Reviews in Endocrine and Metabolic Disorders (2023) 24:633–653
https://doi.org/10.1007/s11154-023-09795-4

Deep neural network for food image classification and nutrient identification: A systematic review

Rajdeep Kaur[1] · **Rakesh Kumar**[1] · **Meenu Gupta**[1]

Accepted: 7 March 2023 / Published online: 28 March 2023

Abstract

Technology impacts human life in both the aspects such as positive and negative, which helps in better communication and eliminating geographical boundaries. However, social media and mobile devices may lead to severe health conditions such as sleep problems, depression, obesity, etc. A systematic review is conducted to analyze health issues by tracking food intake by considering positive aspects using Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) Guidelines. The major scientific databases (such as Web of Science, Scopus, and IEEE explore) are explored to search the image recognition and analysis articles. The search query is applied to the databases using keywords like "Food Image," "Food Image Classification," "Nutrient Identification," "Nutrient Estimation," and using "Machine Learning," etc. 771 articles are extracted from these databases, and 56 are identified for final consideration after rigorous screening. A few investigations are extracted based on available food image datasets, hyperparameters tuning, a technique used, performance metrics, and challenges of Food Image Classification (FIC). This study discusses different investigations with their proposed FIC and nutrient estimation solution. Finally, this intensive research presents a case study using FIC and object detection techniques to estimate nutrition with food image analysis.

Keywords Nutrients · FIC · TL · DL · Pre-trained models · CNN · Fine tuning

Abbreviations

PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analysis
FIC	Food Image Classification
TL	Transfer Learning
DL	Deep Learning
CNN	Convolutional Neural Network
AF	Activation Function
RPN	Region Proposal Network
SVM	Support Vector Machines
VLCKD	Very-Low-Calorie Ketogenic Diet
SEAD	Southern Europe Atlantic Diet
UNIMIB2015	University of Milano-Bicocca 2015
UNIMIB2016	University of Milano-Bicocca 2016
FAMS	Food Annotation Management system
RUMTL	Regularized Uncertainty based Multi-Task Learning model
YOLO	You Only Look Once
Mask R-CNN	Mask Region-Convolutional Neural Network
Mask R-DSCNN	Mask Region-Depthwise Seperable Convolutional Neural Network

✉ Meenu Gupta
gupta.meenu5@gmail.com

Rajdeep Kaur
rajdeep.kaur6291@gmail.com

Rakesh Kumar
rakesh77kumar@gmail.com

[1] Department of Computer Science & Engineering, Chandigarh University, Punjab, India

1 Introduction

Food and lifestyle play an essential role in the emergence of chronic health disorders, including metabolic syndrome, obesity, cancer, hypertension, depression, and cardiovascular disease [1]. A healthy diet and proper nutrition are essential for the health of present and future generations. Further, most people live a sedentary lifestyle, significantly impacting the global population's health. Physical inactivity is the fourth most significant worldwide mortality risk factor, accounting for 6% of global death [2]. Diet, exercise, and behavioral interventions are the most effective approaches to weight loss [3]. Figure 1 shows the sedentary

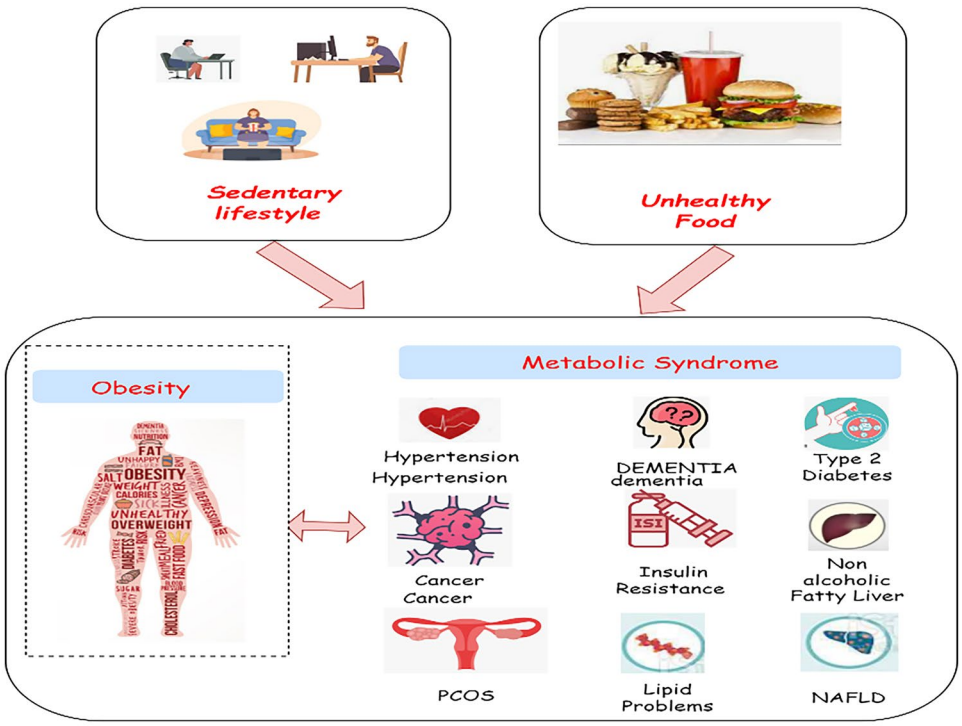

Fig. 1 Impact of Sedentary lifestyle and unhealthy diet on health

lifestyle, obesity, and unhealthy food linked to metabolic syndrome, which increases the risk of developing several other chronic diseases.

To live a healthy life and overcome obesity, everyone needs to maintain their daily food calorie requirement [4]. The intake of calories depends upon individual body requirements, as excess calorie consumption may turn into fat. Generally, Calories are broken down into two parts—macronutrients and micronutrients. Macronutrients consist of carbohydrates, protein, and fat, whereas micronutrients comprise water, vitamins, and minerals. A lack of macronutrients and micronutrients makes people more likely to get bad health and may cause various disorders [5]. There are a variety of diets, low carbohydrate, high protein, keto, etc., to manage overweight and obesity. VLCKD is one of the diets that gives promising results in managing body weight and improving hypertension, dyslipidemia, and hypertension [6]. SEAD is a diet that is consumed explicitly in the northwestern of the Iberian Peninsula. This type of diet is available based on different environmental factors, such as fresh and local seasonal products. This diet is rich in cereals, fruits, dairy products, vegetables, high intake of fish and seafood [7]. With the significant increase in health disorders due to obesity and diabetes in adults and children, appropriate dietary tracking is required daily [8]. Initially, the food consumed by an individual was monitored manually, and it was prone to inaccuracy due to delayed reporting to the practitioner. Technology advancements allow for more precise and user-friendly solutions, such as taking an image of the food with a smartphone and having the meal's calorie content computed.

In the present era, food tracking is simple due to modern devices such as cell phones, tablet devices, webcams, and notebooks. These devices are used to capture and record food images which can further be utilized to track the food consumed by individuals using classification techniques. Food classification is a challenging problem because of the complex structure of food images. Conventionally, image analysis was done using complex image feature extraction or handcrafted discriminant features. However, the recent development of DL algorithms helps provide better predictions to improve image analysis than handcrafted engineering [9]. But DL algorithms such as CNN need a large amount of data, which leads to the data scarcity problem. This data deficiency problem can be resolved using TL techniques.

The primary aim of TL is to use previously learned models for addressing new problems. CNN can yield enhanced recognition outcomes by using TL and feature extraction methods. This study intended to investigate TL with pretrained CNN models (VGG, ResNet, EfficientNet B0-B7, etc.) for food image classification. Further, object detection is another significant part of computer vision. Various image detention models (R-CNN, YOLO series, etc.) are available to locate the objects (foods) in the images. Different food image datasets, such as FOOD-101, FOOD-100, FOOD-256, etc., are available to train the models mentioned above. These food classification and detention models may help practitioners to track the nutrients and calories consumed

by individuals. In addition, this process can be helpful to the practitioner for the recommendation of a balanced nutrient diet to the individuals. It helps to overcome health-related disorders owing to the deficiency of nutrients. Figure 2 depicts the terms, keywords, and phrases essential for FIC and nutrient estimation.

This work is further classified into different sections. Section 2 presents the review process used to extract the relevant articles from different databases as per PRISMA Guidelines. Section 3 presents the previous work done by the researcher in FIC with different techniques. Section 4 discusses the proposed solution to the investigations. Further, a case study based on food image analysis using the EfficientNetB4 model is discussed in Sect. 5. Finally, this study is concluded in Sect. 7 with its future aspects.

2 Review process

In this section, PRISMA guidelines are used to perform the review process. The research articles relevant to the problem statement (FIC and nutrient identification) are considered in this study. All the phases of the systematic review for FIC are presented in Fig. 3.

In this section, the articles (i.e., 771) are collected from the three bibliographic databases, Scopus, IEEE Explorer, and Web of Science, using the keywords mentioned in Table 1. In this process, the articles published in 2016 and later are considered. In the screening phase, the 771 articles are reviewed manually, and 146 are excluded based on duplicate records, not in English language, pure review analysis, and not relevant to the problem statement. Further, the left articles (771–146 = 625) are reviewed with abstract and title, and 468 are excluded due to not being related to the problem statement. In the next eligibility phase, the rest of the articles (625–468 = 157) full text is accessed, and removed 115 articles that are not related due to missing primary data and not related to the scope. After completing the process, only 42 (257–115) articles are considered relevant to the FIC and nutrient identification. In addition to the above-mentioned research, more articles from other general keywords (related to the food image datasets and pre-trained models) are considered for review. Finally, 56 research articles are included in this systematic review process. The year-wise analysis of selected articles is shown in Fig. 4.

