function, preprocess_input by Keras built in function. Data augmentation is commonly used to prevent overfitting and improve model generalization. In order to generate new datasets, methods including horizontal flipping, scaling, and rotation are used. The rotation range was set to 20, along with the width and height shift ranges, was set to 0.2. The images are also horizontally flipped and fillest mode is set to nearest to fill the empty space on the corner and enhance the image, in addition, a shear range and zoom range of 0.2. Both the size of the training dataset and the model's ability to learn characteristics that are invariant across small changes are enhanced by this.

Proposed methodology

We started with the input image from Food-101 dataset which is resized to 380 × 380 m pixels prior to feature extraction and then fine-tuned for enhancing the performance. MLP is used to predict the nutritional content after the food image has been identified correctly as shown in Figure *2*. The detailed structure of working of EfficientNetB4 and MLP to predict the nutritional values is represented in Figure *3*.

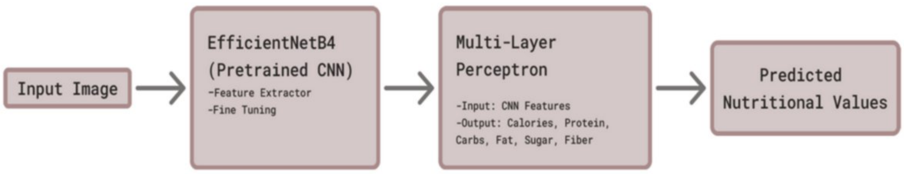

FIGURE 2: Framework for the estimation of nutritional values

In this paper, data augmentation and an image size of (380,380,3) are used for the initial pre-processing of the dataset. The EfficientNetB4 model was initially set up by Keras, and the feature selection output from the model is used to launch a Global Average Pooling layer and then dropout layer to drop 50% of neurons. A dense layer of 512 units with an activation function set to 'relu' is subsequently added. Next, we incorporate the dropout layer into the model, setting the dropout rate to 30%. A dense layer with 101 classes and an activation function set to 'SoftMax' is used to take the output. Then, during feature extraction for 15 epochs, the pre-trained model is frozen in order to stop learning from all of the layers. Thereafter, the last 150 layers are unfrozen for fine tuning the model over 10 epochs to measure appropriate metrics for loss and accuracy. The model is then compiled with the learning rate of 1e-5 using Adam optimizer along with various callback functions such as EarlyStopping over metric val_accuracy to prevent overfitting, ReduceLROnPlateau to reduce the learning rate when val_accuracy stops improving, and ModelCheckpoint to preserve best model. The use of transfer learning, which involves pre-trained models on vast datasets and then fine-tuning them on Food-101, greatly improves the performance of these models. Using the models' design to extract features useful for food item classification and nutritional content analysis is the key to applying them to the Food-101 dataset. The scalability of EfficientNet can be especially useful in this case because it enables the model to be fine-tuned to accommodate different sized and complicated food images without affecting performance. The accuracy of food item identification and portion estimation is of the utmost importance in nutritional analysis.

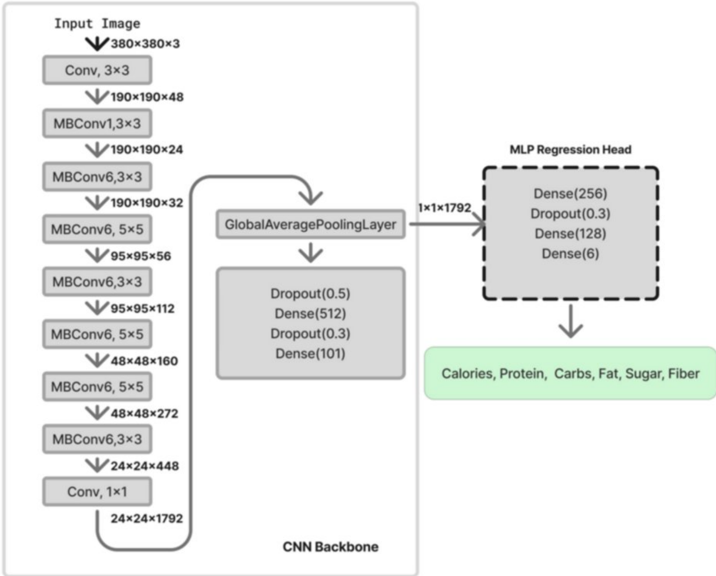

FIGURE 3: Proposed architecture

MLP is further used for nutritional value assessment, which takes input, i.e., shape from Global Average Pooling layer of CNN backbone and then a dense layer with 256 units, subsequently adding dropout layer dropping 30% neurons. The dense layer serves as the output, and its six neurons stand for the following nutritional values: calories, protein, carbohydrates, fat, sugar, and fiber, as shown in Figure *3*. To predict the nutritional values, data is being sourced from external sources and maintained in a CSV file with food class as an index. The names of food classes are used to map class indices using the train generator. In order to forecast dietary goals, the MLP uses a pre-trained feature extractor based on CNN to retrieve information from training images. Each training image's class index is mapped using the names of food classes and subsequently their nutritional value. We used Mean Squared Error (MSE) and Mean Absolute Error (MAE) for evaluating the regression performance of the nutritional value predictor. By combining the power of CNNs for coarse-grained feature extraction with that of MLPs for final classification tasks, this hybrid model improves the precision and consistency of predictions for practical use. When combined with MLPs, CNN EfficientNet offers a robust set of tools for extracting nutritional information from food photos. By using these designs to datasets like Food-101, researchers in the fields of nutrition and dietetics could gain invaluable insights from data and improve the accuracy of classification.

