

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,700

Open access books available

182,000

International authors and editors

195M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

Visual Recognition of Food Ingredients: A Systematic Review

*Michail Marinis, Evangelos Georgakoudis, Eleni Vrochidou
and George A. Papakostas*

Abstract

The use of machine learning for visual food ingredient recognition has been at the forefront in recent years due to its involvement in numerous applications and areas such as recipe discovery, diet planning, and allergen detection. In this work, all relevant publications from 2010 to 2023 were analyzed, including databases such as Scopus, IEEE Xplore, and Google Scholar, aiming to provide an overview of the methodologies, challenges, and potential of this emerging field. Challenges, such as visual differences and complicated ingredient composition, are highlighted, along with the importance of data preprocessing, image preparation methods, and the use of deep learning techniques for state-of-the-art performances. The potential applications of this technology in the fields of automation and robotics are explored, and existing datasets are provided. Research concluded that among the several machine learning techniques being used, the reported performances of convolutional neural networks (CNNs) rate them on top of all approaches that are currently being used.

Keywords: visual recognition, food ingredient recognition, support vector machines (SVM), convolutional neural networks (CNNs), feature extraction, computer vision

1. Introduction

Food and nutrition industry is only one of the several industries that have benefited from the recent breakthroughs in computer vision and machine learning [1]. Visual recognition of food ingredients [2] is a promising topic of study since it has the potential to promote the food industry, as well as endorse health monitoring and nutritional analysis. Artificial intelligence (AI) and image processing have allowed for the development of visual recognition systems that can accurately identify and categorize food items based solely on their outward appearance [3].

There are many potential outcomes stemming from the automatic recognition of food ingredients from photographs [4]. Consumers can expect enhanced dietary options, tailored nutrition suggestions, and easier administration of food allergies and intolerances. Visual recognition systems can improve food quality control, speed up the identification of ingredients, and streamline stock management in the food business. These methods would allow scientists to study public's health, investigate dietary patterns, and evaluate the nutritional value of food on a massive scale.

A convolutional neural network (CNN) [5] is a powerful machine learning algorithm that has played a significant role in the advancement of visual recognition systems for food items. By using layers of convolutional and pooling processes, CNNs provide a type of deep learning model that is particularly effective at extracting meaningful characteristics from images. Several computer vision tasks, such as picture classification and object identification, have achieved astounding success on their part.

Since people are becoming more interested in being able to see what's in food and considering the benefits associated with it, are the main reasons that motivated the current study. The goal of this systematic study is to provide a full analysis of the most up-to-date methods, datasets, evaluation standards, and problems that come with recognizing food ingredients by sight. Through careful analysis and synthesis of the available literature, this review aims to identify research gaps, point out promising methods, and make suggestions for future research areas. The following are the primary aims of this analysis:

1. Study visual recognition of food items, encompassing methods such as image capture, preprocessing, feature extraction, and classification using neural networks.
2. Provide all available datasets that are used to train and test food item recognition algorithms and assess their quality.
3. Assess the accuracy and robustness of visual recognition systems by analyzing the performance indicators and using evaluation methodologies.
4. Consider all related challenges such as illumination changes, occlusions, and the presence of similar-looking substances, while discussing the difficulties and restrictions of visual recognition of food items.
5. Investigate how visual recognition technologies might improve public health, nutrition, and the food sector.

This systematic study seeks to provide a comprehensive picture of the present status of visual recognition of food ingredients by consolidating the existing knowledge. The results will add to the existing body of literature and will be able to provide useful insights for researchers, practitioners, and policymakers interested in applying computer vision and AI, particularly CNNs, to the analysis and nutrition of food.

2. Material and methods

2.1 Review methodology

In this work, we used a systematic review methodology to locate, evaluate, and synthesize studies that were applicable to the study of “visual food recognition.” The primary goal was to analyze the previous research in this field reflectively and critically. The review was conducted in accordance with the following standards:

1. Formulating search criteria:

- All articles had to be written in English.
- All articles had to be published between January 2010 and April 2023.

Over the past 13 years, there has been a surge in academic interest in exploring the potential benefits of vision computing. As a result, we limited our analysis to papers published during these years, between 2010 and 2023. **Figure 1** shows the breakdown, by year of publication, of the research we gathered.

1. Database Search:

- We conducted searches using the established criteria in prominent databases, including Google Scholar, Scopus, Web of Science, IEEE Xplore, and ScienceDirect. These databases were selected based on their comprehensive coverage of scientific literature in various disciplines.

2. Extraction of Qualitative Research:

- We extracted qualitative research studies that focused on visual food recognition. This included studies that utilized different methodologies, datasets, and machine learning algorithms to analyze and classify food ingredients based on visual cues.

3. Data Extraction:

- Relevant information, including study design, dataset details, methodologies employed, performance metrics, and key findings, was extracted from each selected study. This allowed us to obtain a comprehensive understanding of the approaches used in visual food recognition.

4. Data Analysis and Synthesis:



Figure 1.
Publications per year.

- The extracted data from the selected studies were systematically analyzed and synthesized. Common themes, trends, similarities, and differences in methodologies and results were identified and compared across the studies.

5. Drawing Conclusions:

- Based on the analysis and synthesis of the collected data, we drew conclusions regarding the current state-of-the-art techniques, datasets, evaluation metrics, and challenges in visual food recognition. These conclusions provide valuable insights into the field and could serve as a foundation for future research directions.

The research was conducted using a combination of search terms related to visual food recognition, such as “visual recognition of food ingredients,” “machine learning,” and “food image classification.” The specific search terms were adapted and used across the selected databases to ensure a comprehensive search. We prioritized recent and validated research by utilizing IEEE and Scopus, while Google Scholar provided a broader range of articles.

By following this systematic review methodology, we aimed to provide a comprehensive overview of the existing literature on visual food recognition, offering valuable insights and guidance for researchers, practitioners, and policymakers interested in this field. After the research was carried out in the above manner, for the purpose of this review, only the studies and papers that have been published in journals and were written in English were selected, considered as more valid, with documented results, greater clarity, and argumentation. We also selected papers based on qualitative research, quantitative, and experimental studies as they appear to conduct more valid results.

2.2 Final research material

From the searches resulting by using the above terms, we limited our research to a total of 55 articles between 2010 and 2023. After removing duplicates and rejecting those that did not comply with the predefined criteria, we ended up with 19 papers. **Figure 2** illustrates the PRISMA chart [6], showing the total number of found articles and the selection process of papers to conduct the systematic review.

All 19 studies that were analyzed for inclusion in this systematic review concentrated on various aspects of food recognition. **Figure 3** displays the distribution of our collected data among the various search database engines we used.

3. Data analysis

The many aspects of visual food recognition and their possible implications in the context of ingredient identification and analysis will be discussed below. In the food industry and nutrition profession, understanding these traits is essential for designing effective management measures and limiting harmful effects. To better understand the origins of data on food ingredients and their potential effects, it is helpful to understand the characteristics of visual food recognition.