3 Background study

This section discusses the research on different AI-based approaches proposed by researchers for identifying food nutrients using food images. The review procedure consists of two phases. The first phase presents the comprehensive review analysis of the papers in which the CNN model is

Fig. 2 Word cloud image for FIC

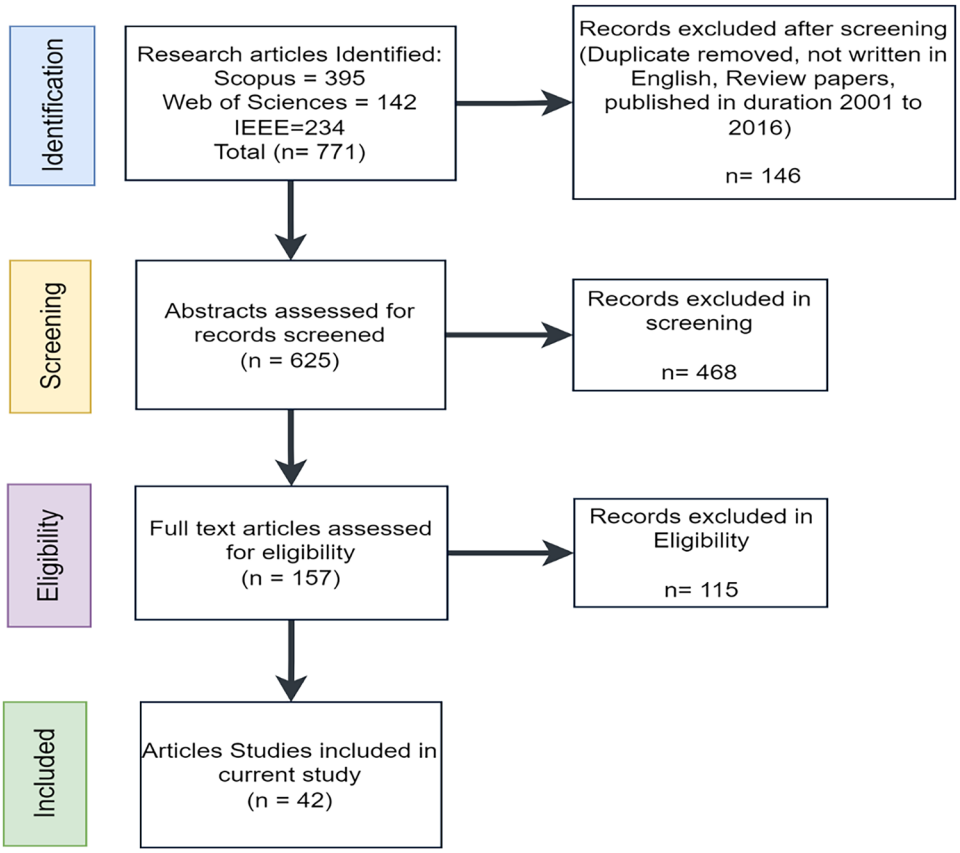

Fig. 3 The systematic review process for FIC & nutrient identification using PRISMA guidelines

designed from scratch, and the second phase discusses the use of the TL approach in the pre-trained CNN model.

3.1 Food image analysis using CNN and object detection models

The availability and processing of vast data have accelerated CNN research by applying different architectural innovations (setting different hyperparameters such as Activation Function (AF), loss functions**,** etc.). This section presents a comprehensive review of the articles designed CNN model (created from scratch) for identifying the food class and attributes of food from the food image.

Wei and Wang [10] proposed a model which uses faster R-CNN architecture for food area detection in the food image and then applied CNN to extract the features of the detected food image area. Experiments were conducted on the food image dataset (Dish-233), which consisted of 233 dishes and 49,168 images. R-CNN was fine-tuned, and the performance was 5% better than other models. Situju et al. [11] discussed an automated technique for estimating food ingredients from food images using multitask CNN. It estimated food calorie content by classifying the food image. A multitask CNN was trained using many food category images, and two-stage TL was applied to enhance ingredient estimation. It experimentally determined the presence of a correlation between food type and salinity content. Xiao et al. [12] proposed a CNN-based food image recognition algorithm. They combined a jumping convolution layer and a traditional convolutional layer to minimize the calculation parameters. Compared to the experimental outcomes of previous DL networks, the suggested method has a beneficial

Table 1 Data collection query and number of articles from different databases

Database	Query	Number of articles (771)
Scopus	("Food Image" OR "Food Image Classification" OR "Nutrient Identification" OR "Nutrient Estimation") AND ("Deep Learning" OR "Machine Learning" OR "Artificial Intelligence" OR "Learning" OR "CNN" "Convolutional Neural Network")	395
Web of Science		234
IEEE Xplore		142

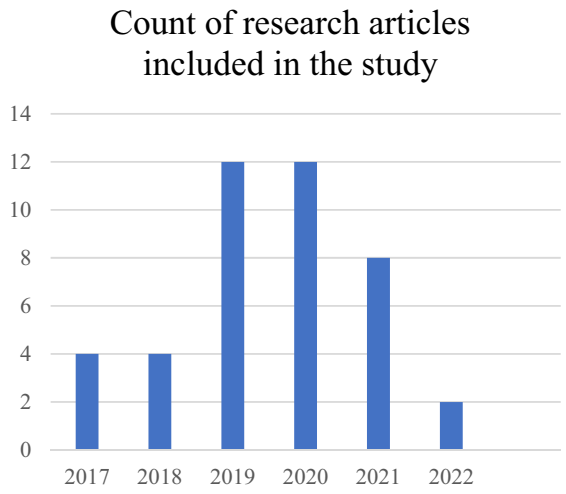

Fig. 4 Year-wise Articles selected for the review

impact that can recognize food rapidly and reduce training time. Jia et al. [13] proposed a framework to detect food images using AI. Large sets of egocentric images were collected from free-living persons using the eButton device. The images in the datasets were classified as food or non-food using the CNN approach. Finally, it was concluded that the AI technique could automatically distinguish foods accurately, minimizing both the data processing burden and privacy concerns.

Jiang et al. [14] proposed a three-step algorithm for recognizing food images based on deep CNN and candidate region. They applied Region Proposal Network (RPN) to generate multiple regions using input images. They also classified each region of the proposals into distinct food groups and mapped each region onto a feature map. Finally, a dietary evaluation report was created based on the amounts of fat, protein, total calories, and carbohydrates. Sahoo et al. [15] used an image recognition algorithm named FoodAI that utilized 400,000 food images from the dataset (756 food classes). FoodAI was trained with previously trained models. FoodAI has been developed as an API service and was one of the components that enabled Singapore's Health Promotion Board's mobile application. It was a simplified food logging model promoting healthy eating habits. Mezgec and Koroušić [16] proposed a revolutionary architecture (deep CNN (NutriNet)) for identifying and recognizing food and beverage images. This architecture was trained on a dataset of foods and drinks images (225,953 images of 520 classes) and achieved an accuracy of 86.72%. The architecture was also tested on self-acquired image captures using a smartphone from Parkinson's disease patients and obtained 55% accuracy. A smartphone application used the model to assess the meals of Parkinson's disease patients.

Min et al. [17] introduced a food image dataset ISIA Food-500, which contains 399726 food images. In addition, the authors proposed a global–local attention network with two sub-networks for FIC. The first subnetwork extracted additional discriminative features, and the second generated the attention regions. Further, the features from these two subnetworks fused for the final FIC. Lohala et al. [18] proposed a DCNN algorithm to improve the prediction accuracy of food images. The proposed model includes a modified loss function, feature extraction by switching between convolutional layers, pooling layers, and a fully connected layer on top of all layers. The assessment metric for algorithm speed was total execution time, while the evaluation metric for algorithm correctness was probability score. The classification accuracy of the fast-food images increased by 5%, and processing time was reduced by 40 to 50 s. Subhi and Ali [19] authors developed a CNN model for detecting and identifying food images. They introduced a new dataset for local Malaysian cuisines consisting of 11 food groups with 5,800 food images. Food-101 (food/non-food categorization) and local Malaysian (food item categorization) datasets were used to analyze the performance of the model. In addition, deep CNN, i.e., 24 weight layers added to the model for FIC. Aguilar et al. [20] proposed a new multitasking model capable of simultaneously predicting multiple food-related tasks, such as cuisine, dish, and food categories. The proposed model includes homoscedastic uncertainty modeling to accommodate both single-label and multi-label classification. In addition, a new multi-attribute food dataset was proposed to measure performance metrics such as accuracy, recall, etc.

Jubayer et al. [21] proposed a study to detect different types of molds that grow on the surfaces of various foods. The authors conducted a case study for the detection of mold detection on food surfaces using the YOLOv5 model. A collection of 2050 food images with mold growing on their surfaces were considered to train the model. The YOLOv5 model outperformed in terms of precision, recall, and average precision than YOLOv3 and YOLOv4. Li et al. [22] proposed a Chinese food image dataset CF-108 to train the Mask R-CNN model. Further, Mask R-DSCNN was introduced by the authors to lower the expensive computation cost. The proposed architecture greatly reduces resource consumption and increases the detection efficiency of Chinese food images. Son et al. [23] developed a DL-based algorithm for detecting foreign objects. It suggested a technique for effectively acquiring a DL training dataset for application in food quality evaluation and production processes. U-net architecture is applied to predict binary classification pixel-by-pixel. The F1-score of the model trained on the synthetic dataset of almonds at 360 lux light intensity was 0.82.

3.2 Role of TL in the analysis of food images

Creating a CNN model from scratch is a challenging task as well as it requires lots of time and effort to combine its layer with a hyperparameter. In this context, TL is an optimal solution that utilizes previously trained models to address new issues. The section presents an overview of the published research articles on image processing using TL.

Tasci [24] presented a food image recognition approach using DL models such as ResNet, GoogleNet, VGGNet, and InceptionV3. Six voting combination rules were applied, such as the product of probabilities, minimum, median, maximum, and an average of probabilities weighted were applied as voting combination rules for ensemble methods. The food image datasets were used to train the models such as FOOD-101, UEC-FOOD100, and UEC-FOOD256. The proposed ensemble voting system with TL provided promising outcomes compared to other state-of-the-art methods. Ma et al. [25] published a dataset named ChinaMartFood-109, the first food image database from the Chinese market. It has 10,921 food images and 23 nutritional components of 18 major dietary groups. In the result analysis, the accuracy achieved from the model formulation was 78% using the enhanced Inception V3 model. Hu et al. [26] proposed a system in which DenseNet121, Xception, and ResNet50 DL models were trained on the food images datasets such as FOOD-101, UEC-FOOD256, ChineseFoodNet, and UEC-FOOD100. The authors claimed that the top-1 accuracy achieved using the proposed models was 75.9%, 84.0%, 79.0%, and 84.5%, respectively. Further, an automatic prototype system was developed for diet recommendations implemented in Tongji University's canteen, and their results were outperformed. Qiu et al. [27] proposed an approach to predict an individual's food intake with food eating sequence. A TL strategy was developed and trained on 4200 food images to determine the foods consumed by the individual. The authors concluded that the proposed method for evaluating each person's diet was reliable. Metwalli et al. [28] proposed a model (named as DenseFood) based on a densely connected CNN architecture with multiple layers. SoftMax and center loss were used during the training phase to decrease variation within the same category and maximize variation between different categories. The different models (ResNet50, DenseFood, and DenseNet121) were trained using the VIREO-172 dataset to compare their performance. The result analysis revealed that the proposed DenseFood model has an accuracy of 81.23 percent. Hassan et al. [29] used the pre-trained model Inception V3 to label the Yelp dataset. The inception V3 model was trained using the Food-101 dataset. Wu et al. [30] designed an AI visual checkout system (consisting Kinect camera, a desktop computer comprising the prototype, and a light source) for a Bento buffet (consisted 22 distinct meal items) to reduce customer wait time. The CNN architectures (i.e., AlexNet, DenseNet, VGG, ResNet, and Inception v4) were used for food recognition with the collected dataset. The inception v4 model achieved 99.11% average validation accuracy on food identification, although the maximum training and recognition time was very high. AlexNet model achieved 94.5% accuracy with minimal training and recognition time.