MLPs are known for their pattern recognition, modelling, and prediction abilities across fields. CNNs are best at processing picture data but can also handle other jobs in numerous areas with advanced machine learning and deep learning frameworks. Visualizing input and output pictures may reveal CNN performance, especially in image restoration. Despite CNNs' dominance in picture tasks, MLPs continue to see heavy use in many domains because of their versatility and capacity to mimic complex nonlinear interactions. With the help of CNN features, MLPs can categorize nutritional values when CNNs extract broad traits. As a result of their ability to learn intricate decision boundaries free of structural biases, MLPs are able to categorize alongside CNNs.

EfficientNeB4 is a part of the EfficientNet family used for classification of images with an input resolution of 380 × 380 for improved accuracy with fewer parameters. EfficientNet, a complex model design, drastically changes how Convolutional Neural Networks (ConvNets) are scaled to improve performance and economy. Innovative scaling that balances network depth, breadth, and resolution with a single compound coefficient is EfficientNet's foundation. This balanced approach improves model performance more than typical scaling strategies, which concentrate on one or two dimensions and ignore others. The architecture's movable inverted bottleneck convolution and deep separable convolutions provide efficient feature extraction with good accuracy, recall, and F1-scores and low losses. EfficientNet demonstrates superior performance in document image classification tasks, providing a more lightweight and effective model relative to conventional, heavier CNNs. Its transfer learning features enable effective performance on smaller in-domain datasets, demonstrating its adaptability and efficiency across many contexts.

Results And Discussion

Experimental setup

The experiments are carried out on desktop machine running on Windows 11 OS, AMD Ryzen 5 3550H CPU with Radeon Vega Mobile Gfx (2.10 GHz), and 16 GB of RAM. The training and interpretation are performed on CPU; no GPU was used to have the lightweight setup. Although the EfficientNetB4 was computationally heavy, it was still deployable on CPU.

The feature extraction and fine-tuning of model are done over 25 epochs in total. The throughput of the pipeline was increased by the use of effective data generators and batch processing; the batch size was set to 32. Hyperparameters and loss function are used. StandardScaler() ensures faster convergence of the MLP during training. This is implemented on Python v3.9.12 alongside libraries Tensorflow(v2.19.0), Numpy(v2.0.2), Pandas (v2.2.3), Matplotlib (v3.9.4), and scikit-learn (v1.6.1).

Feature Extraction Results

The proposed model EfficientNetB4 has been trained for 15 epochs for feature extraction, and corresponding metrics have been calculated for visual representation of model's performance during training. Figure *4* and Figure *5* represent the accuracy and loss incurred during training and on validating, respectively. It is noted that the validation accuracy reached 70.29% at 15th epoch, as shown in Figure *4*, with validation loss dropping to 1.102, as demonstrated in Figure *5*.

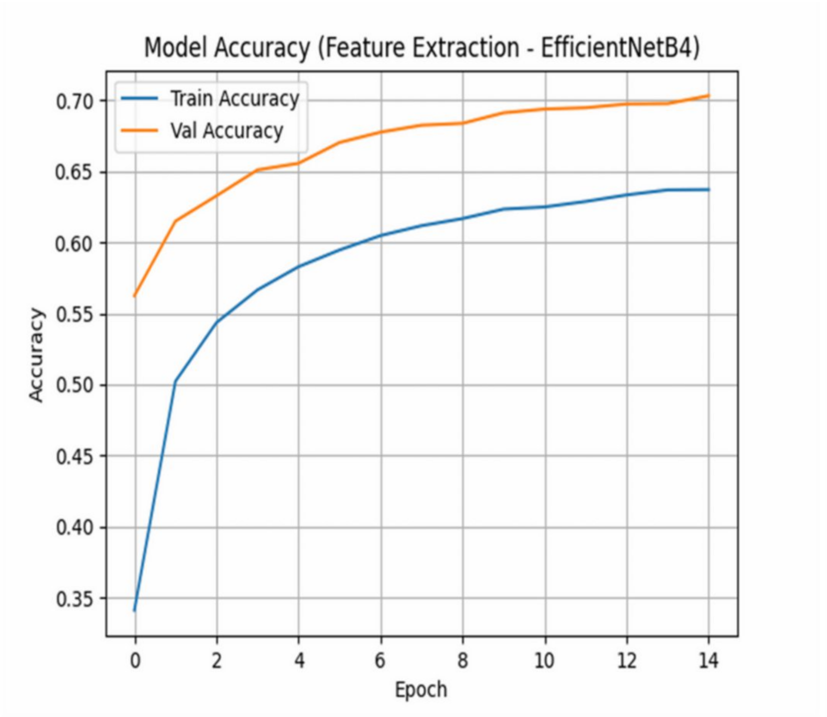

FIGURE 4: Training vs validation accuracy during feature extraction of EfficientNetB4 model