Data collection and analysis in visual food recognition rely heavily on machine learning techniques, such as support vector machines and CNNs. These methods

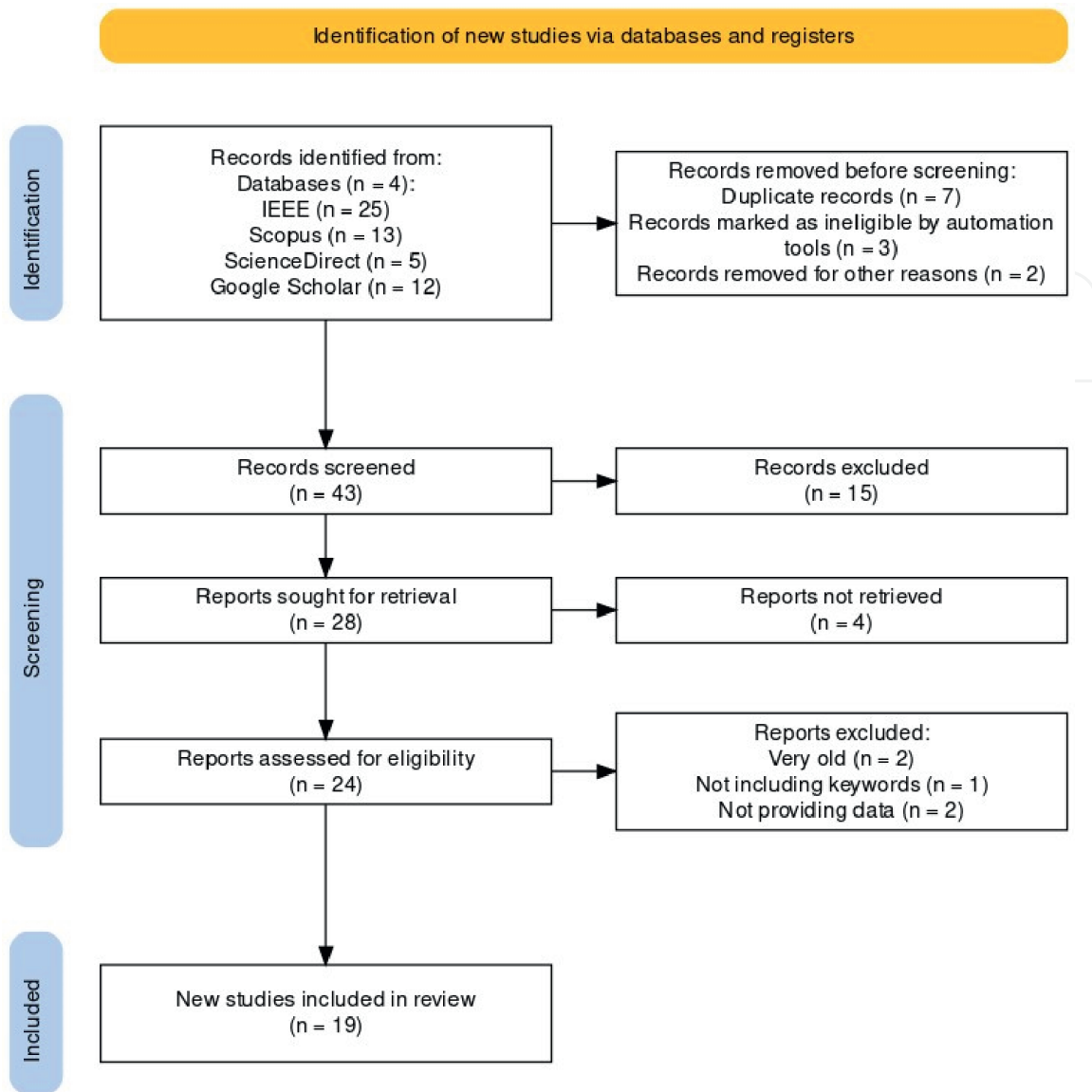


Figure 2.
Prisma diagram.

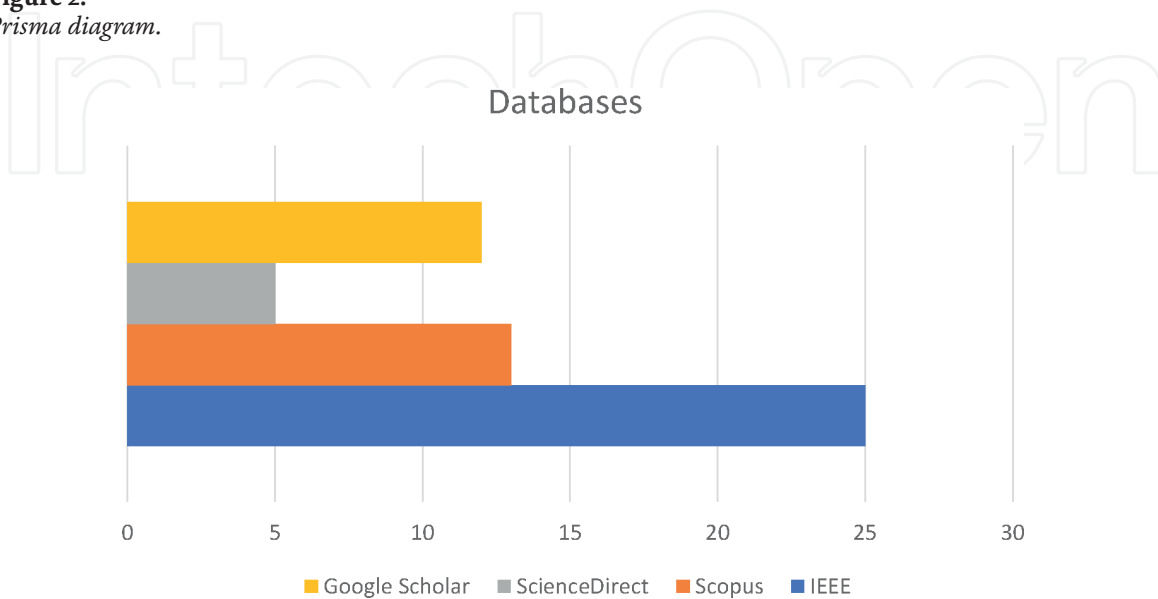


Figure 3.
Classification of selected papers by database.

permit us to better classify ingredients by allowing us to extract useful features and build classifiers based on input properties. In addition, using more specific input features can help to reduce processing time, which, in turn, improves recognition efficiency.

The extraction of key features and construction of robust classifiers are crucial for the visual identification classification of food items. With the aid of feature extraction and classification models, we may learn more about this crucial subject. To further our understanding of what goes into our food and how to improve its nutritional value, we may use machine learning algorithms to harness the full potential of visual food recognition.

During our research, we looked at several different studies concerning the recognition of food images and the identification of ingredients. During our investigation, we came across a few prominent studies that contributed significant new information to the existing body of research. These studies have implemented a diverse selection of deep learning models, datasets, and evaluation standards in order to determine how successfully their methodologies work. In the parts that are to follow, we will provide a summary of the findings of these studies, with an emphasis on the methodologies, datasets, and performance metrics that were used. We anticipate that by examining these findings, we will be able to give a comprehensive assessment of existing approaches for food image recognition and component detection, shedding light on the successes and failures of this rapidly developing field of study. **Table 1** includes all selected papers, along with their used model, and their corresponding results.

As in most image processing applications, the dominant base model used for feature extraction and prediction is a CNN model. Across all 19 papers we analyzed, all of them used an existing model, which was then extended using transfer learning [11] or used as a backbone for a brand-new model [14]. Discussion and analysis of research findings are provided in the upcoming section.

4. Food datasets

The availability of diverse datasets that have been meticulously annotated has led to significant advancements in the field of food photo identification in recent years. Machine learning models in the area of visual food recognition could benefit greatly from using these datasets as training and evaluation resources.

- Food-101 [9, 11, 13, 27]: A popular benchmark for food picture recognition systems. It has 1000 photos for each of 101 different food categories for a total of 101,000. Fruits, vegetables, desserts, beverages, and a wide variety of entrees are all represented in the dataset. Both unprocessed materials and finished dishes are included in the dataset. It provides a large set of photos for testing and training food identification models, which helps to speed up the process of creating reliable technologies. This dataset was the most used dataset among all papers and provided the best results.
- Food-11 [11]: The culinary-11 collection includes roughly 9000 photos of various food products from 166 different culinary categories. It includes a wide variety of foods, from sweets to fruits to vegetables to entrees and beyond. The dataset includes a wide variety of foods from a variety of different categories, making it useful for testing and training food recognition systems.