Yunus et al. [31] proposed a mobile-based application that took images of the meal and predicted the nutritional value with food availability. They have applied the Inception DL model for the identification of food accurately. 85% accuracy was achieved with the proposed model to recognize the food items. Farooq and Sazanov [32] designed a model to track the nutrients consumed by the individual with analysis of food images (collected using wearable devices). The proposed method was used to classify food images from the Pittsburgh dataset using CNN and linear SVM models. AlexNet was utilized to extract features from food images. The classification task was divided into two further sub-tasks. In the first task, images were classified into 61 classes with an accuracy of 70.1%, and in the second task, images were classified into seven classes with an accuracy of 94.01%. Ruenin et al. [33] proposed a model to estimate the number of calories consumed by the elderly in the hospital. It consists of three major components: food weight estimation, food detection and classification, and an application for calorie prediction. A ResNet-50 and R-CNN models were trained on the SH-FID dataset (Suandok Hospital Food Photos Dataset), containing 16,067 food image samples classified into 39 classes. Further, they developed a web application to display the calories consumed by the elderly and computed food weight into calories using the hospital reference table. Memis et al. [34] compared the performance of several DL approaches for classifying food images. Various pre-trained models (ResNet-50, Inception-V3, Densenet-121, Resnet-18, ResNext-50, and Wide Resnet-50) were trained on the food image dataset (UEC Food-100). ResNet-50 yielded the most accurate classification result, with an accuracy of 87.7%.

Xu et al. [35] developed a model by combining pre-trained models (Mobile NetV2, VGG16, and ResNet50) with Convolutional Block Attention Module (CBAM) to improve the accuracy of Asian FIC. In addition, the mixed data augmentation technique (Mixup) was used to have a more refined capacity for classification. The proposed model achieved the Top-1 accuracy of 87.33%. Özsert Yiğit and Özyildirim [36] proposed a deep CNN architecture that was trained and compared its performance with other pre-trained models (such as Caffenet and Alexnet). They used Food11 and Food101 datasets for model formulation. As a result, it was concluded that the pre-trained models performed better, although the performance of the proposed model was adequate.

3.3 Proposed investigations

This study explores food image analysis and other machine-learning approaches employed for nutrient identification. The following investigations useful for assessing the research study are discussed below.

Investigation 1. What are the various food image datasets available for FIC?

Investigation 2. What are the different hyperparameters required to enhance the performance of a model?

Investigation 3. How CNN architecture delivers a solution to the FIC problem?

Investigation 4. How fine tune and feature extraction strategies of TL useful in FIC?

Investigation 5. What metrics are required to measure the performance of a model?

Investigation 6. What are the challenges faced in classifying food images?

The above investigations are retrieved after rigorous analysis of the past work done by researchers in the field of FIC (discussed in Sect. 3). This research methodology is proposed based on these investigations, as discussed in Sect. 4.

4 Research methodology

Classification of food images is challenging when captured in different perspectives and patterns, as many foods have similar colors and shapes. CNN is one of the most extensively used DL techniques for FIC and performs better than traditional image classification methods. A food image dataset is first required to prepare and pre-process for training the CNN model for FIC. This dataset needs to be divided into further sub-datasets for training and testing. Then, different augmentations techniques [37], such as flip, rotation, vertical, horizontal, height shift, width shift, etc., can be applied to the food image dataset to avoid the overfitting problem.

There are two approaches to developing a CNN model. In the first approach, a CNN model can be designed from scratch, passing an image through the different layers (pooling layers, convolutional layers, and fully connected layers) to get the desired result. In the second approach, TL is used where pre-trained models are trained on the newly collected dataset [38]. These pre-trained models are either used for feature extraction or fine-tuned by unfreezing the top layers to classify the images. Further, the model is trained with a training dataset, and the model's performance is evaluated using a test dataset to make the final predictions. Several performance metrics (Accuracy, Recall, Precision, F-1 score, etc.) are available to measure the performance of the CNN model. An FIC model can be used to classify the type of food in the image and then to predict the number of macronutrients (Fat, Protein, Fiber, and Carbohydrate) in the food. Figure 5 presents the FIC process using CNN.

5 Discussion

Based on the proposed research methodology, solutions proposed for retrieved investigations as discussed in Sect. 3.

***Investigation 1.* What are the various food image datasets available for FIC?**

PS The performance of feature extraction and classification approaches depends on the datasets. Several food image datasets, such as FOOD-101, UEC-FOOD-256, UEC-FOOD100, etc., are eventually used as benchmarks to compare new classifiers' performance and current food recognition techniques. Table 2 summarizes some food image datasets with the number of food classes and images and references to the researchers who used them in their work for FIC. Figure 6 represents the sample food images taken from the food image datasets.

***Investigation 2*. What are the different hyperparameters required to enhance the performance of a model?**

PS The hyperparameters are the variables of the CNN that are initialized to optimize the network. Different datasets need unique sets of hyperparameters for successful prediction. There is no definitive solution to the question of which hyperparameters (Learning Rate (LR), number of neurons, optimizer, AF, batch size, epochs, etc.) are optimal for a given dataset. It is essential to tune hyperparameters to determine the optimal collection for a given dataset. Table 3 provides a brief overview of the CNN hyperparameters. There are several approaches, such as manual search, Grid search, Random search, and Bayesian optimization, to fine-tune the hyperparameters of the CNN model, as shown in Fig. 7. Shekar and Dagnew [53] applied a grid search approach to optimize the hyperparameters of the Random Forest (RF) tree to classify the cancer disease. This hyperparameter tuning method provides the optimal parameters for maximum accuracy and minimum error. Table 4 presents the detail of the CNN architectures and object detection models used in literature with the hyperparameters settings.

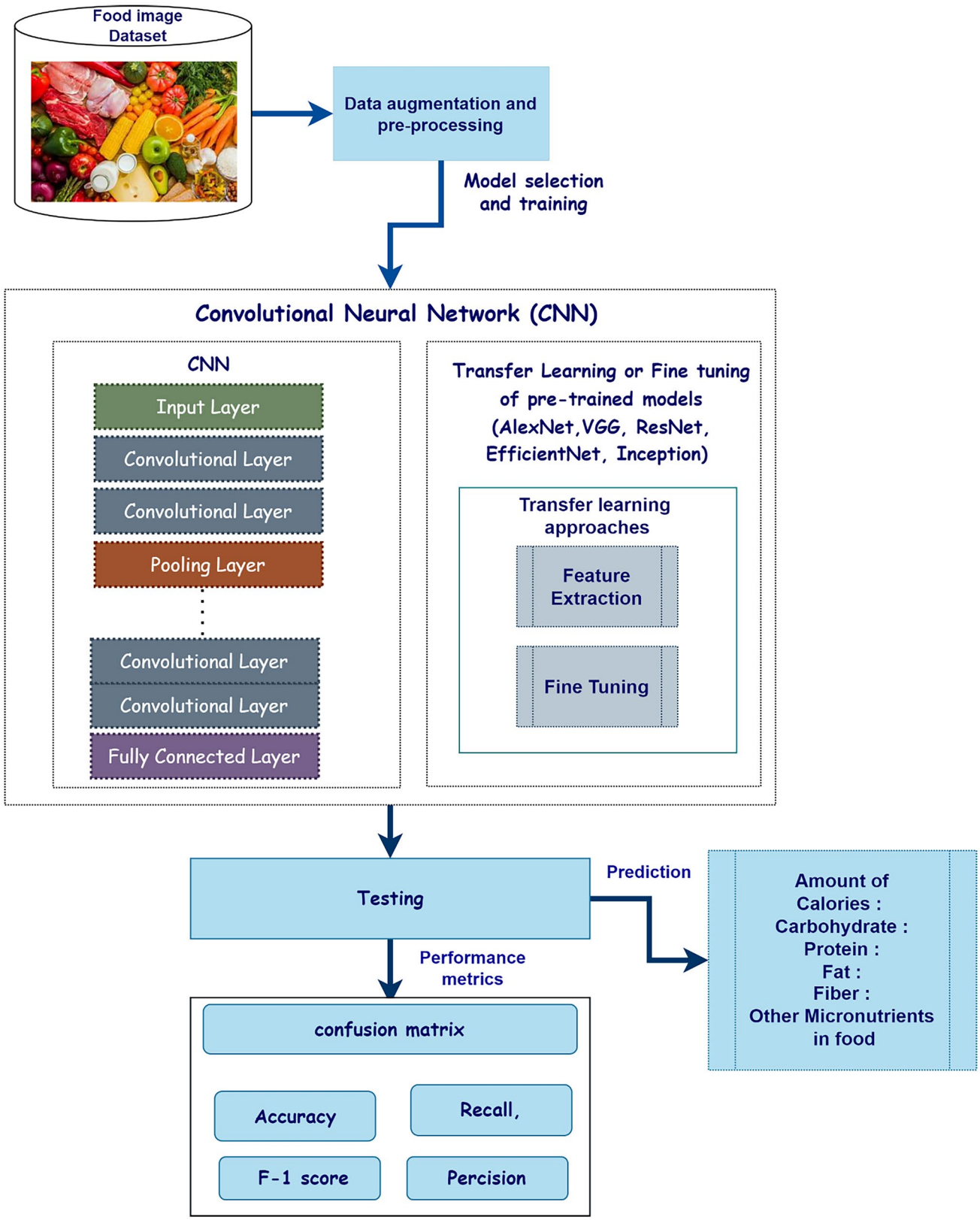

Fig. 5 FIC and nutrient identification work process

Table 2 Summary of Food image datasets with the number of food classes

Food Image Dataset	Year	Number of food classes	Number of images	Articles used the food dataset for classification	Source/Link
UEC-FOOD100	2012	100 food categories	More than 100 images per class	[24, 26, 34]	[39] UEC-Food100 | Kaggle
FOOD-101	2014	101 food categories	10100 food images (1000 images per food class)	[14, 18, 24, 26, 28, 36, 37, 40–42,]	[43] Food Images (Food-101) | Kaggle
UEC-FOOD256	2014	256 food categories	More than 100 images per class	[14, 24, 26, 28, 41]	[44] UEC-Food256 | Kaggle
VIREO-FOOD	2016	172 food classes	110,241 food images	[28, 41]	[45] Vireo-Food 172 dataset (cityu.edu.hk)
UNIMIB2015	2015	15 food categories	2,000 food images	[46]	[47] Food Recognition | Imaging and Vision Laboratory (unimib.it)
UNIMIB2016	2016	73 food classes	3616 food images	[8]	[48] Food Recognition | Imaging and Vision Laboratory (unimib.it)
ChineseFoodNet	2017	208 categories	865 images per category	[26]	[49] https://sites.google.com/view/chinesefoodnet/
FoodX-251	2019	251 food classes	158K images	[50]	[50] iFood - 2019 at FGVC6 | Kaggle
ChinaFood-100	2021	100 food categories	10,074 images	[51]	[51]
Vireo Food-251	2021	251 food classes	169673 images	[52]	[52] Vireo-Food 251 dataset (cityu.edu.hk)
ChinaMartFood-109	2022	109 food classes (clustered in 18 groups and 23 nutrients for each class)	100 images per class	[25]	[25]
Dish-233	2020	233 food image classes	49,168 food images	[10]	[10]
ISIA Food-500	2020	500 food image categories	399,726 food images	[17]	[17] Food Computing:: Home

Investigation 3. How CNN architecture delivers a solution to the FIC problem?

PS Nowadays, proper nutrition assessment and analysis techniques give people more options to learn about their daily eating habits, find nutrition patterns, and maintain a healthy diet. Food image recognition is very challenging and complex, as many foods' shapes, sizes, and colors are the same [54]. CNN is one of the most powerful AI techniques for solving the problem of computer vision and extracting complex and unique features from the image [55]. Figure 8 shows the basic building blocks of CNN for image analysis in which these layers are stacked in a specific order to get improved results.

