IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 18 April 2023, accepted 3 May 2023, date of publication 8 May 2023, date of current version 15 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3274475

TOPICAL REVIEW

A Study on Food Value Estimation From Images: Taxonomies, Datasets, and Techniques

JAMALIA SULTANA[1], BENZIR MD. AHMED[1,2], MOHAMMAD MEHEDY MASUD[3], A. K. OBIDUL HUQ[4], MOHAMMED EUNUS ALI[1], AND MAHMUDA NAZNIN[1], (Member, IEEE)

[1]Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
[2]Department of Computer Science and Engineering, United International University, Dhaka 1212, Bangladesh
[3]College of IT, United Arab Emirates University, Al Ain, United Arab Emirates
[4]Department of Food Technology and Nutritional Science, Mawlana Bhashani Science and Technology University, Tangail 1902, Bangladesh

Corresponding author: Benzir Md. Ahmed (benzir@cse.uiu.ac.bd)

This work was supported by the Institute for Advanced Research (IAR) at United International University under Grant UIU/IAR/01/2021/SE/11.

ABSTRACT Monitoring nutritional values in food can help an individual in planning a healthy diet. In addition, regular dietary assessment can improve and maintain the physical and mental health of individuals. Recent advancement in computer vision using Deep Learning has enabled researchers to develop various techniques for automatic food nutrition estimation frameworks. Researchers have also contributed to prepare large food image datasets consisting of various food classes for this purpose. However, automatic estimation of nutritional values from food images still remains a challenging task. This review paper critically analyzes and summarizes existing methodologies and datasets used for automated estimation of nutritional value from food images. We first define the taxonomies in order to categorize the existing research works. Then, we study different methods to detect the food value estimation from food images in those categories. We have critically analyzed existing methods and compared the performance of various approaches for estimating food value using conventional performance metrics such as Accuracy, Error Rate, Intersection over Union (IoU), Sensitivity, Specificity, Precision, etc. In particular, we emphasize the current trends and techniques of Deep Learning-based approaches for food value estimation from images. Moreover, we have identified the ongoing challenges associated with automated food estimation systems and outlined the potential future directions. This review can immensely benefit researchers and practitioners, including computer scientists, health practitioners, and nutritionists.

INDEX TERMS Deep learning, food classification, food image, nutrition value estimation.

I. INTRODUCTION

Identifying food values such as carbohydrate (CHO), protein, calorie, etc. are essential for a healthy living. In particular, it is crucial for a person (or a patient) to estimate the calorie intake from the food as overindulgence can lead to various life long diseases such as obesity, diabetes, heart-disease, etc. Automation of estimating food values from food images would be beneficial in maintaining physical and mental health. Recent development of smart phone based applications [1], [2], [3] has made it possible to deploy an efficient real-time automated nutrition estimation framework [4]. The general framework of the food value estimation from food images comprises of identification of the food items in the image, estimation of the volume of the identified food items, and retrieval of the nutritional information of food items, as shown in Figure 1. Moreover, for other smart health applications, food item identification, calorie approximation, etc. from meal images have attracted researchers' attention. The performance of our food value estimation framework depends on the results of the intermediate major steps along with several other factors such as quality and diversity of the food image dataset, relevant information to enhance the performance of the frameworks, etc. However, these tasks are

The associate editor coordinating the review of this manuscript and approving it for publication was Amin Zehtabian.

challenging because of the varieties of food classes, variance of the results due to the impact of color, light, and viewing angles on food images, etc. Therefore, estimating food values from the meal image needs significant research effort.

We observe considerable research activities [5], [6] in this area. In early research works, most of the studies like [4] and [5] have used traditional Machine Learning (ML) methods to calculate the nutritional value from food images. However, from 2014, we have found a shift in utilizing *Deep Learning* (DL) based frameworks [7], [8]. Recently, the researchers are using optimization methods such as *Genetic Algorithm* (GA) [9], *Fuzzy Clustering* for data filtering [10], *Particle Swarm Optimization* (PSO) [9], etc. to improve the Deep Learning based frameworks for food classification. In case of food segmentation, which is a pre-processing step of food item identification, we find that researchers are mostly concerned with segmenting single food item from the serving plate [4], [5]. However, with the improved computational methods, researchers [11] are now involved in segmenting food images from the images of multiple food items. In the volume estimation step, researchers [5], [12] have commonly used reference objects in the images. In recent years, researchers like [13] have computed volume without reference in the images. In this particular work, researchers have used Generative Adversarial Networks (GAN) to map energy distribution in food images. Finally, to estimate the food value, researchers lookup the corresponding nutritional facts from some databases, e.g., US Department of Agriculture (USDA) [14], [15]. Recently, in a few studies [16], the caloric values of food images are crowd sourced. However, this method is highly error-prone. A comprehensive literature review is greatly needed to assist the researchers due to their significant research activities in the area of food value estimation.

There are only a few review papers related to food value estimation methods from image datasets. Min et al. [17] have conducted a study on food computation in 2019. In their review, they have included quite a few things including food dataset acquisition, food perception, food recognition, food data retrieval, food recommendation, and prediction and monitoring of social issues. The food datasets include food images, food relevant texts, and multi-modal data of image and text. In their food recognition part, they have discussed only the food classification methods using mostly *Machine Learning* (ML) based techniques on hand-crafted features of meal images. Subhi et al. [18] have presented a literature review on existing food image datasets, food image segmentation, food item classification, and volume estimation. In the food classification part, they describe feature selection, traditional ML techniques, and Deep Learning techniques. Estimating food value using Deep Learning techniques directly from food images has not been covered in their work. In another review work by Chopra and Purwar [19] in 2022, they have focused their review on different techniques only for image segmentation task. In another work, Dalakleidi et al. [20] have presented different methodologies only for food item recognition. On the other hand, our review looks at the whole workflow of the calorie estimation framework from food images and it includes the major steps needed for food calorie estimation.