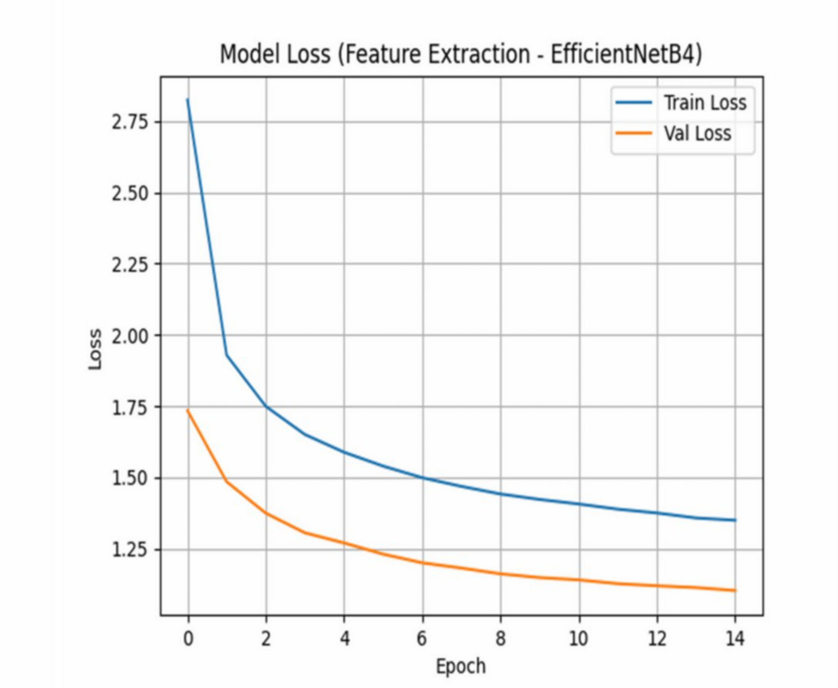

FIGURE 5: Training vs validation loss during feature extraction of EfficientNetB4

Fine-Tuning Results

Furthermore, for fine-tuning, the feature-extracted model is trained for 10 epochs using pre-trained weights to enhance performance in food image classification and improve accuracy. The fine-tuned model demonstrates great improvements in validation accuracy, increasing from 70.29% to 79.68%, as shown in Figure *6*. Correspondingly, Figure *7* illustrates a reduction in validation loss during fine-tuning, indicating that the model is effectively learning new features.

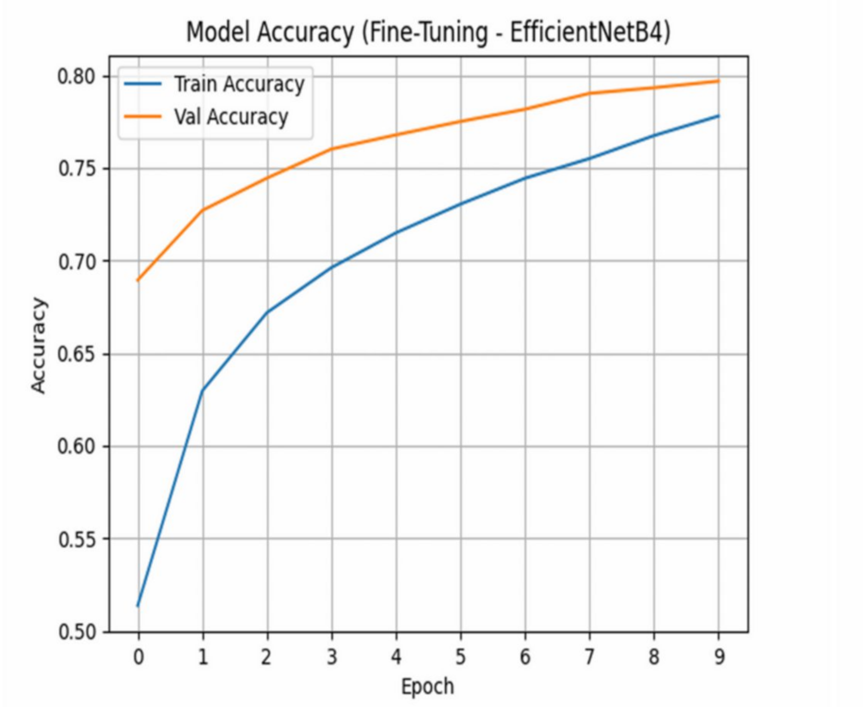

FIGURE 6: Training vs validation accuracy during the fine-tuning of feature-extracted EfficientNetB4 model