CNN architectures like AlexNet, ResNet, Inception, VGGNet, EfficentNet, and GoogleNet models are developed for image recognition applications. The researchers fine-tuned these pre-trained models according to their problem statement to get the desired output. AlexNet is the architecture for DL that promoted CNN. It is the first significant CNN architecture developed to win the ImageNet Large Scale Visual Recognition Challenge in 2012. It comprises 5 convolutional layers with max-pooling layers, three fully connected, and two dropout layers. Finally, a SoftMax function is used in the output layer [56]. ResNet is a residual network and pre-trained CNN that trains deeper networks more quickly than conventional CNNs. These networks may perform well with substantially more in-depth and are simpler to tune. There are 101 layers in ResNet101. The network's input dimensions are 224 by 224 [57]. Simonyan and Zisserman [58] proposed a deep CNN architecture, i.e., VGGNet It has two versions, namely VGG16(which consists of 16 layers) and VGG19(which consists of 19 layers), and the input image size is 224 by 224. Three fully linked layers and one SoftMax layer were added to the convolutional layer stack. It has been trained on millions of images from ImageNet LSVRC-2014 to classify images into 1000 item categories. This network's setup utilizes small convolution filters, such as three-by-three. The EfficientNet is a CNN architecture that employs an efficient composite coefficient to equally scale the depth, breadth, and resolution of the model to dramatically enhance the model's accuracy. This mechanism helps to improve the accuracy as the size of the image increases, neural network requires a greater number of layers to capture more fine-grained patterns from large-size images [59]. Earlier to EfficientNets, ConvNets were

Fig. 6 Sample food images

typically scaled by one of three dimensions: depth (number of layers), width (number of channels), or picture resolution (image size). On the other hand, the EfficientNet model scale all three dimensions of the network uniformly. GoogleNet has been trained on millions of images from the ImageNet Challenge 2014 and categorized them into 1000 classes. This network, which is 22 layers thick, is designed to maintain a consistent computing load as its depth and width increase. The network's input dimensions are 224 by 224 [60]. Different CNN architectures and object detection models used by the researchers with hyperparameter values are discussed in Table 4.

Investigation 4. **How fine tune and feature extraction strategies of TL useful in FIC?**

PS Two TL strategies, Feature extraction and Fine Tuning are used to train the pre-trained models with the new dataset. Feature extraction is applied for the small datasets where all the layers are frozen with the same features, and added new classifier on the top layer of the model. Fine Tuning is applied to train the pre-trained models with that large dataset. In this concept, the pre-trained model involves training some layers instead of freezing all the layers [65]. The feature learning of layers varies in fine-tuning approach. The bottom layers

Table 3 Overview of the CNN hyperparameters

Hyper-parameters	Description
AF	AF introduced non-linearity in the output of the neuron—for example, ReLU, Softmax, Sigmoid, Tanh, and Leaky ReLU
Total layers	The number of layers is added in between the input and output layer until the test error not improves
Epochs	The number of epochs determines the number of times a network weight is updated
Batch Size	It defines the sub-samples to train the model with sizes such as 32, 64, 128, etc
LR	The LR determines how frequently the optimization algorithm updates the weight—for example, Adam, RMSProp, SGD, Adagrad, and AdaDelta
Dropout	Dropout is a method of regularization used to prevent overfitting. It is used between 20 to 50 percent of neurons depending upon the type of the problem

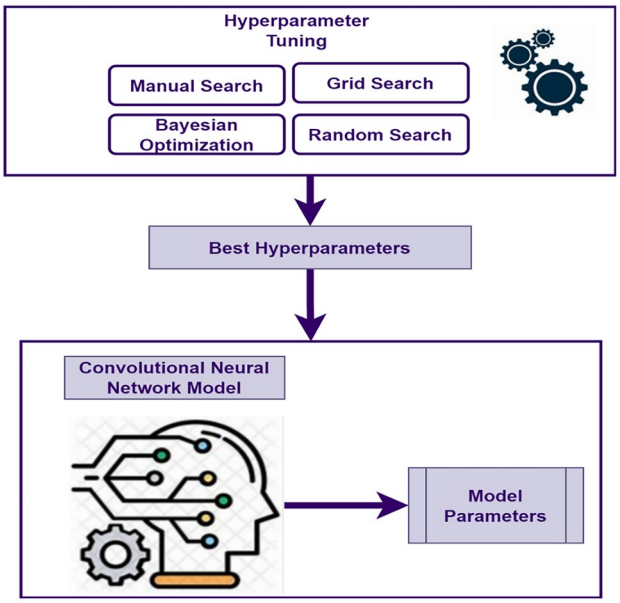

Fig. 7 Tuning of Hyperparameter of CNN model to improve the system performance

of a CNN learn low-level features (edges, texture, etc.), and the top layers learn the features more specific to the dataset (more detailed textures and patterns). Furthermore, freezing and training of the layers are manual tasks that are problematic in optimizing the networks with hundreds or thousands of layers. Figure 9 depicts fine-tuning the pre-trained models, replacing the top layers with the modified ones.

Investigation 5. **What metrics are required to measure the performance of a model?**

PS Evaluation metrics are key to assessing whether food recognition models are appropriate. In the literature, the researchers have discussed different metrics concerning the nature of the problem statement.

Confusion matrix

The Confusion Matrix is a common way to describe how well a classification model works in ML. Classification accuracy alone can sometimes be misleading when there are more than two classes in a dataset or when the number of observations in each class is unequal. The confusion matrix compares the predicted label and the actual label. The confusion matrix is represented using the terms TN (True Negative), TP (True False), FP (False Positive), and FN (False Negative) to calculate the performance metrics such as accuracy, precision, recall, and F1 score.

Accuracy

The accuracy rate is the assessment metric to measure how well the suggested ensemble classification algorithm performs. The ratio of the total number of true positives and true negatives over the total number of cases is used to measure classification accuracy shown in Eq. (1).

$$Accuracy = \frac{TN + TP}{TN + TP + FP + FN} \times 100 \quad (1)$$

Precision

Precision is defined as the ratio between true positive (TP) and all positives (TP + FP), as shown in Eq. (2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall

A recall measures correctly identifying the true positives, as shown in Eq. (3).

$$Recall = \frac{TP}{TP + FP} \quad (3)$$

Investigation 6. **What are the challenges faced in classifying food images?**

PS CNN works exceptionally well in classifying food images, although insufficient training samples may result in an overfitting problem. This issue may prevent the model from classifying the new data accurately. The overfitting problem can be resolved using data augmentation or adding multiple variants to the image during the training phase. Augmentation techniques (Rotation, cropping, padding, flipping, etc.) can be used to add more training samples [66]. Augmentation techniques are applied to the dataset to add more samples. The outcomes of the FIC task are significantly dependent upon the respective datasets. Existing datasets are insufficient and only include a small number of food characteristics, such as varied camera angles, lighting conditions, and backdrops [67]. Food image datasets include food groups or classes with food composition detail (macronutrient or micronutrient) required to develop food classification and nutrient identification using the DL model. Next, the FIC model's primary challenge is tuning the hyperparameters. In this, the values of the hyperparameters (LR, Batch size, Number of epochs, etc.) need to change, improving the model's overall performance. High hardware configuration (many GPUs with high RAM) is also required to train the model with large datasets.

6 Food image analysis and detection: a case study

Food identification is gaining more attention in various applications, including nutrition management and self-service restaurants. This section presents a case study on

Table 4 CNN architecture with detailed hyperparameters

Author	Year	Classification/Object detection techniques used	Experiment setup/Hyperparameters	Outcome
Wei and Wang [10], **2020**	2020	Faster R-CNN for food detection in the image and CNN for feature extraction	Initial learning rate = 0.001 (cnaged with iteration), Momentum = 0.9, weight attenuation = 0.0005, and the number of epochs was 60	The proposed model performed better than CNN-G, CNN-G-F, Faster-R-CNNG
Situju et al. [11]	2019	Xception Model	The LR was set as 0.001 (gradually decreased to 0.0001), dropout 0.5, Batch size 16, number of epochs 100, layers added in the architecture are: global average pooling layer, fully connected layer with dropout, and output layer	The accuracy of the proposed model (multi-task CNN) is better than single task CNN
Xiao et al. [12]	2021	Proposed a CNN-based model which compared with other pre-trained models (VGG16, AlexNet, GoogleNet)	Using a jumping convolution to reduce the computation parameters, the ReLu activation function used the global pooling layer instead of the fully connected layer	The proposed algorithm recognizes the food quickly and reduces calculation parameters, which also reduces the time
Jia et al. [13]	2018	CNN (clarifai CNN) to classify food/non-food images	Images from the two datasets (Food-5 K and eButton) were processed using clarifai CNN	The performance of the AI algorithm compared on the datasets and the performance on the dataset FOOD-5 K was better
Jiang et al. [14]	2020	R-CNN and deep CNN	Multiple regions of proposal generated using RPN (derived from a faster R-CNN model) and deep CNN for object classification. Dataset split is training: testing 80:20, momentum 0.9, weight decay rate = 0.0005	A new type of food dataset was generated using FOOD-101 with a bounding box
Sahoo et al. [15]	2019	FoodAI development, experiment study using ResNet-50, ResNet-101, ResNeXT-50, SENet with ResNeXT-50,	Used focal loss with all the models	A combination model (SENet with ResNeXT-50) achieved the best accuracy (top-1 80.86% and top-5 95.61%), and ResNet-101 did not perform well compared to other models
Mezgec and Koroušić [16]	2017	NutriNet and other CNN models (AlexNet, GoogLeNet, ResNet)	Used three solvers: SGD, NAG, and AdaGrad Hyperparameters AlexNet: batch size 256 with a learning rate of 0.02 NutriNet: batch size 128 with a learning rate of 0.01 GoogLeNet: batch size 64 with a learning rate of 0.005, ResNet: batch size 16 with a learning rate of 0.00125 All the models were trained using 150 epochs	ResNet with solver NAG achieved the best classification accuracy at 87.96%
Min et al. [17]	2020	Pre-trained models (VGG-16, GoogLeNet, ResNet-152, DenseNet-161, etc.) trained on food image dataset ISIA Food-500	Optimizer = SGD, batch size = 80, momentum = 0.9, learning rate 10^{-2}	The proposed work presented a large-scale food image dataset ISIA Food-500 and compared it with other food image datasets
Lohala et al. [18]	2021	Deep CNN	Softmax function, Modified loss function	The presented scheme has enhanced accuracy by 5.1% and processing time from 0.03 to 0.105 s
Subhi and Ali [19]	2018	Deep CNN	Number of layers 24 (deep CNN includes 21 convolutional layers and 3 fully connected layers), used ReLu activation function	A new food image dataset with 11 food categories and 3300 food images is proposed, which includes Malaysian food images
Aguilar et al. [20]	2019	Proposed RUMTL architecture based on ResNet-50	Softmax activation function for a single-Label task, sigmoid for a multi-label task, categorical cross-entropy loss function, batch size 20, learning rate 2e-4, decay 0.2 for every 8 epochs	Proposed model for single label and multi-label classification and presented a new food image dataset with multi-attribute food