Reference	Model	Dataset	Results
Chen et al. [7]	DCNN (Deep Convolutional Neural Network) for known ingredient classifying, mRGCN (multi-relational Graph Convolutional Network - proposed new model) for unknown ingredient prediction	VIREO Food 172, UEC Food-100 (110,241 and 14,136 images, respectively)	Top-K hit ratio: <ul style="list-style-type: none"> • Hit@10: 47.4% unseen ingredients on VIREO, 24.3% on UEC • Hit@20: 48.8% unseen ingredients on VIREO, 42% on UEC
Chen et al. [8]	Multi-task DCNN model for food ingredient recognition and single-task DCNN model for ingredient label (10 ingredients) prediction at image regions	VIREO Food-251 for 251 food categories and 406 ingredient labels	Macro-F1 score: <ul style="list-style-type: none"> • Up to 61.74% for multi-task learning • Up to 95.7% for single-task learning
Alahmari and Salem [9]	CNN (cascaded two-head for multiple recognitions, state and food type, and non-cascaded just for just state)	7563 images for 17 commonly used ingredients	<ul style="list-style-type: none"> • Non-cascaded model: 81% accuracy, 82% precision, 81% recall, 81% F1 score • Cascaded multiheaded model: 87% accuracy, precision, recall and F1 score for food state, 71.35% accuracy, 72% precision, 71% recall, 70% F1 score for food ingredient type
Ishichi et al. [10]	U-Net (convolutional network architecture U-net: Convolutional networks for biomedical image segmentation) with 30 epochs, batch size 8, categorical cross-entropy loss function, Adam optimizer, learning rate: 0.001	10,000 images, generated under three different transparency conditions	<ul style="list-style-type: none"> • Conditions A: ~72.1% average correct answers • Conditions B: ~88% average correct answers • Conditions C: ~92.3% average correct answers
Morol et al. [11]	CNN using transfer learning from ResNet50	Custom dataset, including data from Food101, Fruit 360 and UECFOOD256, 9856 images in total	<ul style="list-style-type: none"> • 99.71% accuracy on training dataset • 92.6% on validation dataset
Christian et al. [12]	MobileNet (CNN-based models for use in mobile and embedded applications - A mobile application for food and its ingredients detection using deep learning), retrained using different gradient descent optimizers	Custom dataset, created <i>via</i> Firefox add-on, scrapping images from Google and Bing Images, 32,914 images in total	<ul style="list-style-type: none"> • Average accuracy: 49.4% • Min. accuracy: 42% (RMSProp Optimizer) • Max. accuracy: 58% (Adam Optimizer)
Pan et al. [13]	CBNet (Combinational Convolutional Network) – a new proposed model, based on VGGNet, ResNet, and DenseNet	Food-41 dataset, 4100 images in total	Fine-tuning last the layer: <ul style="list-style-type: none"> • CBNet-VR: 88.90% accuracy • CBNet-VD: 89.47% accuracy • CBNet-RD: 88.33% accuracy Fine-tuning the whole network: <ul style="list-style-type: none"> • CBNet-VR: 94.03% accuracy • CBNet-VD: 95.00% accuracy • CBNet-RD: 95.28% accuracy

Reference	Model	Dataset	Results
Zhu and Dai [14]	CNN-based model (1x1convnet), consists of 1x1 convolutional layers, using ResNet50 and AlexNet as backbones of the framework	Custom dataset, 4131 images in total	<p>F1-score:</p> <ul style="list-style-type: none"> Level 1 hierarchical segmentation: 52% seafood, 97% crop, 57% livestock Level 2 hierarchical segmentation: 9% nuts, 16% fruits, 77% vegetables, 28% cereals, 23% pulses, 15% fungi, 17% potatoes Level 3 hierarchical segmentation: 40% stems, 28% fruits, 39% leaves, 0% flowers, 34% roots Level 3 non-hierarchical segmentation” 18% stems, 52% fruits, 28% leaves, 0% flowers, 8% roots <p>Precision:</p> <ul style="list-style-type: none"> Level 3 hierarchical segmentation: 27% stems, 49% fruits, 28% leaves, 0% flowers, 27% roots Level 3 non-hierarchical segmentation: 42% stems, 48% fruits, 35% leaves, 0% flowers, 35% roots <p>Recall:</p> <ul style="list-style-type: none"> Level 3 hierarchical segmentation: 42% stems, 48% fruits, 35% leaves, 0% flowers, 34% roots Level 3 non-hierarchical segmentation: 88% stems, 19% fruits, 68% leaves, 0% flowers, 1% roots
Pan et al. [15]	A proposed framework combining a two-level CNN for feature extraction, PCA , CFS , IG for feature evaluation, SMO (Sequential minimal optimization, improvement of SVM) for training the model	MLC-41 dataset, 41 food labels, 100 images for each, 4100 images in total, based on the MLC dataset by Mealcome	<ul style="list-style-type: none"> Best deep learning/classifier model accuracy: ResNet/SMO: 87.781% average accuracy
Hoashi et al. [16]	SVM using multiple kernel learning (MKL) to integrate various kinds of image features. Features include color, BoF, Gabor, and gradient histogram	Custom dataset built from the Internet for 85 kinds of food, each represented by 100 images, 8500 images in total	<p>Classification rate:</p> <ul style="list-style-type: none"> 61.34% for 50-kind food classification 62.52% for 85-kind food classification 45.3% for cellular-phone camera photos (users were not instructed on how to take a proper photo)

Reference	Model	Dataset	Results
Qayyum and Sah, [17]	InceptionV3 CNN model provided by Keras , converted to CoreML model for use in application development	5000 images in total, 15 images/class in the training set, 5 images/class in the testing set	Accuracy ranging between 80% and 97% across 101 classes
Zhang et al. [18]	SRN (Spatial Regularization Network) model, similar to ResNet101 when it comes to general prediction net	MV80-Market Dataset: Custom dataset of multi-labeled vegetable images, 80 classes from Market, authors aim to solve the lack of robustness of available lab-controlled image datasets, 15,798 images in total	SRN results: <ul style="list-style-type: none"> • mAP (mean average precision over classes): 77.2% • macro/micro precision (P-C/P-O): 73.7% and 77.3%, respectively • macro/micro recall (R-C/R-O): 70.7% & 74.7%, respectively • macro/micro F1-measure (F1-C/F1-O): 72% and 76%, respectively
Liu et al. [19]	AFN (Attention Fusion Network) and the food-ingredient Joint learning module: <ul style="list-style-type: none"> • AFN: Divided attention part, which preserves more discriminating features for recognition, and fusion part, which generates feature embeddings for fine-grained food and ingredient recognition. VGGNet and ResNet Backbone • Food-ingredient Joint Learning Module: A balance focal loss function, used to tackle the imbalance of multi-labels of ingredients in a dish and enhance learning ability 	VIREO Food-172 dataset: 172 food categories and 353 ingredient categories, 110,241 images in total	Used accuracy, Micro-F1 and Macro-F1 metrics to measure performance, with different backbones and different methods. Above metrics for performance comparison on ingredient recognition: Accuracy: 34.29% (Best achieved with the proposed method and ResNetSt269 backbone) Micro-F1: 74.1% (Best achieved with the proposed method and ResNetSt269 backbone) Macro-F1: 58.8% (Best achieved with the proposed method and ResNet152 backbone)
He et al. [20]	SVM , using SIFT features for performance comparison	Custom dataset, 15,262 images in total, 55 American food categories <i>via</i> Google Image search	Multi-view kernel SVM: ~90% accuracy Single-view kernel SVM: ~68% accuracy Texture-based SVM: ~49% accuracy SIFT-based nearest neighbor classifier: ~40% accuracy
Madival and Jawaligi, [21]	DBN classifier , Textual features, SIFT and deep features, weight tuning using Improved TDO (ITDO) model	Recipes5k (University of Barcelona)	<ul style="list-style-type: none"> • Results of DBN + ITDO: • F1-score: 94.825% • Accuracy: 93.944%

Reference	Model	Dataset	Results
M. Zhang et al. [22]	NN-based model , double-flow feature fusion module (DFFF), reinforcement learning is achieved by a hybrid loss function, dual learning used to boost the model performance of sequential ingredient recognition	Recipe 1 M, after pre-processing, 361,308 images in total	Results vary by method, the best scores of F1 are around 75%
Sahoo et al. [23]	CNN with transfer learning	FoodAI-756, ~400,000 images in total	Average accuracy: 80.09%
Mezgec and Seljak, [24]	DCNN, AlexNet as the backbone	520 categories, 225,953 images in total	Average accuracy: 55%
Park et al. [25]	DCNN	23 categories, 92,000 images in total	Average accuracy: 91.3%
Cornejo et al. [26]	CNN	36 categories, 3600 images in total	Average accuracy: 85%

Table 1.