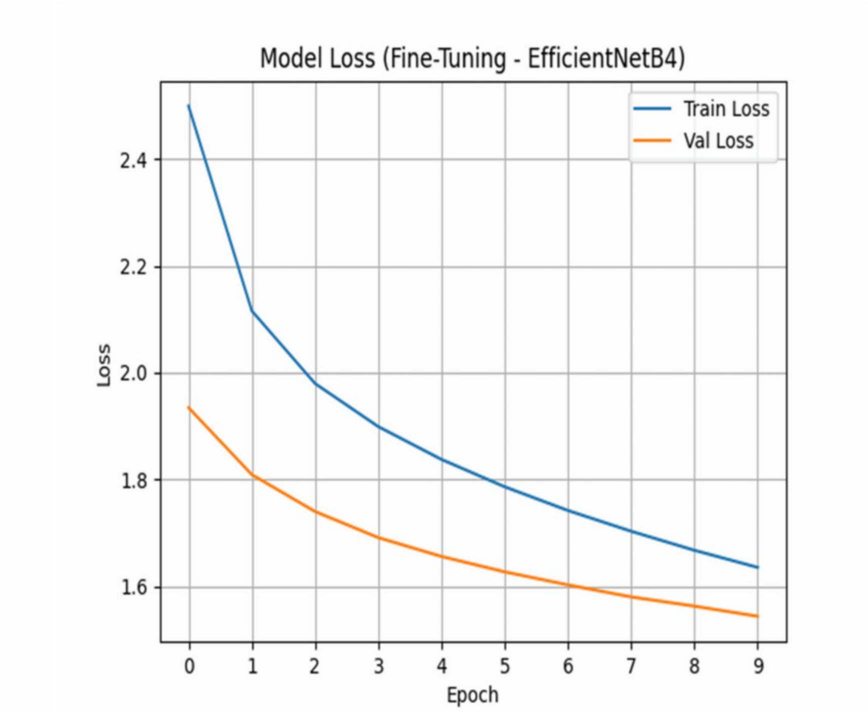

FIGURE 7: Training vs validation loss during the fine-tuning of feature-extracted EfficientNetB4 model

Confusion Matrix Performance

The confusion matrix in Figure 8 indicates how the classification model is performing by calculating correct and incorrect predictions of each respective class labelled 0-100 of the Food-101 dataset. The classification report has been generated for a detailed summary of the model's performance, including metrics such as precision - 0.85, recall - 0.85, and F1-score - 0.85 for each class in the dataset.

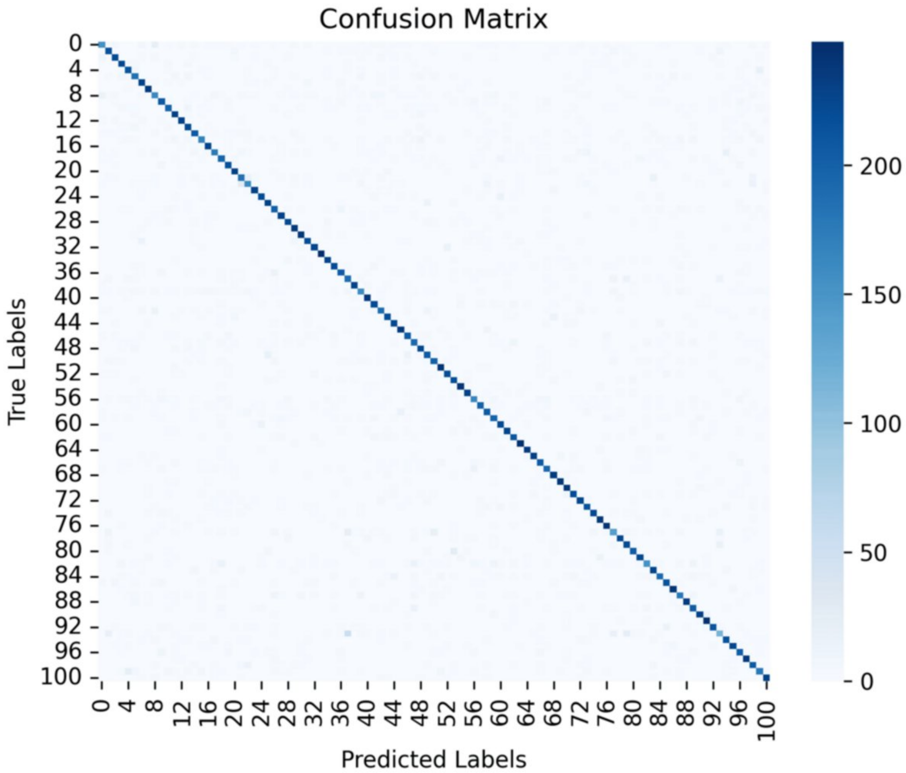

FIGURE 8: Confusion matrix of EfficientNetB4 on Food-101

Food classes Bibimbap, Edamame, Macarons, and Seaweed salad at labels 7, 33, 63, and 88, respectively, have shown great precision, recall, and F1-score, as indicated in Table *1*.

Class	Index in dataset	Precision	Recall	F1-score
Bibimbap	7	0.94	0.94	0.94
Edamame	33	0.99	0.99	0.99
Macarons	63	0.95	0.98	0.97
Seaweed salad	88	0.95	0.94	0.94

TABLE 2: Precision, recall, and F1-score of EfficientNetB4 for few items in Food-101 dataset

For final evaluation 25,250 images are used for testing, and an accuracy of 85% is achieved for classification of images of dataset using EfficientNetB4 coupled with MLP.