Table 4 (continued)

Author	Year	Classification/Object detection techniques used	Experiment setup/Hyperparameters	Outcome
Jubayer et al. [21]	2021	YOLOv3, YOLOv4, YOLOv5	Image size = 640, Batch size = 10, number of epochs = 250, YOLO model = YOLOv5s.yaml	YOLOv5 performed better than the other YOLO series models (YOLOv3 and YOLOv4) in detecting mold on the food surfaces
Li et al. [22]	2020	Mask R-CNN	Optimizer = SGD, learning rate = 0.001, momentum = 0.9, batch size = 128, epochs = 200000	Proposed a new dataset of Chinese food dishes and applied mask R-CNN architecture to reduce the time consumption
Son et al. [23]	2021	U-Net	ReLu activation function, cross-entropy loss function, Optimizer = SGD, learning rate 0.001, momentum = 0.9, batch size 5, weight decay 0.0005, no of epochs = 50	The proposed approach focused on raw materials of food and background detection to detect foreign objects
Tasci [24]**, 2020**	2020	voting-based ensemble approach used a combination of different CNN architectures (VGG16, VGG19, GoogleNet, ResNet101, InceptionV3)	Stochastic Gradient Descent with Momentum (SGDM) optimizer, Batch size = 64, LR (10^{-3}) and momentum coefficient (0.9), epochs (50 for ResNet101, InceptionV3 and 20 for VGG16, VGG19, GoogleNet) and other parameters are set as default value	Voting based ensemble approach gives the highest accuracy (84.28%)
Ma et al. [25]	2022	VGG, ResNet, Inception, WISeR	Default parameters	Inception V3 outperformed (78% Top-1 accuracy and 94% top-5 accuracy)
Hu et al. [26]	2019	DenseNet121, Xception, ResNet50	Adam Optimizer was used for the first five epochs. After five epochs, the SGD optimizer (0.9 momentum and initial LR was 0.005)	Deep learning models are compared on different Food datasets, and the highest achieved accuracy is 84.5% on FOOD-101
Metwalli et al. [28]	2019	DenseNet121, ResNet50, DenseFood (Model trained from scratch)	The initial learning rate was 0.01. LR changed to 0.001, and used cosine decay to reduce the LR. The loss function used as a combination of the center loss function and SoftMax categorical cross entropy	The highest accuracy was 81.23% (DenseFood model)
Won [41]	2020	ResNet50	SGD optimizer, Batch size = 16, initial LR 0.001, dropout 0.5,	The highest accuracy was 91.34%
Chen et al. [52]	2021	VGG	LR 0.001, Batch size 64,	Two scale CNN based on ResNet50 achieved better accuracy on the food image datasets (FOOD-101, UEC FOOD-256, Vireo Food-172)
Kumari and Signh [61]	2019	ResNet50, LSTM architecture	Adam optimizer, different loss functions evaluated, and cross-entropy perform better	Achieved approximately 70% of compression (High cosine)
Islam et al. [62]	2019	AlexNet, GoogleNet, ResNet-50	-	The highest accuracy was 61.3% (GoogleNet)
Ege and Yanai [63]	2018	VGG-16	SGD optimizer with momentum 0.9, Batch size (mini-batch) 8, used batch normalization, compared different loss functions (L_{re}, L_{ab}, $L_{re} + L_{ab}$)	Multi-task CNN improved the correlation coefficient by 0.039
Tai et al. [64]	2020	Modified EfficientNetB0	Swish activation function, Used dropout layer	Achieved an accuracy of 92.33%

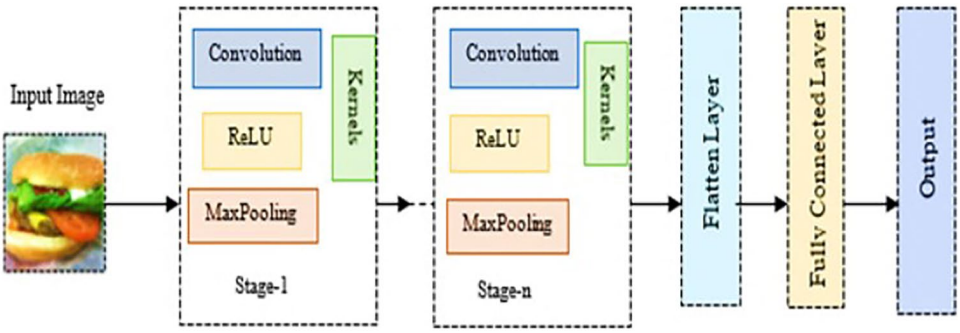

Fig. 8 CNN's basic building blocks

food image analysis using EfficientNet B0-B6 and food item detection using YOLOv5.

6.1 Food image analysis using EfficientNet B0-B6

This work presents a comprehensive review of the research articles which develop models for classifying food images. A case study is conducted using previous research to classify the food images. A dataset with 14 food classes (1000 images per class) derived from the FOOD-101 is considered to train the EfficientNet B0-B6 family models, as shown in Table 5. Further, food images of all the food classes are divided into 80:20 data ratio of the dataset, 80% training, and 20% testing data, respectively.

We fine-tuned the pre-trained CNN models EfficientNet B0-B6 by adjusting the number of 14 output classes in the final fully connected layer to match the number of categories in the corresponding dataset. The hyperparameters values of the model are shown in Table 6.

For the first 10 epochs, all the layers are frozen. After this, for the following ten epochs (after 10 epochs), the top 5 layers are unfrozen to extract the specific features. Then, for the subsequent five layers (from 16 epochs onwards), the following top 5 layers are unfrozen, and this process is repeated. The EfficientNet models B0, B1, B2, B3, B4, B5, B6 achieved the accuracy 87.7%, 88.1%, 88.9%, 88.49%, 89.57%, 89.31%, 89.28% respectively. The performance of all the variants of the EfficientNet model is similar, but the EfficientNetB4 model achieved the highest accuracy of 89.57%. Further, the classification results are evaluated using the confusion matrix shown in Fig. 10. Each cuisine category has 200 test images. However, the food classes (apple_pie and bread_pudding food) are most misclassified, as shown in the confusion matrix. The food classes apple_pie and bread_pudding food have similar shapes in images. Due to this,

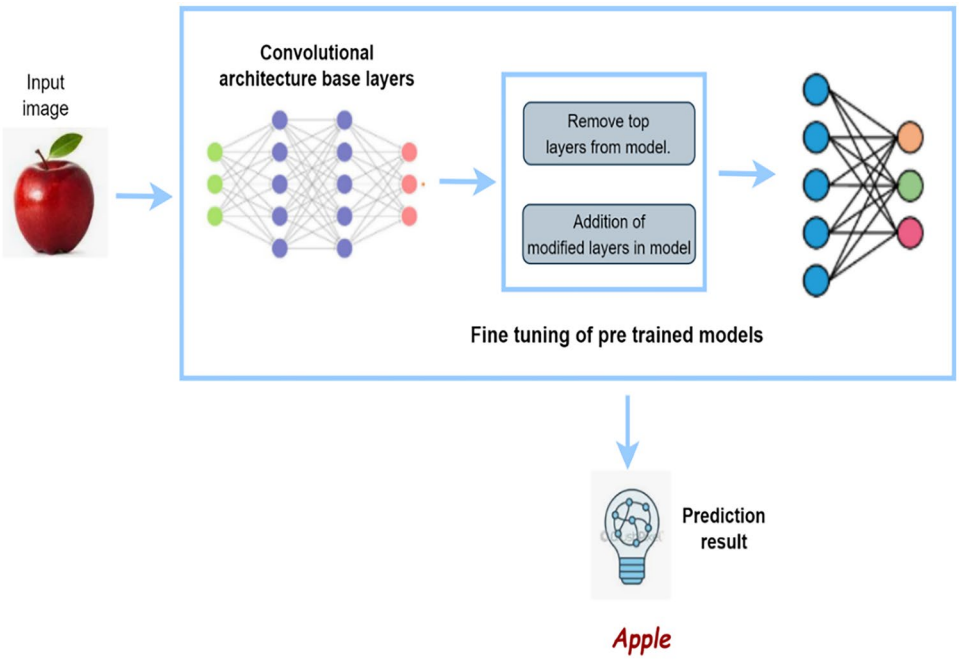

Fig. 9 TL to fine-tune pre-trained models

Table 5 Food items considered for the training

Sno	Food_item	Sno	Food_item
1	apple_pie	8	french_onion_soup
2	bread_pudding	9	french_toast
3	breakfast_burrito	10	fried_rice
4	chicken_wings	11	grilled_cheese_sandwich
5	eggs_benedict	12	pizza
6	fish_and_chips	13	pulled_pork_sandwich
7	french_fries	14	spring_rolls

Table 6 Hyperparameters and their values set to implement the model for FIC

Hyperparameter	Values
Optimizer	SGD with a momentum of 0.9 and an initial LR set of 0.1, which then gradually decreased to 0.0001
Dropout	0.3
Batch Size	32
Loss function	Categorical Cross entropy (from_logits = True)

most of the misclassified bread_pudding images (i.e., 29) are identified as apple_pie. Similarly, apple_pie misclassified images (i.e., 44) are identified as bread_pudding. Figure 11 presents the training and validation accuracy and loss curves of the EfficientNet B4 model, which show the training pattern with a fine-tuning strategy.

6.2 Object detection in food image using YOLOv5

Object detection identifies a specific type of object in the images with a certain degree of confidence and the location of the objects identified using the bounding box. Detection is defined by three attributes (the class of the object, the matching bounding box, and a confidence score). The confidence score is the probability that the algorithm successfully identified the object in the image. Object detection methods are divided into two groups. The R-CNN family (i.e., R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, and so on) and the YOLO families [21]. In this study, YOLOv5 is implemented to recognize the foods in the food images.