Analysis of selected papers, used model, dataset, and performance results.

- UEC-FOOD100 [13]: Dedicated solely to Japanese cuisine. There are 100 different types of cuisine shown with a total of 13,000 pictures. The collection includes photographs of a wide range of authentic Japanese cuisine taken from a variety of vantage points, including straight on and from the side. It also uses pictures taken in a variety of lighting conditions to represent real-life situations. The UEC-FOOD100 dataset is a curated photo archive useful for researching and identifying elements of the Japanese culinary tradition.
- Food-5 K [10]: It was created to test food recognition systems under realistic conditions. It has 5000 pictures of food, split up among 250 different categories. There are 20 pictures in each category. These photos were taken in a wide variety of settings, each with its own lighting, backdrop, and scale. The dataset's varied visual attributes and difficult scenarios are designed to put food identification models through their paces.
- ChineseFoodNet collection [28]: It has 192,000 photos of Chinese cuisine, organized into 208 categories, making it the largest image dataset for Chinese food categorization to date.
- Instagram800K [29]: This dataset is generated by using Instagram API. A total of 808,964 pictures are included, all of which have either general food-related tags or pictures of specific foods attached to them. Included in the dataset are the top 43 most-used food-related tags, such as #lunch and #foodie. It also features 53 of the most searched for foods, such as #pasta and #steak, with accompanying photographs. The collection includes not just photos of food but also metadata about the images and the food itself, which may be used for analysis and research.

- AIFood [30]: The dataset includes a great diversity of cuisines, recipes, and ingredients. It is designed to be applicable to the creation of models that can reliably recognize and classify different types of food, and it attempts to cover a wide range of culinary cultures and dietary preferences.
- ChinaFood-100 [31]: It has been developed aiming to better categorize Chinese cuisine. The calories, protein, fat, carbs, vitamins, and micronutrients for each food group are all included in this dataset.

These datasets have been crucial to the development of food picture recognition technology. For the purpose of training and evaluating machine learning models for accurate and efficient food recognition tasks, they supply researchers with tagged photos across multiple food categories.

5. Food ingredient recognition stages

5.1 Problem description

The main reasons for food ingredients recognition are the following:

- For food safety, consumers demand safe products for their health. Food recognition can ensure consumer-based testing of food ingredients for safe consumption, for example, of allergy-free, gluten-free products.
- For issues related to standards and regulations guidelines, governments impose regulations related to food analysis, regarding specific compositions and nutrients, for example, to detect unwanted compounds, determine the authenticity of products.
- For food quality control, food providers need to test the quality of their products before releasing them to the market, for example, for raw, defective, rotten ingredients.
- For promoting further research, food ingredient recognition may constitute the first step for further advancements in food industry, for example, for visual identification of food chemical compositions, personalized nutrition, sustainable food production toward reduction of food waste, food recommendation [32, 33].

The main stages of food ingredient recognition are three: (1) the preprocessing step, (2) the food segmentation stage, and (3) the food recognition stage. All stages are analyzed thoroughly in the upcoming sections. The general flow diagram of food recognition is illustrated in **Figure 4**.

The preprocessing step includes image processing techniques toward improving the image quality and, thus, facilitating the next steps of the process. Image segmentation refers to the process of dividing an image into segments that can be further processed separately. Image segmentation in food images is used to locate the food ingredients and their boundaries, to properly separate them, and thus, to reduce image complexity and enable the further processing of each segment, that is. food ingredient, separately. Food recognition involves a trained classification algorithm able to identify each segmented food ingredient. The algorithm first extracts feature

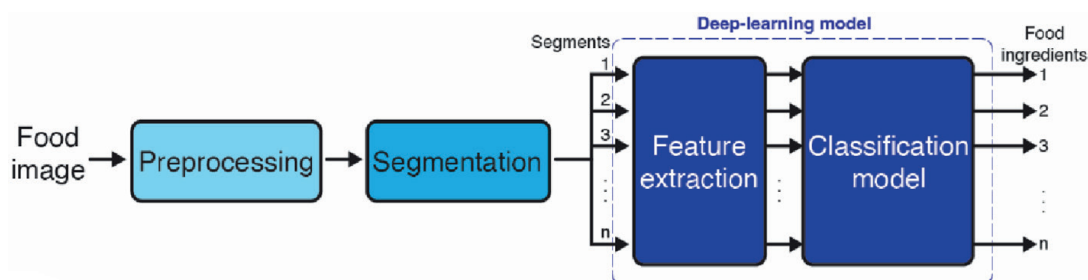


Figure 4.
General flow diagram of food recognition stages.

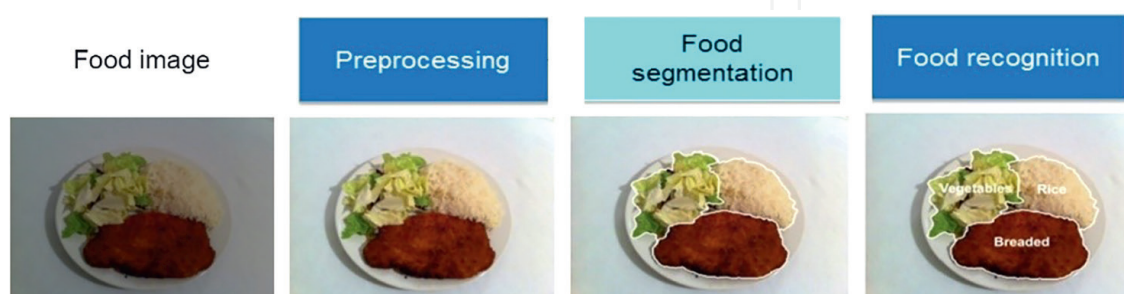


Figure 5.
An illustrative example of food recognition stages [35].

from the image segments, for example, shape, color, texture, and then identifies the food ingredient based on the relevance of extracted features with the features of the labeled training data. Labeling of the ingredients of food dishes is usually done manually, and it is an exhaustive and time-consuming process, especially for multiple labels and large-scale datasets. Feature extraction and model classification can be employed simultaneously by adopting deep learning model architectures [34].

There are many obstacles to overcome while attempting visual food recognition [14]. The great variety in how different foods or even the same foods may look is a huge challenge. This includes differences in color, texture, form, and even presentation. As a result, it is challenging to effectively divide food items from complicated backgrounds and identify individual dishes due to these considerations. The segmentation process is further complicated by occlusions such as utensils, plates, or overlapping ingredients. Dish detection also necessitates that the model would be able to recognize and localize several different food items inside a picture, frequently of variable sizes and configurations. The latter requires reliable item identification methods that can accommodate a wide range of food types. In addition, it is difficult to train effective and generalizable models due to the scarcity of large-scale annotated datasets created for food recognition. **Figure 5** graphically illustrates an example of the food recognition stages [35], including image preprocessing (adjustment of brightness levels), food segmentation, and food recognition.