MLP Training Results

Furthermore, MLP training of model for 50 epochs is done to predict nutritional values using various metrics as represented in Figure *9* and Figure *10*.

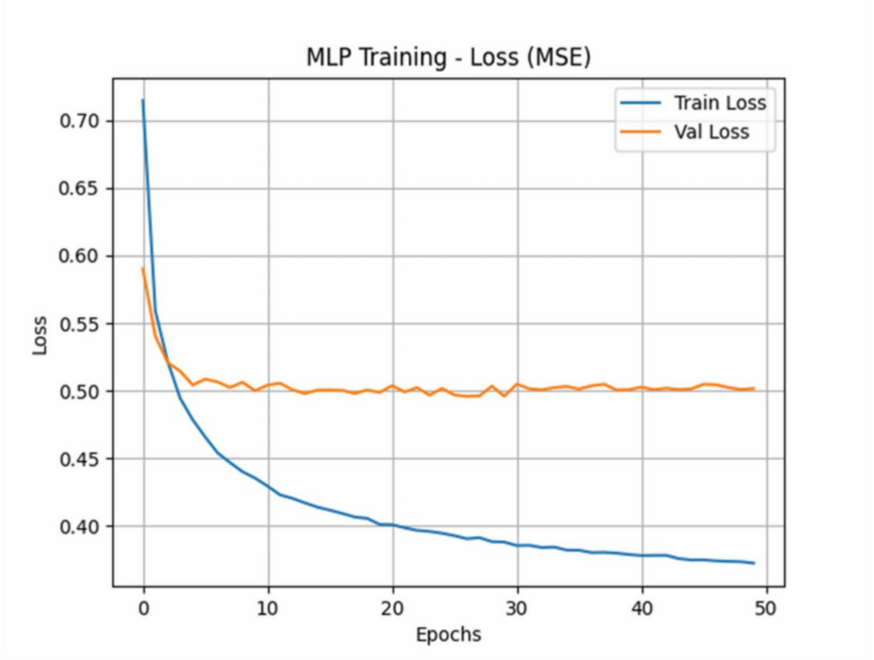

FIGURE 9: Mean squared error during MLP training for EfficientNetB4

MLP, Multilayer Perceptron

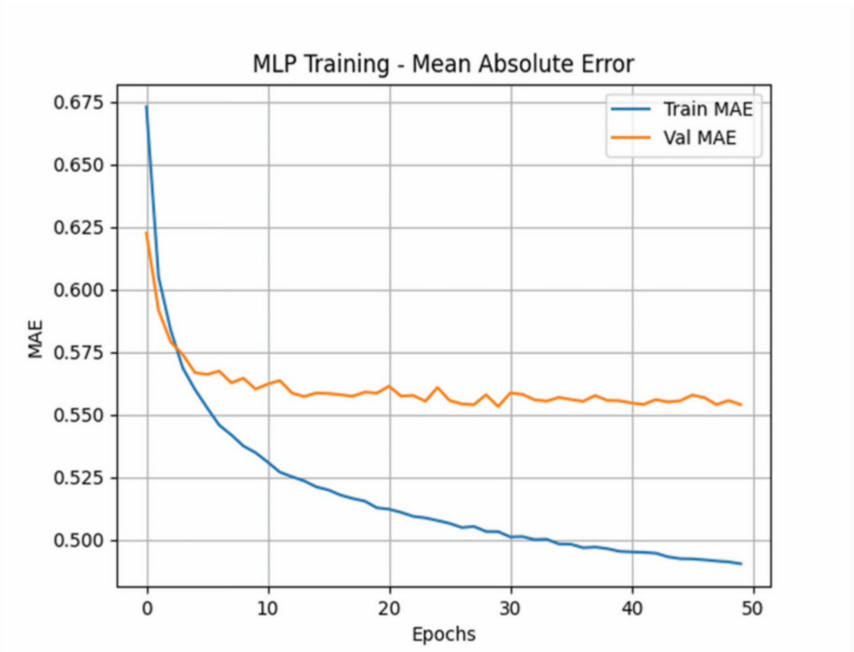

FIGURE 10: Mean absolute error during MLP training for EfficientNetB4

MLP, Multilayer Perceptron

Predicted Sample Results

The predicted results of food image classification through EfficientNetB4 and nutritional estimation by MLP are shown in Figure *11* and Figure *12*. The figures shows that we have test image, and we have got corresponding predicted image through CNN and estimation through MLP, which can be helpful in tailoring dietary recommendations for individuals to avoid various chronic diseases.