The work by Amugongo et al. [21] in 2023 discusses the potential of mobile computer vision-based applications to monitor daily food consumption. Their review has included 22 articles that primarily focuses on recognizing food, estimating volume and calories, and providing dietary recommendations. In another work in the year of 2022 by Konig et al. [22], the authors have focused on smartphone-based dietary assessment tools. However, their review requires to include textual data in addition to food images as inputs to track nutritional intake. Our review includes the food images as the only input data. If using extra texts with food images as input data can be avoided, then the huge cost of data labeling can be saved. Therefore, research using only food images as input data has significant impact in the field of estimating nutrition. Table 1 presents the comparison between the existing review articles and our research in the food computing field for food value estimation from images.

Although existing works have covered some steps of food value estimation, they are not comprehensive that can be observed in Table 1. In addition, none of the previous review articles covered food value estimation directly from image datasets by Deep Learning techniques. To address these gaps, we present our comprehensive literature review on food nutritional value estimation from food image dataset. We include the major steps in the workflow for the nutrition estimation framework along with the description of publicly available food image datasets in our review. The major contributions of our paper are listed below.

1) We have conducted a comprehensive literature review on food nutritional value estimation directly from the food image dataset. This also includes the estimation of food value directly from the image dataset using Deep Learning techniques.
2) We have categorized the reviewed studies based on the steps needed for developing an automated food value estimation framework that includes *food item classification*, *volume estimation*, and *nutrition estimation*.
3) We observe the trends and scientific development in applying different Deep Learning techniques for designing frameworks for estimating food nutrition.
4) We have analyzed the relationship between the frequently used traditional Machine Learning based methods for classifying food items, and the extracted handcrafted features from meal images.
5) We have presented some research challenges and opportunities for future work in this domain.

We organize the rest of the paper as follows. We provide an overview of the review: the methodology of conducting the review and an overview of the food nutrition estimation system in Section II. In Section III, we discuss different types of food image inputs that are used for nutrition estimation frameworks. We narrate the food classification

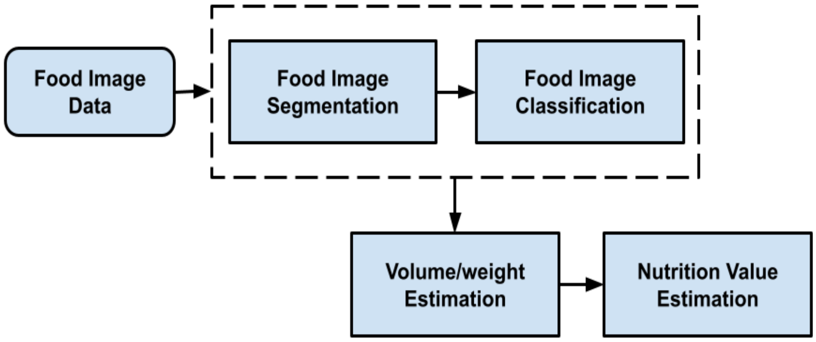

FIGURE 1. A generalized framework for food nutrition estimation from food images.

TABLE 1. Differences between the existing review articles and this article.

Criteria	Min et al. [17]	Subhi et al. [18]	Chopra et al. [19]	Dalakleidi et al. [20]	Amugongo et al. [21]	Konig et al. [22]	This Study
Study year range	2009-2018	2009-2018	1972-2019	2016-2021	2010-2022	1960-2020	2011-2023
No. of Articles in Review	300	NA	66	78	22	117	79
Image Segmentation	✗	✓	✓	✓	✗	✗	✓
Image Classification	✓	✓	✗	✓	✓	✓	✓
Volume Estimation	✗	✓	✗	✗	✓	✓	✓
Nutrition Estimation	✗	✗	✗	✗	✓	✗	✓
Dataset Acquisition	✓	✗	✗	✓	✗	✓	✓
ML methods, features mapping	✗	✗	✗	✗	✗	✗	✓
Traditional ML features analysis	✗	✗	✗	✓	✗	✗	✓
Method perform. analysis	✓	✓	✓	✓	✓	✗	✓
Direct Image-based Food Value Estimations	✗	✗	✗	✗	✗	✗	✓

frameworks along with different applied methodologies in Section IV. We illustrate different approaches that are being used for estimating volume or weight from food image data in Section V. Section VI describes different processes that are used for estimating nutritional values from the image datasets. In Section VII, we summarize the findings of this paper and present the current challenges and potential future research works. Finally, in Section VIII, we draw the conclusion of this review work.

II. OVERVIEW OF THE REVIEW

In recent years, the study on food nutrition estimation has become popular. The primary goal of our review is to study different methods and frameworks that have been used to estimate the nutrition values from food images. In addition to this, we analyze different food datasets used in these studies to understand the mapping between the input data and the frameworks. In this section, we describe the methodology of our reviewing process. We briefly discuss the major components of a standard food nutrition system as well.

A. SCOPE OF THE REVIEW

We present a *food analysis system* that uses computational techniques to automatically compute nutrition values from input food data in this section. For many people who keep track of their diet, it is crucial to track the approximate nutritional content of foods. With the rise in obesity and malnutrition-related disorders, concerned researchers are interested in automation of the estimation of food nutrition estimation using food images as input. In this work, we mainly focus on these studies. Estimating nutrition from an image may necessitate a number of intermediate steps. The first step of food nutrition estimation is to segment the food items in the image and then to classify *food items* of the target image. Researchers then determine the amount (or volume) of the food items in the image to find the nutritional values. Hence, we broadly categorize the reviewed articles into three major groups: *food image classification*, *volume/weight estimation*, and *nutrition estimation*, shown in Figure 2. A brief description of these major groups is given as follows.