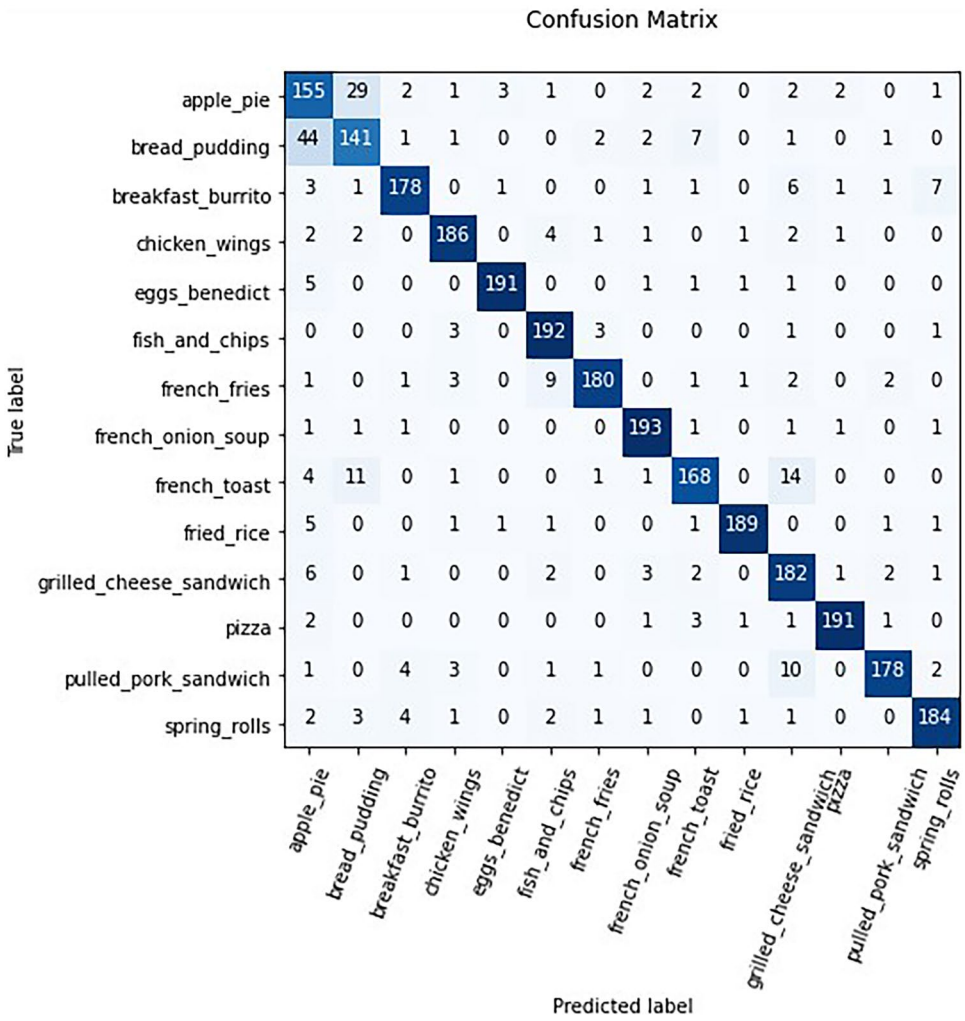

Fig. 10 Confusion matrix of CNN model for FIC

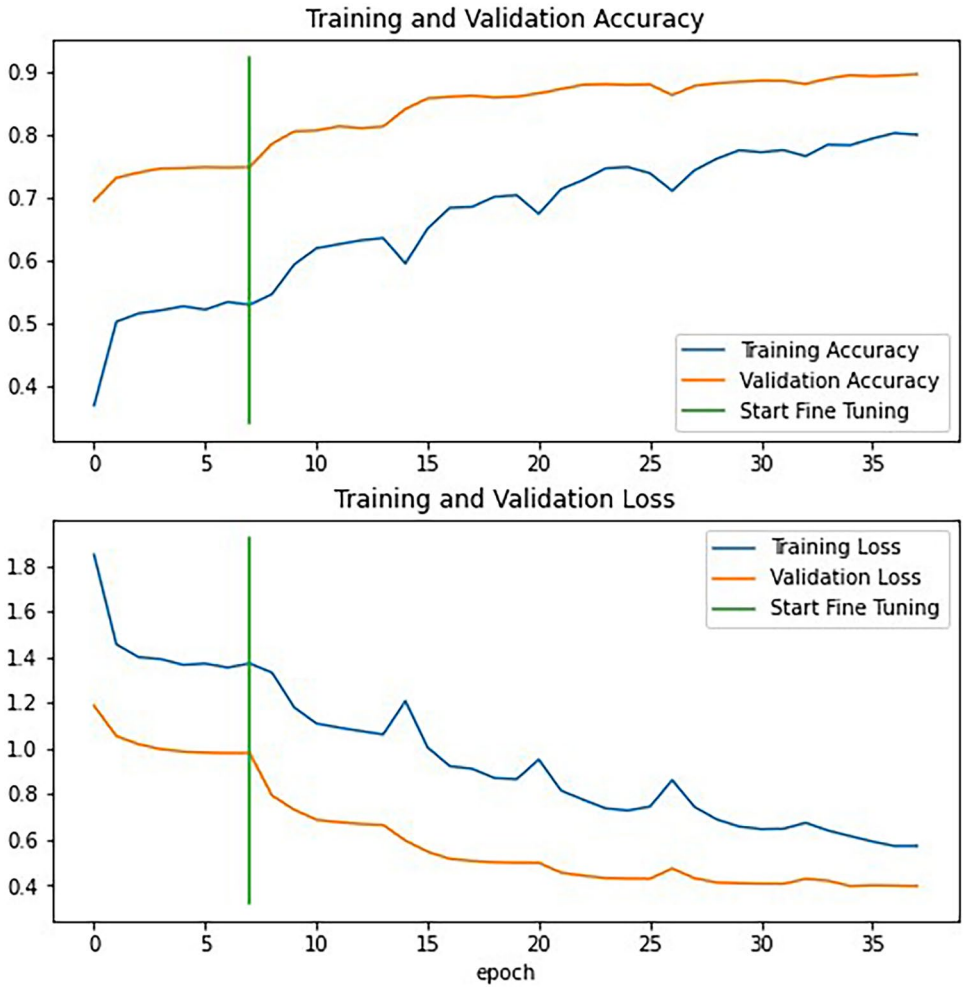

Fig. 11 Training and validation accuracy curve and loss curve of EfficientNet B4

The first step for the YOLOv5 implementation is the data preparation which consists of two types of data (images and annotations). For the food image, a dataset with 3 food classes (500 food images) is derived from FOOD-101. This customized dataset is prepared in YOLOv5 format with bounding box annotations. Roboflow tool and provided notebook [68] are used to annotate the food images. Image size is selected as 416*416, and the hyperparameters set optimizer as SGD with lr = 0.01, momentum = 0.937 and, weight_decay = 0.0005, batch size = 16, number of epochs as 250. Sample food images with a label, bounding box, and confidence score are shown in Fig. 12.

Recall, precision and mAP (Mean Average Precision) metrics are used to evaluate the performance of the object detection models. The performance metrics of the YOLOv5 model for food (french_fries, pizza, spring_rolls) are presented in Table 7. The PR curve, which shows a change of recall for a given precision, is shown in Fig. 13, and the F1-Confidence curve for food detection is presented in Fig. 14. Different model evaluation metrics curves for training (box loss, object loss, class loss, precision, recall) and validation (box loss, object loss, class loss, precision, recall) are presented in Fig. 15.

Table 7 Recall, precision, and mAP of YOLOv5 for food detection model

Class	precision	recall	mAP50
french_fries	0.67	0.65	0.67
pizza	0.72	0.77	0.77
spring_rolls	0.61	0.46	0.52
all classes	0.67	0.63	0.66

Figure 16 presents a sample food image taken UNIMIB2016 dataset, annotated with multiple foods. The object detection model can also be used to detect multiple foods in a single image.

7 Conclusion and future aspects

A dietary evaluation and healthcare monitoring system are essential to monitor an individual's nutrient intake and develop dietary predictions to track their nutritional needs. The main goal of this systematic review is to provide readers and ML practitioners with valuable insights that help to decide how to build CNN models and tune them appropriately in consideration of food image processing. In this work, investigations are identified from the previous researcher's results. Based on investigations, a research methodology is designed to solve the problem of FIC. Further, the research methodology proposes a solution for all the encountered investigations. In this analysis, several food image datasets are discussed, with several