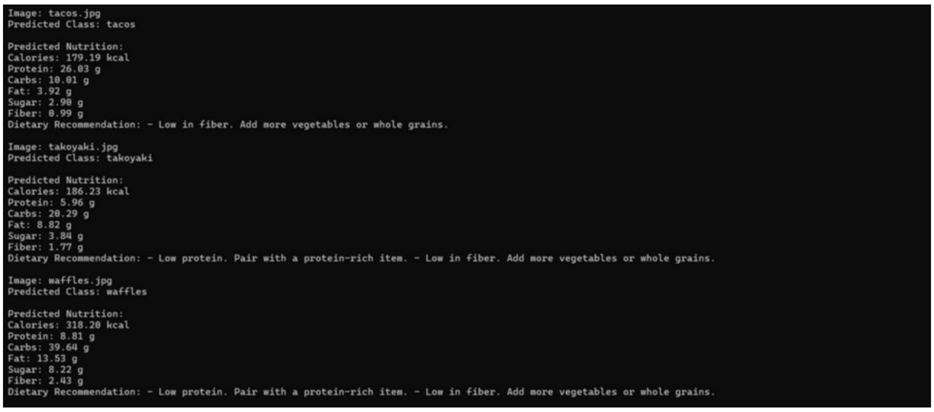
```
Image: tacos.jpg
Predicted Class: tacos

Predicted Nutrition:
Calories: 179.19 kcal
Protein: 26.03 g
Carbs: 10.01 g
Fat: 3.92 g
Sugar: 2.90 g
Fiber: 0.99 g
Dietary Recommendation: - Low in fiber. Add more vegetables or whole grains.

Image: takoyaki.jpg
Predicted Class: takoyaki

Predicted Nutrition:
Calories: 186.23 kcal
Protein: 5.96 g
Carbs: 20.29 g
Fat: 8.82 g
Sugar: 3.84 g
Fiber: 1.77 g
Dietary Recommendation: - Low protein. Pair with a protein-rich item. - Low in fiber. Add more vegetables or whole grains.

Image: waffles.jpg
Predicted Class: waffles

Predicted Nutrition:
Calories: 318.20 kcal
Protein: 8.81 g
Carbs: 39.64 g
Fat: 13.53 g
Sugar: 8.22 g
Fiber: 2.43 g
Dietary Recommendation: - Low protein. Pair with a protein-rich item. - Low in fiber. Add more vegetables or whole grains.
```

FIGURE 11: Sample result I

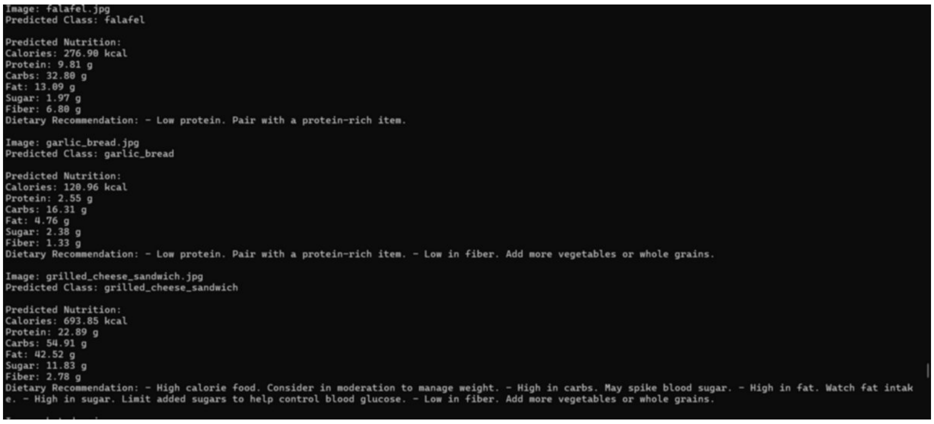
```
Image: falafel.jpg
Predicted Class: falafel

Predicted Nutrition:
Calories: 276.90 kcal
Protein: 9.81 g
Carbs: 32.80 g
Fat: 13.09 g
Sugar: 1.97 g
Fiber: 6.80 g
Dietary Recommendation: - Low protein. Pair with a protein-rich item.

Image: garlic_bread.jpg
Predicted Class: garlic_bread

Predicted Nutrition:
Calories: 120.96 kcal
Protein: 2.55 g
Carbs: 16.31 g
Fat: 4.76 g
Sugar: 2.38 g
Fiber: 1.33 g
Dietary Recommendation: - Low protein. Pair with a protein-rich item. - Low in fiber. Add more vegetables or whole grains.

Image: grilled_cheese_sandwich.jpg
Predicted Class: grilled_cheese_sandwich

Predicted Nutrition:
Calories: 693.85 kcal
Protein: 22.89 g
Carbs: 54.91 g
Fat: 42.52 g
Sugar: 11.83 g
Fiber: 2.78 g
Dietary Recommendation: - High calorie food. Consider in moderation to manage weight. - High in carbs. May spike blood sugar. - High in fat. Watch fat intak
e. - High in sugar. Limit added sugars to help control blood glucose. - Low in fiber. Add more vegetables or whole grains.
```