1) FOOD IMAGE CLASSIFICATION

In *food image classification*, the researchers use images of foods to classify the types of foods or food items. This kind of food classification can be conducted on multi-food item meal type [15], [23] or single food item meal type [15]. For both food image types, *food image segmentation* is used before classification. For a single food item in an image, food image segmentation methods divide the input images into food and non-food data points. Most of our observed studies use *object segmentation algorithms*

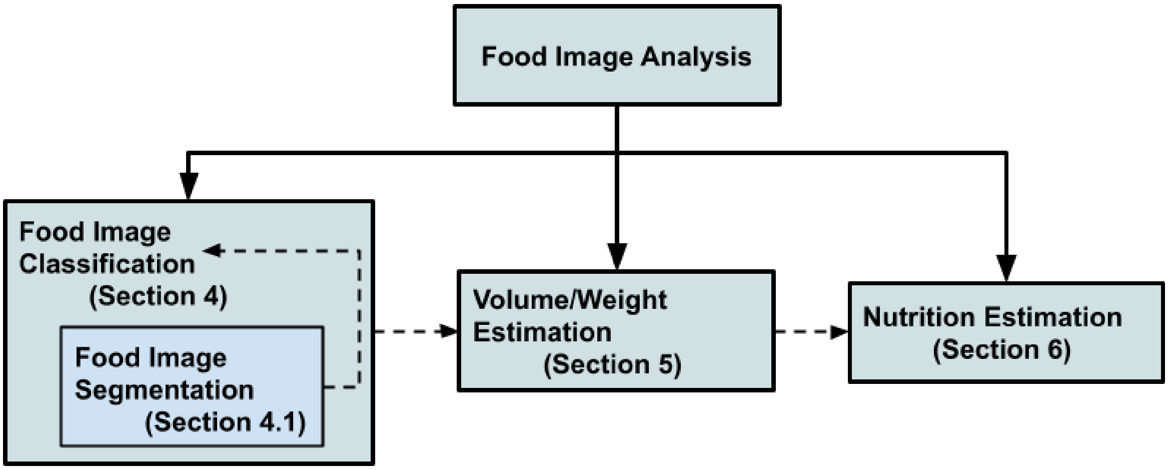

FIGURE 2. A taxonomy of the food image analysis application domain.

for segmenting multiple food items or separating single food items from non-food items [24], [25]. After image segmentation, classification of segmented food items is performed. In this study, we explore different types of ML techniques, including K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Support Vector Regression (SVR), etc., and Deep Learning techniques for food classification. For traditional ML approaches, researchers have extracted different features from the food image data to train the models. Various Deep Learning models use raw food image data as input for food classification. Deep Learning frameworks [26], [27], [28] are used for both food image segmentation and image classification processes. We discuss food image classification techniques in Section IV.

2) VOLUME OR WEIGHT ESTIMATION

Computation of an approximate volume or weight of the food or food item is also one of the precursor steps in food value estimation. In general, the volume estimation step is performed after the food classification step. However, for liquid foods, some studies [29], [30] conduct volume estimation from food images without classifying the food types.

Volume estimation approaches use different food image data collection methods to compute the volume or weight of the food. Few researchers have used special cameras known as *depth cameras* to capture the 3D images of the foods. In some methods, researchers have reconstructed 3D food images from the top & side views of the same food image. In some cases, reference objects such as, thumbs, credit cards, forearms, etc. are placed in the food image so that the researchers can calculate the volume or weight of the food. However, in some studies, researchers have used Deep Learning techniques to estimate volume from 2D food image data points. More detailed discussion on volume or weight estimation techniques is presented in Section V.

3) NUTRITION ESTIMATION

Estimating nutrition from food images is an interesting research field that encompasses other research fields including food item classification, volume or weight estimation, calorie computation, etc. After classifying the food items and estimating the volume of the food items, researchers apply the predetermined nutritional values of the food classes. These nutritional values are determined by the experts of the field, e.g., USDA, or other resources. Different reviewed papers evaluate nutrition in different ways. Some studies show the range of caloric value of the detected food instead of giving an approximate value. Some studies focus more on calculating the value of carbohydrates present in the food image. Majority of the studies focus on computing caloric values from the food images. We discuss nutrition estimation techniques from food images in Section VI.

4) PERFORMANCE METRICS

In our review for food classification approaches, we find that the most of the studies have used the metric, *accuracy* (*acc*). Accuracy is one of the evaluation metrics used as performance measure in classification models. We know that, *accuracy* metric in the model returns the numerical fraction of correctly predicted objects. We see the mathematical form of accuracy in Equation 1, where the number of correct predictions includes both true positives and true negatives.

$$acc = \frac{No.\_of\_Correct\_Predictions}{Total\_No.\_of\_Predictions} \tag{1}$$

We also notice the utilization of different other metrics to measure the performance of the proposed methods for food segmentation, food volume estimation, and calorie estimation. Some of these performance metrics are:

Error Rate: This performance metric refers to a measure of the degree of the prediction error of a model made with respect to the true model. Equation 2 presents the mathematical formula of this performance metric. Here,

TABLE 2. List of full forms and their abbreviations (only the ones that are not mentioned in the text).