Fig. 12 Sample food images for the YOLO5 detection model

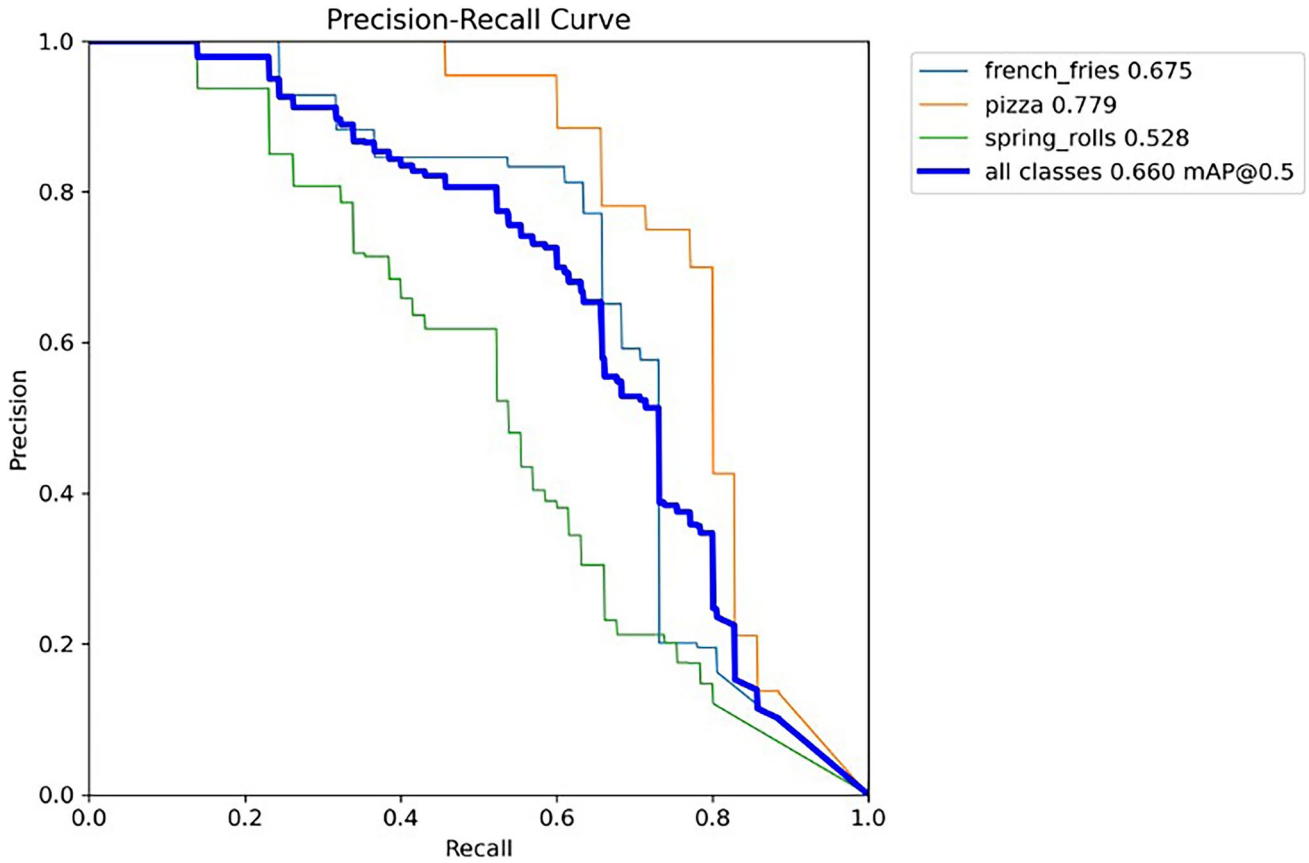

Fig. 13 PR (Precision-Recall) curve of the food detection model

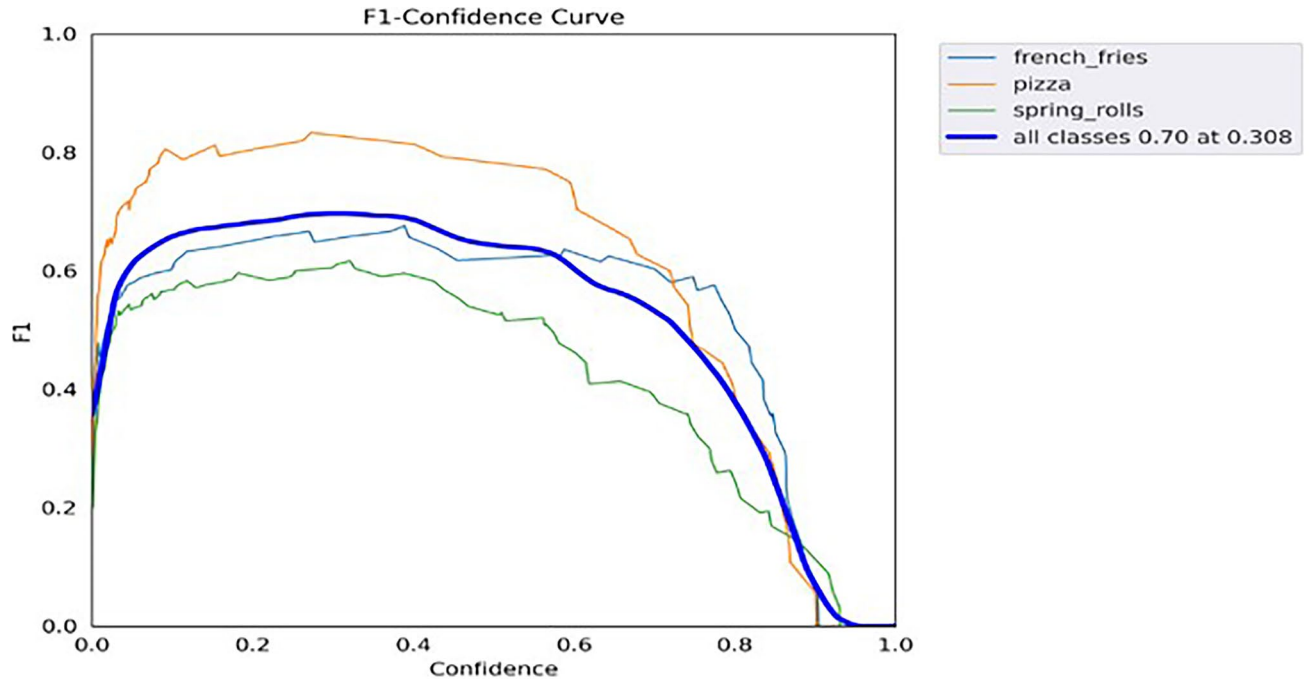

Fig. 14 F1-Confidence curve for food (french-fries, pizza, spring_rolls) detection

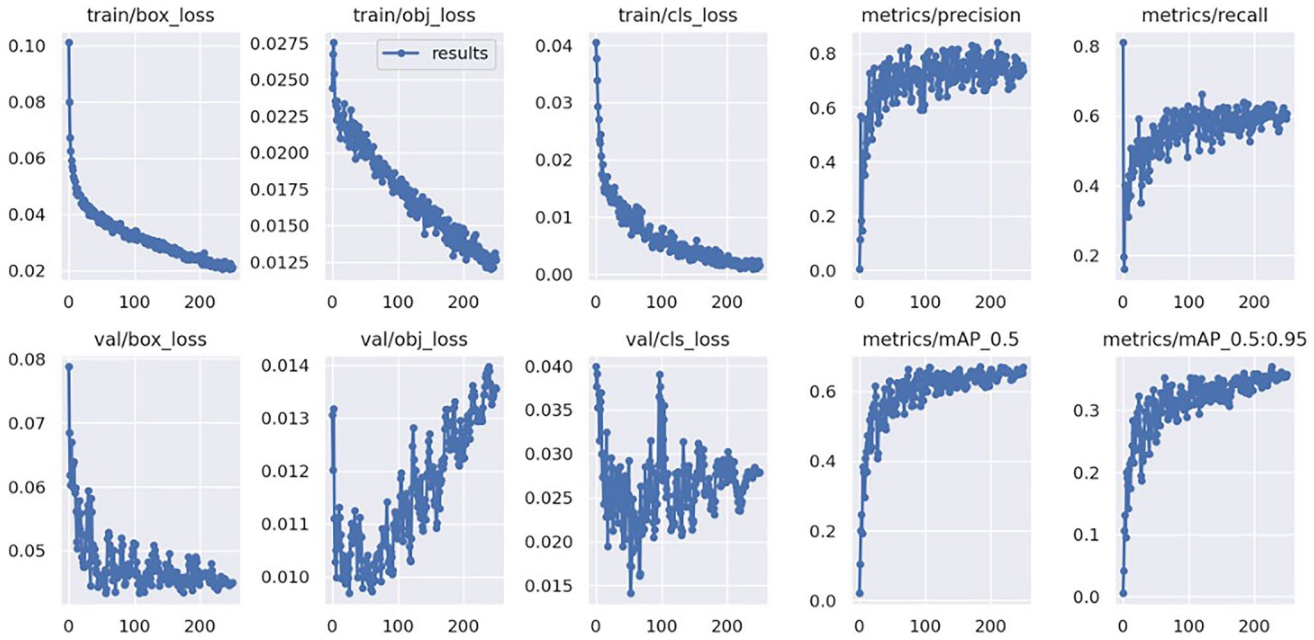

Fig. 15 Training (Box loss, Object loss, class loss, Precision, Recall) and Validation (Box loss, Object loss, class loss, Precision, Recall) curves

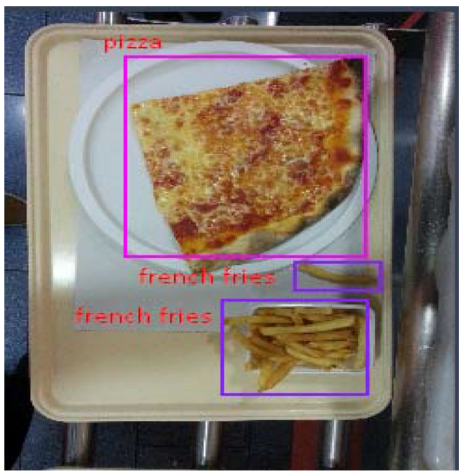

Fig. 16 Sample image of multiple food detection

food classes and total images in every dataset, and the articles used those datasets. A case study is introduced to solve a real-life problem: a food image dataset is used to train the EfficientNet model and achieves 86.5% accuracy. This systematic review will help the researchers to design food image-based diet recommendation systems.

Data availability Source mentioned.

Code availability We can upload as per requirement.

Declarations

Conflict of interest The authors declare no conflicts of interest regarding the present study.

References

1. Browne JD, Boland DM, Baum JT, Ikemiya K, Harris Q, Phillips M, Neufeld EV, Gomez D, Goldman P, Dolezal BA. Lifestyle modification using a wearable biometric ring and guided feedback improve sleep and exercise behaviors: a 12-month randomized, placebo-controlled study. Front Physiol. 2021;12:2094. https://doi.org/10.3389/fphys.2021.777874.
2. Park JH, Moon JH, Kim HJ, Kong MH, Oh YH. Sedentary lifestyle: Overview of updated evidence of potential health risks. Korean J Fam Med. 2020;41(6):365–73. https://doi.org/10.4082/kjfm.20.0165.
3. Celik O, Yildiz BO. Obesity and physical exercise. Minerva Endocrinol (Torino). 2021;46(2):131–44. https://doi.org/10.23736/S2724-6507.20.03361-1.
4. Rahmat RA, Kutty SB. Malaysian Food Recognition using Alexnet CNN and Transfer Learning. In 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). 2021;59–64. https://doi.org/10.1109/ISCAIE51753.2021.9431833.
5. Ege T, Yanai K. Image-based food calorie estimation using knowledge on food categories, ingredients and cooking directions. In Proceedings of the on Thematic Workshops of ACM Multimedia. 2017;367–75. https://doi.org/10.1145/3126686.3126742.
6. Castellana M, Biacchi E, Procino F, Casanueva FF, Trimboli P. Very-low-calorie ketogenic diet for the management of obesity, overweight and related disorders. Minerva Endocrinol (Torino). 2021;46(2):161–67. https://doi.org/10.23736/S2724-6507.20.03356-8.