5.2 Image preprocessing

Preprocessing is the first essential phase of food image identification since it improves the image quality and facilitates further analysis. To enhance the image for better ingredient recognition and categorization, preprocessing employs several methods. During preprocessing, a wide variety of operations, such as scaling,

normalizing, coloring, and noise reduction, are employed. Whenever working with datasets containing photos of varied resolutions, it is crucial that the images are scaled to a uniform size. The image's brightness and contrast can be normalized to remove these differences and facilitate comparisons across photos. It is possible to improve color accuracy and fix color imbalances by using color adjustment procedures. Filtering algorithms and other noise reduction techniques can be used to minimize distracting background noise and pixelation in an image.

The preprocessing goal is twofold. Its primary goal is to enhance the quality of the image, bringing into sharp focus the elements that truly matter. Preprocessing improves the image quality to reduce the effects of factors such as noise, blur, or illumination irregularities that could otherwise hinder precise ingredient detection. To better recognize and categorize individual ingredients, preprocessing seeks to remove any irrelevant or distracting elements from the image. Preprocessing aids analysis by reducing distractions, such as clutter and occlusions, on the food and its constituents.

Preprocessing is crucial because it prepares the image for further analysis with CNNs and other advanced algorithms using scaling, normalizing, color adjustment, and noise reduction techniques. It lays the groundwork for more precise ingredient recognition, segmentation, and classification, which, in turn, boosts efficiency and effectiveness in the field of food picture recognition.

5.3 Food segmentation

In the articles that we looked through, we came across several segmentation approaches that had been utilized for food recognition [10, 19], listed in the following:

- Color-based segmentation

There is an assumption in color-based segmentation that clusters of pixels with similar color attributes represent meaningful objects. One limitation of these approaches is that they may not be able to discriminate between food items that share a color with the plate or background and those that do not.

- Texture-based segmentation.

Separating areas of an image according to their texture patterns is the goal of texture-based segmentation, a method used in image processing. It classifies and segments areas based on texture analysis and machine learning methods.

- Graph-based segmentation

Graph-based segmentation divides images into sections based on pixel similarities. It partitions a graph with nodes representing pixels and edges representing similarity.

- Grid-based segmentation

Grid-based segmentation separates images into grids or cells. Segmentation is easy since each grid or cell is an area. It works well when the image has uniform or regular structures, and you want to divide it into grid-like portions.

- Edge boxes

Edge boxes can provide bounding boxes around objects of interest in a photo during object detection. Edge data locates likely item locations. Edge boxes find probable bounding boundaries for elements in a picture. This strategy narrows the search region, allowing object detection algorithms to focus on productive areas.

- Super pixel-based methods

Supersixel methods fragment images in a more intuitive way. Supersixels preserve image boundaries and structure. Clustering pixels with similar color, texture, and other visual qualities creates a compact image representation.

5.4 Food recognition

5.4.1 Features and dimensionality

The qualities and characteristics that are unique to each of the many food sites contribute to the complexity of the problem of correctly classifying the different kinds of foods. A numerical value that is used to characterize some aspects of the appearance of an image is referred to as a feature or descriptor. In the subject of food picture ingredient detection, strategies for feature extraction play a significant role in the process of gleaning information that is both valuable and identifiable from photographs of raw food [36]. After conducting an exhaustive search of the relevant research literature, a number of feature extraction strategies that are specifically suited to this subject have been uncovered. Included in this package are several color characteristics such as red-green-blue (RGB) [7, 19], hue-saturation-value (HSV), and lightness, as well as the scale-invariant feature transform (SIFT), local binary patterns (LBP) [37], Gabor filters, and CNNs [38, 39] designs, such as ResNet50. Considering that the primary objective of relevant research is to identify the components of food in images, it is essential to further investigate and evaluate a variety of feature extraction strategies in order to successfully capture and depict the typical components of food products.

In what follows, a detailed description of the most popular feature extraction approaches takes place, as compiled from the voluminous scholarly literature. In the field of food image ingredient recognition, these feature extractors have been extensively studied and applied, demonstrating their efficiency in collecting crucial properties and permitting precise analysis of food photographs.

- CNNs are a type of deep learning model developed expressly for the purpose of analyzing images. In order to effectively recognize and classify complicated patterns, they use multiple layers of convolutional filters to learn and extract hierarchical features from images.
- RGB, HSV, and LAB color space features [39]: Information regarding the frequency and range of colors in an image is captured by color-based features. HSV captures the hue, saturation, and value components of an image, while RGB stands for the color channels (red, green, and blue). The letters “L”, “A,” and “B” stand to represent the two complementary colors in the LAB color space. The SIFT transform is a well-known method for extracting features from images that can preserve their structure regardless of transformations such as scaling, rotation, or brightness. Image matching and recognition is a common application of this technology.

- A texture-based feature descriptor known as LBP uses neighboring pixel intensities to characterize local texture patterns. It is widely used in numerous image analysis applications due to the compact representation of texture information it provides.
- Gabor filters are a sort of linear filter that simulates the way cells in the human visual system react to variations in spatial frequency and orientation. By inspecting an image's local frequency content and orientation, these methods excel at capturing texture details.
- ResNet50, a subset of CNN [40], uses residual connections to circumvent the vanishing gradient issue. Amazing success in picture recognition challenges and widespread use in transfer learning for other visual recognition applications are two of its most notable achievements to date.

A wide variety of techniques, such as deep CNNs (DCNNs) learned high-level representations, color characteristics, texture patterns, and local image descriptors, are available for use in feature extraction for food photos. Each approach has its own advantages and can help with identifying and analyzing food ingredients. ResNet (2015), AlexNet (2012), and GoogleNet (2014) are the most popular CNNs utilized for feature extraction.

Food picture recognition requires dimension reduction, especially on low-processing mobile devices. The bag-Of-features (BOF) model [14] reduces feature vector dimensions to improve classification accuracy. Based on codeword frequency, the BOF model displays a picture. Fisher Vector is an effective BOF model modification [10] that encodes patches according to their dissimilarity from a universal Gaussian mixture model. This approach compresses and classifies effectively, even with linear classifiers.

As dimensionality reduction methods, autoencoders and principal components analysis (PCA) have shown promise in food image recognition. PCA uses a linear transformation to locate the most effective orthogonal components to minimize feature space dimensionality. However, autoencoders are neural network models that can compress input data and recreate the whole dataset from this internal representation. PCA and autoencoders can reduce feature vector dimensionality without losing classification information.

Dimensionality reduction methods including the BOF model, Fisher Vector approach, principal component analysis, and autoencoders can increase image-based food recognition accuracy and speed. Reducing feature space dimensionality improves computation speed, classification accuracy, and resource use.

5.4.2 Classification techniques

There are several different categorization approaches that have been investigated in the published research for their potential to accurately recognize and place food items into certain categories. Both the descriptors that were used and the hyperparameters that were selected for the classifiers had a significant impact on the final outcomes of the food image categorization. In addition, for the classification results to be adequate, the quality and variety of the food image datasets that are used to train the algorithms are essential. In order to ensure accurate descriptor selection, hyperparameter optimization, and the utilization of high-quality training datasets, the

designer of an image-based food recognition system (IBFRS) is required to consider the variables. It is the intention of the researchers that by considering these qualities, food picture identification algorithms may be made more accurate and robust, which could then have applications in fields such as dietary assessment, nutritional analysis [41], and personalized meal recommendation.

DCNNs have been established as an effective method for identifying dishes in photographs [7]. DCNNs excel at capturing the rich patterns and textures seen in food photos due to their capacity to automatically acquire hierarchical features from raw pixel data. DCNNs are able to accurately categorize foods thanks to their use of several convolutional layers, pooling layers, and nonlinear activation functions. Large-scale food image datasets are often used to train the network architecture, which then generalizes well to new photos by learning discriminative characteristics.