Full Forms	Abbreviation
Scalable Color Descriptor	SCD
Color Structure Descriptor	CSD
Dominant Color Descriptor	DCD
Color Layout Descriptor	CLD
Gradient Orientation Spatial Dependence Matrix	GoSDM
Entropy-Based Categorization and Fractal Dimension Estimation	EFD
Gabor Based Image Decomposition and Fractal Dimension Estimation	GFD
Locality-constrained Linear Ccoding	LLC
Region-based Convolutional Neural Networks	RCNN
Deep Neural Networks	DNN
Speeded Up Robust Features	SURF
Bag of Features	BoF
Segmentation-based Fractal Texture Analysis	SFTA
Multidimensional Scaling Fitness	MDSFIT
Hue, Saturation, and Brightness	HSB
Not Available	NA

FP and *FN* mean False Positive and False Negative, respectively. *P* means all the positive samples in the model, and *N* means all the negative samples in the model. The summation of *P* and *N* notes the true model, and the summation of *FP* and *FN* means the prediction error of the model.

$$ErrorRate = \frac{FP + FN}{P + N} \tag{2}$$

Intersection over Union (IoU): This performance metric evaluates the object segmentation methods by estimating the percentage of intersection between the predicted image mask and the actual image [31]. Equation 3 presents the mathematical formula for the IoU metric where *TP*, *FN* and *FP* are True Positive, False Negative and False Positive, respectively. The summation of *TP*, *FN* and *FP* means the Area of Union and only *TP* means the Area of Overlap.

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

Sensitivity: This metric is known as true positive rate and measures the proportion of the positive instances a model is able to identify correctly. Equation 4 shows the mathematical formula of sensitivity. Here, *TP* and *FN* are True positive and False Negative, respectively.

$$Sens = \frac{TP}{TP + FN} \tag{4}$$

Specificity: This performance metric gives the numerical fraction of the True Negatives that are correctly predicted by the model. Equation 5 presents the mathematical formula where *TN* and *FP* mean True Negative and False positive, respectively.

$$Spec = \frac{TN}{TN + FP} \tag{5}$$

Precision: This metric calculates the ratio of correctly identified positive samples to the total number of identified positive samples. Equation 6 shows the formula for the precision in machine learning models. Here, *TP* signifies True Positive and *FP* is noted for False Positive. The summation of *TP* and *FP* is the total number of identified positive samples.

$$Prec = \frac{TP}{TP + FP} \tag{6}$$

B. REVIEW METHODOLOGY

An extensive search has been conducted across multiple databases including Google Scholar, ResearchGate, and PubMed to collect published research papers in the field of food image processing and analysis, and calorie estimation from food images. We have explored the papers published from the year of 2011 to 2023 for our comprehensive study. All of the selected papers are written in English and peer-reviewed in high impact journals and conferences. Our review works encompass all types of modeling techniques with various handcrafted extracted features used for nutrition estimation frameworks. In this paper, we provide a comparative analysis of the changing trend in the field of food image processing and calorie measurement within the last eleven (11) years. Our analysis entails the feature extraction methods, classification approaches, calorie estimation frameworks, and performance metrics used for evaluations. The keywords that we have used for our exploration are - 1) *food image segmentation*, 2) *food image classification*, 3) *volume estimation from food images*, and 4) *calorie estimation from food images*.

We have found a total of 465 peer-reviewed papers after our initial search on web-based Google Scholar (221), ResearchGate (182), and PubMed (62). After removing the duplicate articles, we are left with 387 articles. We have then screened the titles and abstracts of these papers and excluded the articles based on the following criteria: 1) studies on food calorie analysis application domain using multi-modal food datasets, for example, some studies have utilized text information from recipes along with the food images as inputs [32], and 2) out-of-the scope of this study, for example, some studies are about food classification and segmentation to detect diseased areas of the food [33], leaving only