7. Tejera-Pérez C, Sánchez-Bao A, Bellido-Guerrero D, Casanueva FF. The Southern European Atlantic diet. Minerva Endocrinol (Torino). 2021;46(2):145–60. https://doi.org/10.23736/S2724-6507.20.03381-7.
8. Aslan S, Ciocca G, Schettini R. Semantic food segmentation for automatic dietary monitoring. In 2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin) 2018;1–6. https://doi.org/10.1109/ICCE-Berlin.2018.8576231.
9. Darapaneni N, Singh V, Tarkar YS, Kataria S, Bansal N, Kharade A, Paduri AR. Food Image Recognition and Calorie Prediction. In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) 2021;1–6. https://doi.org/10.1109/IEMTRONICS52119.2021.9422510.
10. Wei P, Wang B. Food image classification and image retrieval based on visual features and machine learning. Multimedia Syst. 2022;28:2053–64. https://doi.org/10.1007/s00530-020-00673-6.
11. Situju SF, Takimoto H, Sato S, Yamauchi H, Kanagawa A, Lawi A. Food constituent estimation for lifestyle disease prevention by multi-task CNN. Appl Artif Intell. 2019;33(8):732–46. https://doi.org/10.1080/08839514.2019.1602318.
12. Xiao L, Lan T, Xu D, Gao W, Li C. A simplified CNNs visual perception learning network algorithm for foods recognition. Comput Electric Eng. 2021;92:107152. https://doi.org/10.1016/j.compeleceng.2021.107152.
13. Jia W, Li Y, Qu R, Baranowski T, Burke LE, Zhang H, Bai Y, Mancino JM, Xu G, Mao ZH, Sun M. Automatic food detection in egocentric images using artificial intelligence technology. Public Health Nutr. 2019;22(7):1168–79. https://doi.org/10.1017/S1368980018000538.
14. Jiang L, Qiu B, Liu X, Huang C, Lin K. DeepFood: Food image analysis and dietary assessment via deep model. IEEE Access. 2020;8:47477–89. https://doi.org/10.1109/ACCESS.2020.2973625.
15. Sahoo D, Hao W, Ke S, Xiongwei W, Le H, Achananuparp P, Lim EP, Hoi SC. FoodAI: Food image recognition via deep learning for smart food logging. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2019; 2260–68. https://doi.org/10.1145/3292500.3330734.
16. Mezgec S, Koroušić SB. NutriNet: a deep learning food and drink image recognition system for dietary assessment. Nutrients. 2017;9(7):657. https://doi.org/10.3390/nu9070657.
17. Min W, Liu L, Wang Z, Luo Z, Wei X, Wei X, Jiang S. ISIA Food-500: a dataset for large-scale food recognition via stacked global-local attention network. In Proceedings of the 28th ACM International Conference on Multimedia 2020;393–401. https://doi.org/10.1145/3394171.3414031.
18. Lohala S, Alsadoon A, Prasad PW, Ali RS, Altaay AJ. A novel deep learning neural network for fast-food image classification and prediction using modified loss function. Multimed Tools Appl. 2021;80:25453–76. https://doi.org/10.1007/s11042-021-10916-x.
19. Subhi MA, Ali SM. A deep convolutional neural network for food detection and recognition. In 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES). 2018; 284–87. https://doi.org/10.1109/IECBES.2018.8626720.
20. Aguilar E, Bolaños M, Radeva P. Regularized uncertainty-based multi-task learning model for food analysis. J Vis Commun Image Represent. 2019;60:360–70. https://doi.org/10.1016/j.jvcir.2019.03.011.
21. Jubayer F, Soeb JA, Mojumder AN, Paul MK, Barua P, Kayshar S, Akter SS, Rahman M, Islam A. Detection of mold on the food surface using YOLOv5. Curr Res Food Sci. 2021;4:724–8. https://doi.org/10.1016/j.crfs.2021.10.003.
22. Li Y, Xu X, Yuan C. Enhanced mask r-cnn for chinese food image detection. Math Probl Eng. 2020;2020:1–8. https://doi.org/10.1155/2020/6253827.
23. Son GJ, Kwak DH, Park MK, Kim YD, Jung HC. U-Net-based foreign object detection method using effective image acquisition system: a case of almond and green onion flake food process. Sustainability. 2021;13(24):13834. https://doi.org/10.3390/su132413834.
24. Tasci E. Voting combinations-based ensemble of fine-tuned convolutional neural networks for food image recognition. Multimed Tools Appl. 2020;79(41–42):30397–418. https://doi.org/10.1007/s11042-020-09486-1.
25. Ma P, Lau CP, Yu N, Li A, Sheng J. Application of deep learning for image-based Chinese market food nutrients estimation. Food Chem. 2022;373:130994. https://doi.org/10.1016/j.foodchem.2021.130994.
26. Hu L, Zhang W, Zhou C, Lu G, Bai H. Automatic diet recording based on deep learning. In 2018 Chinese Automation Congress (CAC) 2018;3778–82. https://doi.org/10.1109/CAC.2018.8623474.
27. Qiu J, Lo FP, Lo B. Assessing individual dietary intake in food sharing scenarios with a 360 camera and deep learning. In 2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN) 2019;1–4. https://doi.org/10.1109/BSN.2019.8771095.
28. Metwalli AS, Shen W, Wu CQ. Food image recognition based on densely connected convolutional neural networks. In 2020 international conference on artificial intelligence in information and communication (ICAIIC) 2020;027–32. https://doi.org/10.1109/ICAIIC48513.2020.9065281.
29. Hasan HM, Khan H, Asif T, Hashmi S, Rafi M. Towards a transfer learning approach to food recommendations through food images. In Proceedings of the 3rd International Conference on Machine Learning and Soft Computing 2019;99–105. https://doi.org/10.1145/3310986.3310990.
30. Wu MY, Lee JH, Hsueh CY. A framework of visual checkout system using convolutional neural networks for bento buffet. Sensors. 2021;21(8):2627. https://doi.org/10.3390/s21082627.
31. Yunus R, Arif O, Afzal H, Amjad MF, Abbas H, Bokhari HN, Haider ST, Zafar N, Nawaz R. A framework to estimate the nutritional value of food in real time using deep learning techniques. IEEE Access. 2019;7:2643–52. https://doi.org/10.1109/ACCESS.2018.2879117.
32. Farooq M, Sazonov E. Feature extraction using deep learning for food type recognition. In Bioinformatics and Biomedical Engineering: 5th International Work-Conference, IWBBIO 2017, Granada, Spain, 2017; 464–72 (Springer International Publishing). https://doi.org/10.1007/978-3-319-56148-6_41.
33. Ruenin P, Bootkrajang J, Chawachat J. A system to estimate the amount and calories of food that elderly people in the hospital consume. In Proceedings of the 11th International Conference on Advances in Information Technology 2020;1–7. https://doi.org/10.1145/3406601.3406613.
34. Memiş S, Arslan B, Batur OZ, Sönmez EB. A comparative study of deep learning methods on food classification problem. In 2020 Innovations in Intelligent Systems and Applications Conference (ASYU) 2020;1–4. https://doi.org/10.1109/ASYU50717.2020.9259904.
35. Xu B, He X, Qu Z. Asian food image classification based on deep learning. J Comput Commun. 2021;9(03):10. https://doi.org/10.4236/jcc.2021.93002.
36. Özsert Yiğit G, Özyildirim BM. Comparison of convolutional neural network models for food image classification. J Inf Telecommun. 2018;2(3):347–57. https://doi.org/10.1080/24751839.2018.1446236.
37. Phiphiphatphaisit S, Surinta O. Food image classification with improved MobileNet architecture and data augmentation. In Proceedings of the 3rd International Conference on Information Science and Systems 2020;51–6. https://doi.org/10.1145/3388176.3388179.
38. Rajayogi JR, Manjunath G, Shobha G. Indian food image classification with transfer learning. In 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS) 2019;1–4. https://doi.org/10.1109/CSITSS47250.2019.9031051.

39. Matsuda Y, Yanai K. Multiple-food recognition considering co-occurrence employing manifold ranking. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012) 2012;2017–20.
40. Tan RZ, Chew X, Khaw KW. Quantized deep residual convolutional neural network for image-based dietary assessment. IEEE Access. 2020;8:111875–88. https://doi.org/10.1109/ACCESS.2020.3003518.
41. Won CS. Multi-scale CNN for fine-grained image recognition. IEEE Access. 2020;8:116663–74. https://doi.org/10.1109/ACCESS.2020.3005150.
42. Hu H, Zhang Z, Song Y. Image based food calories estimation using various models of machine learning. In 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE) 2020;1874–78. https://doi.org/10.1109/ICMCCE51767.2020.00411.
43. Bossard L, Guillaumin M, Van Gool L. Food-101–mining discriminative components with random forests. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VI 13 2014;446–61. https://doi.org/10.1007/978-3-319-10599-4_29.
44. Kawano Y, Yanai K. Foodcam-256: a large-scale real-time mobile food recognitionsystem employing high-dimensional features and compression of classifier weights. In Proceedings of the 22nd ACM international conference on Multimedia 2014;761–62. https://doi.org/10.1145/2647868.2654869.
45. Chen J, Ngo CW. Deep-based ingredient recognition for cooking recipe retrieval. In Proceedings of the 24th ACM international conference on Multimedia 2016;32–41. https://doi.org/10.1145/2964284.2964315.
46. Aslan S, Ciocca G, Mazzini D, Schettini R. Benchmarking algorithms for food localization and semantic segmentation. Int J Mach Learn Cybern. 2020;11(12):2827–47. https://doi.org/10.1007/s13042-020-01153-z.
47. Ciocca G, Napoletano P, Schettini R. Food recognition and leftover estimation for daily diet monitoring. In New Trends in Image Analysis and Processing - ICIAP 2015 Workshops. 2015;334–41. https://doi.org/10.1007/978-3-319-23222-5.
48. Ciocca G, Napoletano P, Schettini R. Food recognition: a new dataset, experiments, and results. IEEE J Biomed Health Inform. 2016;21(3):588–98. https://doi.org/10.1109/JBHI.2016.2636441.
49. Chen X, Zhu Y, Zhou H, Diao L, Wang D. Chinesefoodnet: a large-scale image dataset for chinese food recognition. arXiv preprint: arXiv:1705.02743. 2017. https://doi.org/10.48550/arXiv.1705.02743.
50. Kaur P, Sikka K, Wang W, Belongie S, Divakaran A. Foodx-251: a dataset for fine-grained food classification. arXiv preprint arXiv:1907.06167. 2019;2–6. https://doi.org/10.48550/arXiv.1907.06167.
51. Ma P, Lau CP, Yu N, Li A, Liu P, Wang Q, Sheng J. Image-based nutrient estimation for Chinese dishes using deep learning. Food Res Int. 2021;147:110437. https://doi.org/10.1016/j.foodres.2021.110437.
52. Chen J, Zhu B, Ngo CW, Chua TS, Jiang YG. A study of multitask and region-wise deep learning for food ingredient recognition. IEEE Trans Image Process. 2020;1514–1526. https://doi.org/10.1109/TIP.2020.3045639.
53. Shekar BH, Dagnew G. Grid search-based hyperparameter tuning and classification of microarray cancer data. In 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP). 2019;1–8. https://doi.org/10.1109/ICACCP.2019.8882943.
54. Setyono NF, Chahyati D, Fanany MI. Betawi traditional food image detection using ResNet and DenseNet. In 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). 2018;441–45. https://doi.org/10.1109/ICACSIS.2018.8618175.
55. Ciocca G, Napoletano P, Schettni R. CNN-based features for retrieval and classification of food images. Comput Vis Image Underst. 2018;176:70–7. https://doi.org/10.1016/j.cviu.2018.09.001.
56. McAllister P, Zheng H, Bond R, Moorhead A. Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. Comput Biol Med. 2018;95:217–33. https://doi.org/10.1016/j.compbiomed.2018.02.008.
57. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016;770–8. https://doi.org/10.1109/CVPR.2016.90.
58. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014;1–13. https://doi.org/10.48550/arXiv.1409.1556.
59. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning. 2019;6105–14. https://doi.org/10.48550/arXiv.1905.11946.
60. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2015;1–9. https://doi.org/10.1109/CVPR.2015.7298594.
61. Kumari M, Singh T. Food image to cooking instructions conversion through compressed embeddings using deep learning. In 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW) 2019;81–4. https://doi.org/10.1109/ICDEW.2019.00-31.
62. Islam KT, Wijewickrema S, Pervez M, O'Leary S. An exploration of deep transfer learning for food image classification. In 2018 Digital Image Computing: Techniques and Applications (DICTA). 2018;1–5. https://doi.org/10.1109/DICTA.2018.8615812.
63. Ege T, Yanai K. Image-based food calorie estimation using recipe information. IEICE Trans Inf Syst. 2018;101(5):1333–41. https://doi.org/10.1587/transinf.2017MVP0027.
64. Tai TT, Thanh DN, Hung NQ. A dish recognition framework using transfer learning. IEEE Access. 2022;10:7793–9. https://doi.org/10.1109/ACCESS.2022.3143119.
65. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans Med Imaging. 2016;35(5):1299–312. https://doi.org/10.1109/TMI.2016.2535302.
66. Hattori T, Doman K, Ide I, Mekada Y. Application of data augmentation for accurate attractiveness estimation for food photography. In Proceedings of the 11th Workshop on Multimedia for Cooking and Eating Activities. 2019;33–40. https://doi.org/10.1145/3326458.3326927.
67. Shen Z, Shehzad A, Chen S, Sun H, Liu J. Machine learning based approach on food recognition and nutrition estimation. Procedia Comput Sci. 2020;174:448–53. https://doi.org/10.1016/j.procs.2020.06.113.
68. Dwyer B, Nelson J, Solawetz J, et al. Roboflow (Version 1.0). In: Computer Vision. 2022. https://roboflow.com. Accessed 15 Feb 2023.