FIGURE 12: Sample result II

Conclusions

This paper proposes a model based on the transfer-learning-fine-tuned EfficientNetB4 model, which achieves an accuracy rate of 85% on the Food-101 dataset and also mean absolute error of for nutritional value predictor grounded on MLP. For nutritional analysis on large-scale and diverse datasets like Food-101, a strong framework for dealing with these issues can be achieved by combining EfficientNetB4 with an MLP. The fused model can improve the prediction of food categories and, possibly, nutritional values from known data of food items within those categories by utilizing CNNs to extract detailed visual features and MLPs for nuanced classification. There has been a shift in focus in recent years towards dietary causes of chronic diseases and other health problems. Efforts to create more precise techniques of assessing individual food intake have been spurred by the growing demand for systems that can monitor and advice on dietary intake. In order to keep track of what people eat, a healthcare monitoring system and dietary assessment are required. They plan their meals in advance so they can monitor their nutrient intake. The preceding considerations, however, show that there are substantial obstacles to dietary assessment, especially when it comes to precisely determining the nutrient consumption. There are still many obstacles to reliably estimating nutrient intake, even if present approaches have demonstrated promise in tackling the problem of food item recognition. One difficulty is that, depending on their appearance, not all of the elements in a dish will be easily identifiable by sight or camera. Because of differences in serving sizes and the impact of cooking, calculating the nutritional value is still challenging even when all of the ingredients are known. Furthermore, the current technology for estimating volumes is neither accurate nor robust enough. This is because things like food preparation, cooking, presentation, and consumption can have a significant impact on the final form and appearance of food items. Combining computer vision with artificial intelligence offers a particularly promising technique for food detection and analysis, such as calculating meal portion size and nutrient content. However, accurately measuring dietary intake is still a challenging research problem. Future research should improve by adding a more diverse dataset and developing a more accurate model of food object identification, among other areas where it falls short. Future iterations of this research should suggest a more precise approach for measuring calories. While there is a great need for research into effective diet management strategies in the health and food sciences, this research is just the beginning of a line of inquiry that will have far-reaching positive effects on society. Image data augmentation was a game-changer for this research preprocessing phase, and the model's fine-tuning was the study's most innovative component. Our results on controlled datasets show promising per-image nutrient prediction, but more testing is needed for real-world deployment, particularly for mixed plates and changing portion sizes.

In the future using computer vision without human intervention could help create generalizable models that can sort mixed food items which still possess challenges due to variability in food and handling portion sizes. CNNs are great for sorting food pictures into groups because they can instantly pull out features from photos without having to be manually added and with optimized pipeline real-time dietary feedback can be provided through mobile app integration. The current work shows feasibility of using datasets like Food-101, which has a lot of categories of food, and deep learning models like CNNs (ResNet 50 and EfficientNet) and MLPs to look at the nutritional values in a plate of food, still several potential extensions and improvements can be further incorporated such as including user-specific data like health conditions and dietary restrictions to recommend healthier alternatives. This area is still developing, but it has the potential to change how we recognize and classify foods. This would make sure that whatever we consume comply with the highest nutritional requirements and encourage healthier eating habits through better information.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Anshul Diwakar, Rashi Agarwal

Acquisition, analysis, or interpretation of data: Anshul Diwakar

Drafting of the manuscript: Anshul Diwakar

Critical review of the manuscript for important intellectual content: Anshul Diwakar, Rashi Agarwal

Supervision: Rashi Agarwal

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

This article was previously presented at the International Conference on Next Generation Technology in Science and Engineering (ICNGTSE - 2025) at Pranveer Singh Institute of Technology, Kanpur (PSIT), on 22–23 August 2025.

References

1. Protas É, Bratti JD, Gaya JFO, Drews P, Botelho SSC: Visualization methods for image transformation convolutional neural networks. IEEE Transactions on Neural Networks and Learning Systems. 2019, 30:2231-2243. 10.1109/tnnls.2018.2881194
2. Uckun FA, Ozer H, Nurbaş E, Onat E: Direction finding using convolutional neural networks and convolutional recurrent neural networks. 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey. 2020, 1-4. 10.1109/SIU49456.2020.9302448
3. Passalis N, Tefas A: Training lightweight deep convolutional neural networks using bag-of-features pooling. IEEE Transactions on Neural Networks and Learning Systems. 2018, 30:1705-1715. 10.1109/tnnls.2018.2872995
4. Thakur RS, Yadav RN, Gupta L: State-of-art analysis of image denoising methods using convolutional neural networks. IET Image Processing. 2019, 13:2367-2380. 10.1049/iet-ipr.2019.0157
5. Lu Z, Whalen I, Dhebar Y, Deb K, Goodman ED, Banzhaf W, Boddeti VN: Multiobjective evolutionary design of deep convolutional neural networks for image classification. IEEE Transactions on Evolutionary Computation. 2021, 25:277-291. 10.1109/tevc.2020.3024708
6. Kaur T, Gandhi TK: Deep convolutional neural networks with transfer learning for automated brain image classification. Machine Vision and Applications. 2020, 31:20. 10.1007/s00138-020-01069-2