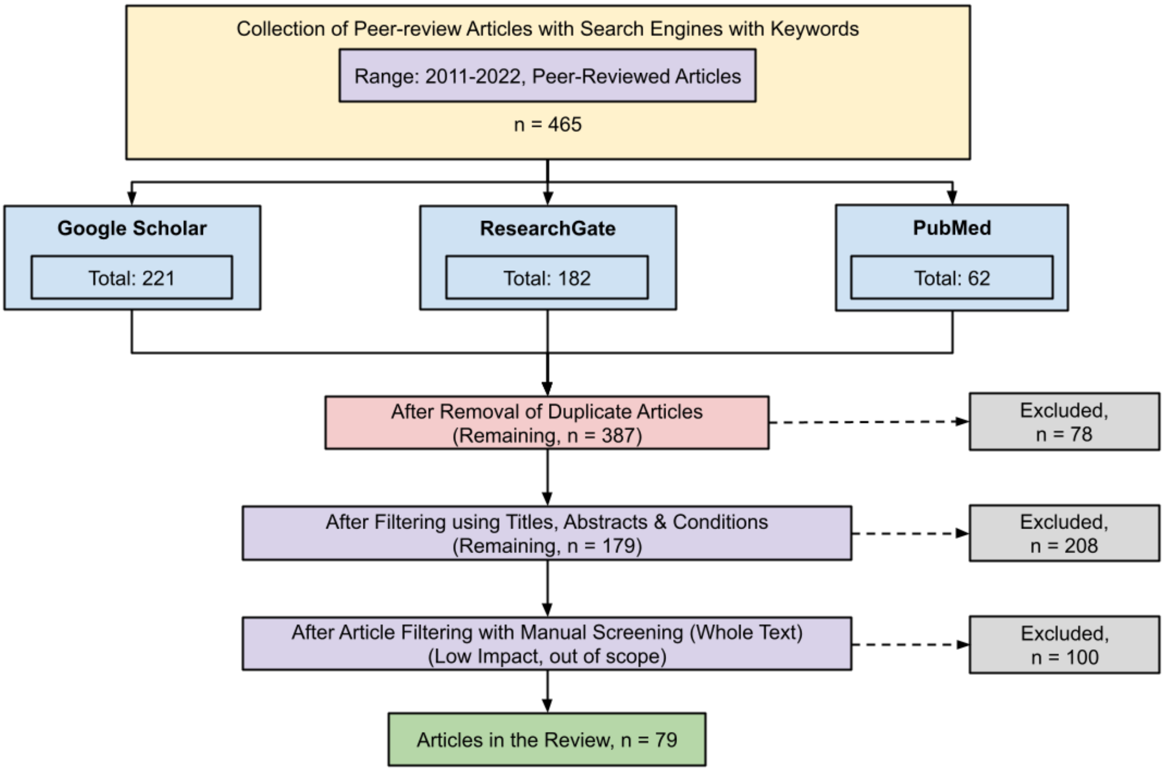

FIGURE 3. Pipeline for our literature review.

179 articles. After following our rigorous full text assessment, we have included 79 (seventy-nine) papers in our review study. Our review methodology is given in Figure 3. We have provided the categorization of our reviewed articles on food nutrition frameworks in Figure 2 in Section II-A.

III. INPUT DATA FOR FOOD IMAGE ANALYSIS

Researchers have used meal images, text description of meals or both (Multimodal) as input data to estimate nutrition of food. In some studies [34], for better performance results researchers have used text data along with the food images. However, our review is limited to input data with food images only. Most of the works that we have studied, do construct a unique food image dataset for their own experimentation. However, there are also some benchmark food image datasets covering different geographic regions with different food classes. Training any food nutrition estimation system requires an extensive dataset containing food images of multiple classes. A standard publicly available dataset can significantly help researchers building different classifiers and compare their results. Several large benchmark food image datasets are publicly available and are summarized in Table 3.

Early datasets [37] have smaller numbers of food images than the recent ones [34], [46]. The datasets representing specific cuisines, e.g., Turkish food image dataset [3] also have a small number of food images. The food image datasets such as [34], [35], and [46] that are mostly created from mixed cuisines e.g., English, Italian, Japanese, Korean, Indian, etc. have a large number of food images and food classes. Datasets with a large number of food images such as *ChineseFoodNet* [8], *Instagram800k* [46], *Food-500* [45] are acquired from scraping the social media or websites e.g., search engines. Some researchers use mobile apps to create datasets. For example, Bossard et al. [35] use the "Foodspotting" mobile app to create the *ETHZ Food-101* benchmark dataset with 101000 images and 101 classes, Xu et al. [42] used the "Dianping" app to create the *Dishes* dataset with 117504 food images and 3832 classes. Some researchers have used previously available datasets to create a new benchmark dataset for training such as Food524DB [49], Food-24 [3], and MaFood-121 [54]. It is apparent that most of the food datasets are curated for a specific task. For example, Food-24 Dataset [3] is made of Turkish cuisine; ChineseFoodNet [8], BTBUFood-60 [55], Dishes [42] are made of Chinese cuisine; UNIMIB-2016 [44] is made of Italian Cuisine; ETHZ Food–101 [35] and UNICT-FD889 [37] are made of a mixture of Eastern (Korean, Japanese) and Western (Italian) food items. Fruits-360 [52] and FruitVeg [51] are built of food items from fruits and vegetables. Some datasets are created for specific purposes such as, Food201-Segmentation [39] dataset is created for food image segmentation purpose and has 12625 food images and 201 classes.

Some researchers [5], [60] have used non-food and easily accessible reference objects in the food image so that they

TABLE 3. Existing benchmark datasets used in the reviewed articles.

Dataset	No. of Classes	No. of Images	Image Collection Method
ETHZ Food-101(2014) [35]	101	101000	Foodspotting app
Diabetes(2014) [36]	11	4868	Web
UNICT-FD889(2014) [37]	889	3583	Smartphone
UEC-FOOD 100 (2015) [38]	100	9060	Web, Manual
Food201-Segmented(2015) [39]	201	12625	Manual
UPMC Food-101(2015) [40]	101	90840	Web (Google)
FooDD(2015) [41]	23	3000	Camera
Dishes(2015) [42]	3832	117504	Dianping app
Menu-Match(2015) [14]	41	646	Social media
UNIMIB-2015(2015) [43]	15	2000	Smartphone
UNIMIB-2016(2016) [44]	73	1027	Smartphone
Food-975(2016) [11]	975	37785	Camera+Yelp
Food-500(2016) [45]	508	148,408	Web+Social media
Instagram800K(2016) [46]	43	808,964	Instagram
Food-11(2016) [47]	50	5000	Social media
UEC-FOOD 256 [16](2016)	256	25088	Web: Google, Bing, Twitter, etc.
UNICT-FD1200(2016) [48]	1200	4754	Camera
Food524DB (2017) [49]	524	247636	Prior Datasets
VegFru(2017) [50]	292	160731	Web
FruitVeg-81(2017) [51]	81	15630	Mobile
ChineseFoodNet(2017) [8]	208	192000	Web
Cookpad(2017) [34]	NA	4748044	Cookpad app
Fruits 360(2018) [52]	103	71125	Camera
Food-24(2019) [3]	24	10825	Prior datasets (Food4 [35], Food15 [53]), image downloaded with Fatkun Batch plugin
MAFood-121(2019) [54]	121	21175	Prior Datasets+Web
BTBUFood-60(2019) [55]	NA	52495	Baidu+Google
Sushi-50(2019) [56]	50	3,963	Images downloaded from google using sushi guide
ISIA Food-200(2019) [57]	200	197,323	Wikipedia listing, Search engines (Google, Bing, Baidu)
ISIA Food-500(2020) [58]	500	399,726	Web scraping from search engines (Google, Bing, Baidu)
Food2K(2021) [59]	2,000	1,036,564	catering website: Meituan

can later use that object to estimate the dimension, volume, or weight of the food. Therefore, based on the presence of the reference objects, we can divide the input food images into two sub-groups. These sub-groups are: 1) only food images 2) food images with non-food reference objects. Some researchers use the thumb or index finger on the edge of the plate alongside the foods in the food images [15]. Use of thumb or index finger for the food images comes with its own limitations as well. For example, finger size varies from person to person. Some researchers use *credit cards* [60] and 3cm X 3cm *card boxes* [23].