Multi-relational graph convolutional network (mRGCN) is a method for classifying images of food by utilizing graph convolutional networks [7]. In this method, food photographs are represented as networks, where nodes stand for different parts of the image or different things, and edges capture the connections between them. Improved classification performance is achieved by mRGCN due to the capture of spatial interdependence and contextual information *via* information propagation through the graph structure. This method stands to the task of identifying multi-ingredient dishes with multiple components that interact with one another.

U-Net is a well-liked architecture for analyzing and categorizing food photos [10]. It uses an encoder and a decoder connected by a fully convolutional network. High-level features are extracted by the encoder and segmentation masks or class predictions are created by the decoder from the input pictures. By incorporating fully connected layers or softmax activation at the output, U-Net can be used for classification and improve performance when segmenting food sections of interest in images. This method allows for precise detection and identification of edibles in cluttered settings.

MobileNet is a small, fast, and lightweight convolutional neural network architecture made specifically for handheld and embedded gadgets [12]. To lessen the computational burden without sacrificing accuracy, it employs depth-wise separable convolutions and parameterized point-wise convolutions. By striking a reasonable balance between model size and performance, MobileNet is well-suited for contexts with limited resources. Its small size and speedy operations make it possible for mobile devices with low central processing unit (CPU) power to classify food images in real time.

To encode high-dimensional features into a compact representation, Compact Bilinear Network (CBNet) uses compact bilinear pooling. It uses an outer product operation to combine the strengths of two feature extractors, typically deep convolutional networks, and to capture their interactions. A classifier is then fed with the resulting condensed bilinear features of food images. With improved accuracy and less processing overhead compared to full bilinear models, CBNet stands promising [13].

Support vector machines (SVM) is a common supervised learning method used to categorize food pictures [16]. A high-dimensional feature space is searched until a separation hyperplane between food types is found. Using kernel functions, SVM can process data that is both linearly and non-linearly separable. If you use the right kernels with SVM, it can capture complex decision boundaries and generalize well to photos of foods you have not seen before. When training an SVM classifier, features can be created by hand or taken from a DCNN model that has already been trained [42].

Random forest (RF) is an ensemble learning system that classifies food images by combining numerous decision trees [43]. The final classification choice is reached based on the majority vote of the individual decision trees, which are each trained on a unique subset of the training data. RFs can handle missing values, non-linear relationships, and high-dimensional feature fields. It is well-respected for its sturdiness, interpretability, and tolerance for noisy data. The RF classifier can be trained using a wide range of features from those created by hand to those derived from DCNN models.

6. Discussion

6.1 Research findings

There have been considerable developments in the use of deep learning techniques for food image identification and ingredient detection. In what follows, we will evaluate the results of the studies included in **Table 1**, having investigated various models and datasets for recognizing ingredients and classifying the status of food. CNN, graph convolutional networks, cascaded models, segmentation methods, transfer learning, and attention fusion networks are just few of the methods used in these investigations. The usefulness and potential limitations of various approaches toward enhancing the accuracy and robustness of food picture recognition systems can be better understood by analyzing the performance measures and outcomes of each study. In what follows, research findings on the examined investigations are provided in further depth:

- Zero-shot ingredient recognition by multi-relational graph convolutional network: Using DCNN to identify known ingredients and mRGCN to identify unknown ingredients yields promising results in identifying unseen ingredients. The hit ratios obtained on the VIREO and UEC datasets demonstrate the model's ability to predict unknown ingredients.
- Food state recognition using deep learning: The cascaded multiheaded model outperforms the non-cascaded model in accuracy, precision, recall, and F1 score for food state and ingredient type categorization. Consideration of food state and ingredient-type dependencies increases system performance.
- Ingredient segmentation with transparency: U-Net accuracy varies with transparency. Transparency affects ingredient segmentation as shown by conditions C's better accuracy.
- Deep learning-based ingredient detection recipe recommendation: The CNN model learns patterns from the training dataset based on its high accuracy. However, the validation dataset's slightly lower accuracy signals overfitting, therefore a larger and more diversified dataset would be helpful.
- Deep learning food and ingredient detection mobile app: Modest custom dataset accuracy suggests potential for improvement. The accuracy differences among optimizers indicate the necessity of optimizer selection for better results.
- Novel combinational convolutional neural network for automatic food-ingredient classification: Fine-tuning the CBNNet model on food-41 yields good accuracy.

Fine-tuning the full network improves accuracy, suggesting that fine-tuning the model can improve performance.

- CNN-based food ingredient segmentation: The model can identify specific food ingredients using hierarchical and non-hierarchical segmentation data. However, precision and memory differences among levels show the need for further refining and enhancement.

Many methods and technological advances in food image recognition and ingredient detection have been documented in recent research articles and are considered in this research. CNNs, DCNNs, mRGCNs, U-Nets, and CBNet are just few of the models that have been shown to be useful in these experiments for effectively categorizing and segmenting food constituents. Accuracy, precision, recall, F1 score, and hit ratios are only few of the evaluation criteria that shed light on these models' general efficacy. Dataset collecting, class imbalance, and generalizing models to new components and food states are all areas that need more investigation as this field develops [44]. Using the insights from this study, we can create food picture recognition algorithms that are more precise, effective, and trustworthy for uses including dietary evaluation, individualized recipe recommendations, and promoting good eating habits.

In conclusion, CNNs' exceptional performance and effectiveness in food image recognition and ingredient detection justify their widespread adoption despite their high requirements in terms of training dataset size, hardware specifications, number of parameters, and execution time. The results of this study, along with other notable datasets, such as VIREO Food-172, UEC Food-100, Food-41, and Recipe 1 M, reveal the great potential of CNNs for a few tasks related to food recognition, including ingredient segmentation, feature extraction, and classification. The ability of CNNs to automatically acquire hierarchical representations from raw input data is a key factor in their success since it allows them to detect subtle yet distinguishing elements in food photographs. Researchers have made use of this ability to create complex models that can properly recognize and classify food products and their contents using datasets such as Food-41, which has 4100 photos, and Recipe 1 M, which contains 361,308 images. CBNet model showed remarkable precision results, with accuracy ranging from 88.90 to 95.28% depending on the fine-tuning strategy.

Researchers have been able to evaluate the generalization capacities of CNNs on unseen or unknown elements using datasets, such as VIREO Food-172 (110,241 photos) and UEC Food-100 (14,136 images). Examples are mRGCN and DCNN models, which achieved hit ratios of 47.4 and 48.8% for unseen compounds on VIREO and 24.3 and 42% on UEC. These results demonstrate the extent and promise of CNNs in food recognition, especially when it comes to dealing with unfamiliar components. CNNs have proven to be useful, but it is crucial to recognize the difficulties they can cause. Training CNNs effectively often requires large-scale datasets containing hundreds of thousands of photos, such as Recipe 1 M and FoodAI-756, and significant computer resources. It might sometimes be difficult to comprehend the reasoning behind CNN models due to their limited interpretability. The availability of larger food image datasets, such as NutriNet's 225,953 photos, and the continued development of deep learning algorithms, however, offer hope for overcoming these obstacles. Future applications will need to make extensive use of CNNs; therefore, researchers should keep looking for new ways to maximize their

potential. Maximizing the potential of CNNs for food recognition tasks requires a concentrated effort on methods such as transfer learning, data augmentation, and network optimization, with the aid of amazing datasets, such as VIREO Food-172 and FoodAI-756. Datasets, such as UEC Food-100 and Food-41, can be used to train CNN models that are more adaptable to individual dietary requirements, food allergies, and cultural norms when researchers, industry professionals, and nutritionists work together. New opportunities in fields, such as customized nutrition, dietary evaluation, and smart food logging, can be unlocked by adopting CNNs and overcoming hurdles such as dataset size, hardware requirements, and model interpretability. These innovations, made possible by exceptional datasets, have the potential to radically alter how we monitor and control our dietary intake, with beneficial effects on the health of people and entire communities. There is no doubt that as the area of food recognition develops further, CNNs and these extraordinary datasets will continue to be at the forefront, pushing innovation, and redefining our relationship with food in the digital age.