7. Rajesh V, Naik UP, Mohana: Quantum convolutional neural networks (QCNN) using deep learning for computer vision applications. 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), Bangalore, India. 2021, 728-734. 10.1109/rteict52294.2021.9574030
8. Wang X, Kumar D, Thome N, Cord M, Precioso F: Recipe recognition with large multimodal food dataset. 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, 2015. 2015, 1-6. 10.1109/icmew.2015.7169757
9. Sengur A, Akbulut Y, Budak U: Food image classification with deep features. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey. 2019, 1-6. 10.1109/idap.2019.8875946
10. Bisong E: The multilayer perceptron (MLP). Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA; 2019. 401-405. 10.1007/978-1-4842-4470-8_31
11. Yüksel E, Soydaner D, Bahtiyar H: Nuclear binding energy predictions using neural networks: Application of the multilayer perceptron. International Journal of Modern Physics E. 2021, 30:2150017. 10.1142/s0218301321500178
12. Isa IS, Saad Z, Omar S, Osman MK, Ahmad KA, Sakim HAM: Suitable MLP network activation functions for breast cancer and thyroid disease detection. 2010 Second International Conference on Computational Intelligence, Modelling and Simulation, Bali, Indonesia. 2010, 39-44. 10.1109/cimsim.2010.93
13. Vaz JM, Balaji S: Convolutional neural networks (CNNs): concepts and applications in pharmacogenomics. Molecular Diversity. 2021, 25:1569-1584. 10.1007/s11030-021-10225-3
14. Carranza-García M, García-Gutiérrez J, Riquelme JC: A framework for evaluating land use and land cover classification using convolutional neural networks. Remote Sensing. 2019, 11:274. 10.3390/rs11030274
15. Anthimopoulos MM, Gianola L, Scarnato L, Diem P, Mougiakakou SG: A food recognition system for diabetic patients based on an optimized bag-of-features model. IEEE Journal of Biomedical and Health Informatics. 2014, 18:1261-1271. 10.1109/jbhi.2014.2308928
16. Yadav S, Alpana, Chand S: Automated food image classification using deep learning approach. 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India. 2021, 542-545. 10.1109/icaccs51430.2021.9441889
17. Islam MT, Karim Siddique BMN, Rahman S, Jabid T: Image recognition with deep learning. 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Bangkok, Thailand. 2018-10, 106-110. 10.1109/ICIIBMS.2018.8550021
18. Subhi MA, Md. Ali S: A deep convolutional neural network for food detection and recognition. 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Sarawak, Malaysia. 2018, 284-287. 10.1109/IECBES.2018.8626720
19. Mezgec S, Seljak BK: NutriNet: A deep learning food and drink image recognition system for dietary assessment. Nutrients. 2017, 9:657. 10.3390/nu9070657
20. Wang Y, Deng H, Wu J, Zeng X: Food image recognition and food safety detection method based on deep learning. Computational Intelligence and Neuroscience. 2021, 2021:1-13. 10.1155/2021/1268453
21. Leclercq C, Turrini A, Valsta LM: Food composition issues - implications for the development of food-based dietary guidelines. Public Health Nutrition. 2001, 4:677-682. 10.1079/phn2001153
22. Katz DL, Njike VY, Faridi Z, Rhee LQ, Reeves RS, Jenkins DJA, Ayoob KT: The stratification of foods on the basis of overall nutritional quality: The overall nutritional quality index. American Journal of Health Promotion. 2009, 24:133-143. 10.4278/ajhp.080930-quan-224
23. Mojduszka EM, Caswell JA: A test of nutritional quality signaling in food markets prior to implementation of mandatory labeling. American Journal of Agricultural Economics. 2000, 82:298-309. 10.1111/0002-9092.00026
24. Hwang J, Lorenzen CL: Effective nutrition labeling of restaurant menu and pricing of healthy menu. Journal of Foodservice. 2008, 19:270-276. 10.1111/j.1748-0159.2008.00108.x
25. Kushi LH, Thun MJ, Mccullough M, et al.: American Cancer Society guidelines on nutrition and physical activity for cancer prevention: Reducing the risk of cancer with healthy food choices and physical activity. CA: A Cancer Journal for Clinicians. 2006, 56:254-281. 10.3322/canjclin.56.5.254
26. Krause M: Food, nutrition and diet therapy. Diabetes. 1958, 7:499. 10.2337/diab.7.6.499
27. Pouladzadeh P, Shirmohammadi S: Mobile multi-food recognition using deep learning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM). 2017, 13:1-21. 10.1145/3063592
28. Yanai K, Kawano Y: Food image recognition using deep convolutional network with pre-training and fine-tuning. 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy. 2015, 1-6. 10.1109/ICMEW.2015.7169816
29. Pandey P, Deepthi A, Mandal B, Puhan NB: FoodNet: Recognizing foods using ensemble of deep networks. IEEE Signal Processing Letters. 2017, 24:1758-1762. 10.1109/lsp.2017.2758862
30. Liu C, Cao Y, Luo Y, Chen G, Vokkarane V, Yunsheng M, Hou P: A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. IEEE Transactions on Services Computing. 2018, 11:249-261. 10.1109/tsc.2017.2662008
31. Bossard L, Guillaumin M, Van Gool L: Food-101 - Mining discriminative components with random forests. Computer Vision - ECCV 2014. Fleet D, Pajdla T, Schiele B, Tuytelaars T (ed): Springer, Cham; 2014. 8694:446-461. 10.1007/978-3-319-10599-4_29