A. ACQUISITION OF FOOD IMAGE DATA

Some studies have built their food image datasets from scratch. Some other studies use pre-existing benchmark datasets for their experiments. The reviewed articles have used different methods to collect or create datasets such as using pre-existing standard food image datasets, using in-house built apps to collect data from users or using *web scraping* to build dataset [16]. The five sources that are used by the reviewed articles for collecting the datasets are given in Table 4.

From Table 4, we can see that use of apps in smartphone devices is a common and widely used method for collecting food images from users. Researchers can capture food images and upload them to the storage system built in the IoT devices. In these methods, researchers can also control the environment in which the food images are collected. Another widely used method for collecting data is web scraping. This method is used for creating a large dataset cheaply for the cuisine of any nationality. With search engines such as Google, and Baidu [55], the researchers can accumulate a large amount of food images for their datasets. Websites are scraped using keywords like the name of the foods.

People tend to share their meal images with food names as tags on social networks. Some researchers have used social media such as Yelp [11], Instagram [46], etc. to collect food image data from users. Some [34] collected data from cooking websites. The additional information such as the food ingredients and the volume or calorie amount from the cooking websites may help researchers to improve the performance of the food nutrition system. Some benchmark datasets like [39] are created by the researchers by capturing the food images in a controlled lab environment. These datasets are generally smaller in size but they have more accurate information about the input images.

B. INPUT FEATURES USED BY FOOD ANALYSIS MODELING TECHNIQUES

The set of features to be extracted from meal images depend on the ML technique. Two main categories of machine learning techniques are Traditional ML techniques

TABLE 4. Acquisition methods of food images.

Methods	Description	Ref
Applications in smartphones	Collects the food images by using the apps in IoT devices	[51], [42], [34], [48], [37], [41], [43], [44], [52]
Social Media	Collects the shared food images from the users of social media	[46], [47], [45], [11], [14]
Recipe Websites	Food image data is collected from specific websites used for sharing recipes	[34], [42]
Web Scraping	Creates large scaled datasets cheaply by scraping the websites and search engines	[16], [38], [16], [54], [55], [8], [50], [45], [40], [36]
Controlled Environment	Dataset is created manually by capturing food images in a controlled lab environment by researchers	[37], [38], [48], [39]

and Deep Learning techniques. In traditional ML methods, researchers handcraft their features from the input data. The performances of the frameworks built on traditional ML techniques largely depend on these carefully selected features. Features extracted from the image data can provide valuable information for fine-tuning an ML model. Researchers select the features based on the goal of their experiments. Few popular features that are usually extracted from food images are discussed below.

Color: Color feature is used in [4], [5], [28], [38], [41], [61], [62], [63], [64], and [65]. It is a very important and intuitive feature in food images. The color of the foods seen from the images can be used to classify the food images.

Scale-Invariant Feature Transform (SIFT): Another popular and derived feature from the food images is SIFT [66]. SIFT, a computer vision algorithm, is used to detect and match local SIFT features in images. SIFT works by extracting key points from the reference food image sets and storing these extracted points in a database. A food class is identified from a new food image in two steps. First, we compare each feature in the new image to the previously constructed image database. Then, we identify the candidate matching features based on the Euclidean distance of their feature vectors. SIFT is used by several researchers [23], [36], [38], [48], [60], [61], [67], [68]. A variation of SIFT is Colored SIFT (CSIFT) which is extracted from an RGB color space. CSIFT is presented as a robust feature against illumination changes. CSIFT is used by [38] and [69]. In the study of Matsuda et al. [38], the authors have used all the features: color, SIFT, and CSIFT that preserve the color of the target food.

Texture: In food image classification and segmentation, *Texture* feature used in [5], [38], and [41] plays a crucial function in visual perception and can be considered as one of the fundamental features of natural images of different food classes. Since the 1950s, texture has been one of the most active research topics in machine intelligence and pattern analysis. Texture is used to discriminate between different patterns of images. It extracts the dependency of intensity between the pixels and their neighboring pixels [70] or obtains the intensity variance across pixels [71]. We have observed that researchers prefer applying Gabor Texture Filter instead of extracting texture from the food image [5], [38], [41], [61]. A Gabor texture feature depicts texture patterns of local regions at various scales and orientations. Histogram is a feature that extracts the texture pattern from the food images [25]. Histogram of Oriented Gradients (HOG) [72] is a feature descriptor used in image processing and computer vision to recognize objects. HOG keeps rough location data by constructing histograms for each dense grid and concatenating them into a single long feature vector [38], [64]. RootHOG is inspired by "RootSIFT" and is an element-wise square root of the L1 normalized HOG [73]. It is shown in the studies that RootHOG leads to better performance than original HOG [4]. Local Binary Pattern (LBP) is one of the methods to extract texture features of the foods [28]. Pairwise Rotation Invariant Co-occurrence LBP descriptor (PRICoLBP) is primarily concerned with encoding spatial co-occurrences and pairwise orientations of well-known LBP features [74]. It maintains the relative orientations of LBP feature pairs to provide rotational invariance [48].