6.2 Limitations, challenges, and future directions

As in any machine learning problem, a quality dataset is the be-all and end-all of a successful experiment. Many researchers focused on creating new datasets in order to increase the robustness of their work as a common issue that arises is that existing datasets are usually created in a controlled environment or laboratories, which, in turn, trains models to more ideal conditions, but making them unable to perform decently in real conditions. However, it is easy to see how a custom dataset performs poorly due to the unbalanced number of training samples, added noise, possible obstructions, or transparent ingredients, which are sometimes difficult to distinguish.

There has also been limited research done on the ingredients' state recognition, which can prove extremely useful in real-world applications, where freshness and possible staleness play a big role in the quality of the dish. Context is also important as some unknown ingredients could be predicted by region-level recognition. Finally, a large number of papers have used transfer learning on existing models, such as ResNet50 or AlexNet, which are usually trained with the ImageNet dataset, which sometimes forces researchers to modify the training by adding their own labeled images, as a lot of ingredients and food categories are missing from the original dataset.

Future work could also focus on known cooking practices for ingredients as this should greatly assist in recognizing ingredients in different states. The latter, however, would require a great amount of time and experience from the person who is cooking.

7. Conclusions

In this work, a systematic literature review is provided on the most up-to-date methods, datasets, performances, and challenges related to visual recognition of food ingredients. Through the analysis and synthesis of the available literature, this work identifies research gaps, points out the most promising methods, and guides future potential research. Research findings aim to add to the existing body of knowledge and to provide useful insights for researchers, practitioners, and policy-makers interested in applying computer vision and AI to the analysis and nutrition of food.

Acknowledgements

This work was supported by the MPhil program “Advanced Technologies in Informatics and Computers,” hosted by the Department of Computer Science, International Hellenic University, Greece.

Conflict of interest


The authors declare no conflict of interest.

Author details

Michail Marinis, Evangelos Georgakoudis, Eleni Vrochidou
and George A. Papakostas*
MLV Research Group, Department of Computer Science, International Hellenic
University, Kavala, Greece

*Address all correspondence to: gpapak@cs.ihu.gr

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Upadhyay S, Goel G. Food computing research opportunities using AI and ML. In: *Image Based Computing for Food and Health Analytics: Requirements, Challenges, Solutions and Practices* [Internet]. Cham: Springer International Publishing; 2023. pp. 1-23. Available from: https://link.springer.com/10.1007/978-3-031-22959-6_1
- [2] Min W, Wang Z, Liu Y, Luo M, Kang L, Wei X, et al. Large scale visual food recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* [Internet]. 2023;**45**(8):9932-9949. Available from: <https://ieeexplore.ieee.org/document/10019590/>
- [3] Lin Y, Ma J, Wang Q, Sun D-W. Applications of machine learning techniques for enhancing nondestructive food quality and safety detection. *Critical Reviews in Food Science and Nutrition* [Internet]. 2023;**63**(12):1649-1669. Available from: <https://www.tandfonline.com/doi/full/10.1080/10408398.2022.2131725>
- [4] Dai J, Hu X, Li M, Li Y, Du S. The multi-learning for food analyses in computer vision: A survey. *Multimedia Tools and Applications* [Internet]. 2023;**82**(17):25615-25650. Available from: <https://link.springer.com/10.1007/s11042-023-14373-6>
- [5] Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*[Internet]. 2022;**33**(12):6999-7019. Available from: <https://ieeexplore.ieee.org/document/9451544/>
- [6] Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta analyses: The PRISMA statement. *PLoS Medicine*. 2009;**6**(6):e1000097. DOI: 10.1371/journal.pmed1
- [7] Chen J, Pan L, Wei Z, Wang X, Ngo C-W, Chua T-S. Zero-shot ingredient recognition by multi-relational graph convolutional network. *Proceedings of the AAAI Conference on Artificial Intelligence*[Internet]. 2020;**34**(07):10542-10550. Available from: <https://ojs.aaai.org/index.php/AAAI/article/view/6626>
- [8] Chen J, Zhu B, Ngo C-W, Chua T-S, Jiang Y-G. A study of multi-task and region-wise deep learning for food ingredient recognition. *IEEE Transactions on Image Processing* [Internet]. 2021;**30**:1514-1526. Available from: <https://ieeexplore.ieee.org/document/9305995/>
- [9] Alahmari SS, Salem T. Food state recognition using deep learning. *IEEE Access* [Internet]. 2022;**10**:130048-130057. Available from: <https://ieeexplore.ieee.org/document/9982452/>
- [10] Ishichi T, Yamabe T, Tsuji T, Hiramitsu T, Seki H. Ingredient segmentation with transparency. In: *2023 IEEE/SICE International Symposium on System Integration (SII)* [Internet]. New York, NY, USA: IEEE; 2023. pp. 1-5. Available from: <https://ieeexplore.ieee.org/document/10039190/>
- [11] Morol MK, Rokon MSJ, Hasan IB, Saif AM, Khan RH, Das SS. Food recipe recommendation based on ingredients detection using deep learning. In: *Proceedings of the 2nd International Conference on Computing Advancements* [Internet]. New York, NY, USA: ACM; 2022. pp. 191-198. DOI: 10.1145/3542954.3542983

- [12] Christian S, Murwantara IM, Lazarusli I. A Mobile application for food and its ingredients detection using deep learning. In: 2022 1st International Conference on Technology Innovation and its Applications (ICTIIA) [Internet]. New York, NY, USA: IEEE; 2022. pp. 1-6. Available from: <https://ieeexplore.ieee.org/document/9935937/>
- [13] Pan L, Li C, Pouyanfar S, Chen R, Zhou Y. A novel combinational convolutional neural network for automatic food-ingredient classification. *Computers, Materials & Continua* [Internet]. 2020;**62**(2):731-746. Available from: <https://www.techscience.com/cm/v62n2/38273>
- [14] Zhu Z, Dai Y. CNN-based visible ingredient segmentation in food images for food ingredient recognition. In: 2022 12th International Congress on Advanced Applied Informatics (IIAI-AAI) [Internet]. New York, NY, USA: IEEE; 2022. pp. 348-353. Available from: <https://ieeexplore.ieee.org/document/9894627/>
- [15] Pan L, Pouyanfar S, Chen H, Qin J, Chen S-C. DeepFood: Automatic multi-class classification of food ingredients using deep learning. In: 2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC) [Internet]. New York, NY, USA: IEEE; 2017. pp. 181-189. Available from: <http://ieeexplore.ieee.org/document/8181494/>
- [16] Hoashi H, Joutou T, Yanai K. Image recognition of 85 food categories by feature fusion. In: 2010 IEEE International Symposium on Multimedia [Internet]. New York, NY, USA: IEEE; 2010. pp. 296-301. Available from: <http://ieeexplore.ieee.org/document/5693856/>
- [17] Qayyum O, Sah M. IOS Mobile application for food and location image prediction using convolutional neural networks. In: 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS) [Internet]. New York, NY, USA: IEEE; 2018. pp. 1-6. Available from: <https://ieeexplore.ieee.org/document/8629202/>
- [18] Zhang L, Zhao J, Li S, Shi B, Duan L-Y. From market to dish: Multi-ingredient image recognition for personalized recipe recommendation. In: 2019 IEEE International Conference on Multimedia and Expo (ICME) [Internet]. New York, NY, USA: IEEE; 2019. pp. 1252-1257. Available from: <https://ieeexplore.ieee.org/document/8784769/>
- [19] Liu C, Liang Y, Xue Y, Qian X, Fu J. Food and ingredient joint learning for fine-grained recognition. *IEEE Transactions on Circuits and Systems for Video Technology* [Internet]. 2021;**31**(6):2480-2493. Available from: <https://ieeexplore.ieee.org/document/9179998/>
- [20] He H, Kong F, Tan J. DietCam: Multiview food recognition using a multikernel SVM. *IEEE Journal of Biomedical and Health Informatics* [Internet]. 2016;**20**(3):848-855 Available from: <https://ieeexplore.ieee.org/document/7078945/>
- [21] Madival SA, Jawaligi SS. Fine tuned DBN model for food ingredient recognition: Introduction to self-improved Tasmanian devil optimization algorithm. In: 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS) [Internet]. New York, NY, USA: IEEE; 2023. pp. 1-8. Available from: <https://ieeexplore.ieee.org/document/10099841/>
- [22] Zhang M, Tian G, Zhang Y, Liu H. Sequential learning for ingredient recognition from images. *IEEE Transactions on Circuits and Systems for Video Technology* [Internet]. 2023;**33**(5):2162-2175. Available