Size & Shape: Extracting the sizes and shapes of food from the image is vital for estimating the calorie of the observed foods from images. Some studies have placed standard accessories to correctly measure the approximate size of the foods from the food images [5], [41]. One example is shown in Figure 4.

FIGURE 4. Food images with reference cards.

We find other additional features such as super-pixels [75], [76], Visual Words [77], Bag of Features [23], [63], etc. extracted from food image data in different studies.

These features can preserve multiple visual descriptors in one feature value, such as *Super-Pixels* [75] extracted from the food images. A super-pixel [76] is a small region formed by splitting an image based on edge and local features, and there are no boundaries among different image objects. These local features can be consistent with both color and texture extracted from the food images. It is possible that different patterns of food classes are present inside the same super-pixel. Another high-level feature that can contain multiple visual descriptors is *Visual Words*, and researchers use Visual Words for retrieving image information [60]. Visual words [77] represent small sections of a food image that include information about the characteristics (such as color, shape, or texture) or changes in the pixels (such as filtering, low-level feature descriptors). The *bag of features* method used in [23] and [63] represents images with orderless collections of local features. Each image is abstracted by numerous local patches after feature extraction. Methods for representing patches as numerical vectors are dealt with in feature representation approaches. These numerical vectors are called *feature descriptors*. To some extent, a decent descriptor should be able to handle intensity, rotation, scale, and affine variations. SIFT is one of the most well-known descriptors. Each patch is converted to a 128-dimensional vector by SIFT. Following this phase, each image is a collection of vectors of the same dimension (128 for SIFT), with no regard for the order of the vectors. The *Fisher Kernel* is a function that compares the similarity of two items using a statistical model and the basis sets of measurements for each object [4]. In a classification framework, the class of a new object (whose true class is unknown) can be estimated by minimizing the average of the Fisher Kernel distance between the new object and the object classes.

IV. FOOD CLASSIFICATION

One of the fundamental steps in food nutrition estimation is to classify the food images. In the food classification step, researchers train their ML models with labeled food images and predict the food classes of the food items of a test image using the trained models. In the initial food classification approaches, most of the studies work on identifying one single food item from the containers by using food image data points. However, in natural settings, it is very common to have multiple food items in one container. It is also difficult to train a machine with image data points consisting of multiple food classes. Therefore, researchers have used image segmentation as a significant first step for classifying food types from images with multiple food items. Image segmentation can also be conducted for segmenting food from non-food such as visual separation between the actual food and the container of the food. After the segmentation of food items, they are classified as the final step. The accuracy for the segmentation and food classification depends on the training of the models with a large dataset with standard food images.

A. FOOD IMAGE SEGMENTATION

Food image segmentation means separating the food items in the same container by using visual features. The methods applied for the food segmentation can be grouped into two categories depending on how the input food image data have been handled. These are: 1) Application of Non-Machine Learning (Non-ML) methods with handcrafted extracted features from food images, and 2) Application of Machine Learning (ML) methods. In the first category, some studies use region detection and separation techniques such as GraphCut [15], GrabCut [27], etc. as their food image segmentation method. These food image segmentation techniques handcraft the extracted image features according to their applied methods. Hence, these Non-ML techniques have difficulties in generalizing the food segmentation process. In recent years, more studies are using ML techniques for image segmentation. So far, most of the Machine Learning based methods are presenting better performance in food image segmentation than Non-ML based approaches. Different food image segmentation techniques that are used in the reviewed articles are given in Table 5.

Some of the studies [38], [39], [78] have conducted image segmentation on food images solely for image segmentation purpose. Most of the studies [4], [5], [48] use food segmentation as an intermediate step of the food classification process. The performance of the food image segmentation can contribute to the cascading errors in food value estimation. Similarly, the performance of the food classification models can be improved with a better performing image segmentation step, as shown in the experiments conducted by the studies in [49] and [62]. However, food image segmentation is a challenging task as many foods have irregular features such as irregular shapes and edges, non-uniform contours, etc. [61]. Food image segmentation can be more difficult when multiple food items are mixed together or placed on top of another resulting in occlusion [79]. From Table 5, we see that the *GrabCut* algorithm [80] is the most used method in the food image segmentation process as used in the studies conducted in [4], [5], [78], and [81]. In general, most of the image segmentation processes work using graph segmentation processes. In this approach, the whole image is represented as a graph. Then, a set of pixels are used to create a super-pixel and this is considered a node or vertex of the graph. These nodes are connected to their neighboring nodes with an edge creating an adjacency relationship in the graph. Then the problem is to find an optimal cut in the edge set that separates the graph into dissimilar sets of nodes and group the similar nodes into one class. In another way, the image segmentation process works by clustering the pixels. For the clustering task, one or more features, such as SIFT, pixel colors, etc., of the food images are used.

According to our review, the earliest research work on image segmentation is conducted by Chae et al. [82], who have used image segmentation in their study to estimate the food volume from food images. Their proposed mathematical model extracts feature points to determine the dimension of the food shape templates and reconstructs the 3D properties of the food shape from a single image. Utilizing this template-based approach, this system segments the image and estimates food volume. Kawano and Yanai [62] have presented a system where the users need to draw a bounding box manually on the food image to select the food area. The food area is then extracted using the *GrabCut* algorithm. The accuracy here is determined by the ability of users to draw accurate boundary boxes. Fang et al. [78] have presented a semi-automatic framework for segmentation where the users draw a bounding box around the food and tag the food properly from the available food list. After these steps, the framework would segment the food from the image using GrabCut technique. The manual drawing of the bounding box is addressed in [81]. Shimoda et al. [81] have proposed a framework where they generate a bounding box using CNN and Distinct Class-specific Saliency Maps (DCSM). The bounded area is then segmented by GrabCut. In another study conducted by Pouladzadeh et al. [5], they have used Graph cut segmentation in their experiment and have the highest overall food classification accuracy of 95%. In [5], the image segmentation using GrabCut is automated by separating the graph representation of food images into two different dissimilar groups by considering the weights on the edge of the adjacent vertices.