from: <https://ieeexplore.ieee.org/document/9934942/>

[23] Sahoo D, Hao W, Ke S, Xiongwei W, Le H, Achananuparp P, et al. Food AI. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining [Internet]. New York, NY, USA: ACM; 2019. pp. 2260-2268. Available from: <https://dl.acm.org/doi/10.1145/3292500.3330734>

[24] Mezgec S, Koroušić SB. NutriNet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients* [Internet]. 2017;**9**(7):657. Available from: <http://www.mdpi.com/2072-6643/9/7/657>

[25] Park S-J, Palvanov A, Lee C-H, Jeong N, Cho Y-I, Lee H-J. The development of food image detection and recognition model of Korean food for mobile dietary management. *Nutrition Research Practice* [Internet]. 2019;**13**(6):521. Available from: <https://e-nrp.org/DOIx.php?id=10.4162/nrp.2019.13.6.521>

[26] Cornejo L, Urbano R, Ugarte W. Mobile application for controlling a healthy diet in Peru using image recognition. In: 2021 30th Conference of Open Innovations Association FRUCT [Internet]. New York, NY, USA: IEEE; 2021. pp. 32-41. Available from: <https://ieeexplore.ieee.org/document/9599959/>

[27] He L, Cai Z, Ouyang D, Bai H. Food recognition model based on deep learning and attention mechanism. In: 2022 8th International Conference on Big Data Computing and Communications (BigCom) [Internet]. New York, NY, USA: IEEE; 2022. pp. 331-341. Available from: <https://ieeexplore.ieee.org/document/10064346/>

[28] Chen X, Zhu Y, Zhou H, Diao L, Wang D. ChineseFoodNet: A

large-scale image dataset for Chinese food recognition. arXiv Preprint. 2017. arXiv:1705.02743. DOI: 10.48550/arXiv.1705.02743. Available from: <http://arxiv.org/abs/1705.02743>

[29] Begum N, Hazarika MK. Artificial intelligence in agri-food systems—An introduction. In: Pattnaik PK, Kumar R, Pal S, editors. *Internet of Things and Analytics for Agriculture*, Volume 3. Studies in Big Data, Volume 99. Singapore: Springer; 2022. DOI: 10.1007/978-981-16-6210-2_3

[30] Lee GG, Huang C-W, Chen J-H, Chen S-Y, Chen H-L. AIFood: A large scale food images dataset for ingredient recognition. In: TENCON 2019-2019 IEEE Region 10 Conference (TENCON) [Internet]. New York, NY, USA: IEEE; 2019. pp. 802-805. Available from: <https://ieeexplore.ieee.org/document/8929715/>

[31] Ma P, Lau CP, Yu N, Li A, Liu P, Wang Q, et al. Image-based nutrient estimation for Chinese dishes using deep learning. *Food Research International* [Internet]. 2021;**147**:110437. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0963996921003367>

[32] Gao X, Feng F, Huang H, Mao X-L, Lan T, Chi Z. Food recommendation with graph convolutional network. *Information Sciences (Ny)* [Internet]. 2022;**584**:170-183. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0020025521010549>

[33] Rostami M, Oussalah M, Farrahi V. A novel time-aware food recommender-system based on deep learning and graph clustering. *IEEE Access* [Internet]. 2022;**10**:52508-52524. Available from: <https://ieeexplore.ieee.org/document/9775081/>

[34] Salim NOM, Zeebaree SRM, Sadeeq MAM, Radie AH, Shukur HM,

- Rashid ZN. Study for food recognition system using deep learning. *Journal of Physics: Conference Series* [Internet]. 2021; **1963**(1):012014. Available from: <https://iopscience.iop.org/article/10.1088/1742-6596/1963/1/012014>
- [35] Aslan S, Ciocca G, Mazzini D, Schettini R. Benchmarking algorithms for food localization and semantic segmentation. *International Journal of Machine Learning and Cybernetics* [Internet]. 2020; **11**(12):2827-2847. Available from: <https://link.springer.com/10.1007/s13042-020-01153-z>
- [36] Tahir GA, Loo CK. A comprehensive survey of image-based food recognition and volume estimation methods for dietary assessment. *Healthcare* [Internet]. 2021; **9**(12):1676. Available from: <https://www.mdpi.com/2227-9032/9/12/1676>
- [37] Do T-H, Nguyen D-D-A, Dang H-Q, Nguyen H-N, Pham P-P, Nguyen D-T. 30VNFoods: A dataset for Vietnamese foods recognition. In: 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT) [Internet]. New York, NY, USA: IEEE; 2021. pp. 311-315. Available from: <https://ieeexplore.ieee.org/document/9530774/>
- [38] Dewantara BSB, Devy AZ, Bachtiar MM, Setiawardhana. Recognition of food material and measurement of quality using YOLO and WLD-SVM. In: 2021 International Electronics Symposium (IES) [Internet]. New York, NY, USA: IEEE; 2021. pp. 545-551. Available from: <https://ieeexplore.ieee.org/document/9593949/>
- [39] Jiang S, Min W, Liu L, Luo Z. Multi-scale multi-view deep feature aggregation for food recognition. *IEEE Transactions on Image Processing* [Internet]. 2020; **29**:265-276. Available from: <https://ieeexplore.ieee.org/document/8779586/>
- [40] Yanai K, Kawano Y. Food image recognition using deep convolutional network with pre-training and fine-tuning. In: 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) [Internet]. New York, NY, USA: IEEE; 2015. pp. 1-6. Available from: <https://ieeexplore.ieee.org/document/7169816>
- [41] Mezgec S, Seljak BK. Using deep learning for food and beverage image recognition. In: 2019 IEEE International Conference on Big Data (Big Data) [Internet]. New York, NY, USA: IEEE; 2019. pp. 5149-5151. Available from: <https://ieeexplore.ieee.org/document/9006181/>
- [42] Zhao H, Yap K-H, Chichung KA. Fusion learning using semantics and graph convolutional network for visual food recognition. In: 2021 IEEE Winter Conference on Applications of Computer Vision (WACV) [Internet]. New York, NY, USA: IEEE; 2021. pp. 1710-1719. Available from: <https://ieeexplore.ieee.org/document/9423157/>
- [43] Song G, Guo X, Wang W, Ren Q, Li J, Ma L. A machine learning-based underwater noise classification method. *Applied Acoustics* [Internet]. 2021; **184**:108333. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0003682X21004278>
- [44] Zhu J, Wang Z, Chen J, Chen Y-PP, Jiang Y-G. Balanced contrastive learning for long-tailed visual recognition. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. New York, NY, USA: IEEE; 2022. pp. 6898-6907. Available from: <https://ieeexplore.ieee.org/document/9878764/>