Matsuda et al. [38] use a circle detection, Felzenszwalb's Deformable Part Model (DPM), and JSEG region segmentation [93] for food image segmentation. Although this study has a low food classification accuracy of 21%, it shows that with only the DPM model, the overall food classification accuracy can be increased. In [49], the authors have also used JSEG region segmentation along with color, saturation, and noise removal. They, however, manually segmented the tray images by drawing polygonal boundaries in the image. The segmentation provides better precision compared to other methods. The work conducted by He et al. [24] has used local variation segmentation algorithms and created a feedback loop for segmentation refinement. They have achieved a better classification accuracy than the normalized cut approach. Kong and Tan [85] has used a perspective distance algorithm for the three image views of the same food and then clustered the features of each one. Then they segment the food image based on the clustered features. For one food-item images, this method has achieved the highest food classification accuracy of 100% and for five food item images, this method has achieved an accuracy of 76%. Sadeq et al. [87] have used K(=3)-means clustering for their food image segmentation. They have demonstrated that food segmentation using clustering decreases the standard error rate for some foods.

Yarlagadda et al. [75] have introduced the concept of superpixel. Their proposed unsupervised method finds the salient missing objects between a pair of food images taken before and after eating. Their goal is to design a class agnostic food segmentation method. They have utilized the after eaten image as the background to calculate the contrast of each pixel with the before eaten image. The contrast and saliency maps were then combined to produce the final segmentation mask of the salient missing objects in the previously eaten image. We can segment the food by recognizing salient objects in the previously consumed image since before eaten food images have salient objects. In recent years, we have seen the use of Deep Learning methods in food image segmentation [81], [88]. In [88], Freitas et al. have used Deep Learning techniques including Fully Convolutional Network (FCN), SegNet, Efficient Neural Net (ENet), DeepLabV3+ and Mask RCNN. They have shown that most of these segmentation methods perform well when the image contains only one food class. For multiple food items in the same image, FCNs outperform other methods.

In the recent year 2022, Generative Adversarial Network (GAN) is used for food image segmentation [91] and has obtained an accuracy of 95.21% in calorie estimation on the "UNIMIB 2016" food dataset. Similar high performance can be seen from the Mask R-CNN based food image segmentation frameworks [92]. Aditama and Munir [92] have used Mask R-CNN based food segmentation framework for 6 food classes in their experiments. They have also included the ResNetXt-101-FPN to aid their framework for better performance. In 2022, Aguilar et al. [90] have proposed to add Bayesian network with the DeepLabV3+ and have achieved a mean IoU of 0.81 for three publicly available datasets: UNIMIB2016, UECFOODPIXComplete, and Food-201. The study conducted by Honbu and Yanai [89] has an accuracy of 90% by using Few-shot and Zero-shot segmentation for the unseen food classes.

One of the limitations of the food segmentation frameworks is the scarcity of annotated food datasets for segmentation. In some previous research studies [38], [82], scientists have utilized food shape templates for segmentation. This technique is not applicable to amorphous-shaped food [23], [87]. Researchers in [13] and [87] have used algorithms such as Canny edge detection to identify the edges of the food shapes for amorphous-shaped food segmentation. In recent times, researchers are utilizing DL frameworks such as Mask R-CNN, SegNet, CNN, DeepLabV3+, and so on for food segmentation but the issue of insufficient food annotation is still prevalent [27].

B. FOOD IMAGE CLASSIFICATION MODELING TECHNIQUES

After food items segmentation, the next step in the food value estimation framework is food classification. Detecting food classes from the target food image is a challenging task. Same food can visually look different along with rotation, occlusion, low resolution, etc. Moreover, differences in food preparation can result in different color, shape, and texture. Training the models with large datasets of multiple food classes is essential to obtain high food classification accuracy. The food classification frameworks used by the researchers can be categorized into two different groups based on the

TABLE 5. Image segmentation methods in the reviewed literature. Majority articles do not report segmentation performances, they rather report the overall food classification accuracy. Dataset column: total number of images followed by number of food classes in the parentheses. Weight Estimation Error (WER).

Ref	Year	Dataset	Method	Food Classification Accuracy
[82]	2011	NA	Mathematical model, food template shape	Average Relative Error: 11.00%
[38]	2012	In-house	DPM, a circle detector & JSEG region segmentation	21.00%
[60]	2012	In-house	Histogram, subtraction based mechanism	92.00%, Volume estimation error: 20.00%
[24]	2013	1453(96)+143(1)	Local variation segmentation, feedback for segmentation refinement	WER: 10.00% compared to 11.00% for normalized cuts
[62]	2013	UEC-Food256	Bounding box, GrabCut	81.60%
[5]	2014	[83]	Graph cut segmentation	95.00%
[84]	2014	In-house: 83	Multiple segmentation hypotheses with segment confidence scores	74.00%
[25]	2014	2500	Semantic segmentation - STFT, GMIM	Nutrition prediction error: 31.80%
[39]	2015	Food-101	Deep Lab Model	76.00%, Recall: 0.33, IoU: 0.25
[81]	2015	UEC-FOOD100	DCSM, Bounding box using CNNs-> GrabCut	49.90%
[85]	2015	PFID	Perspective distance algorithm, cluster segmentation	100% for 1 food-item image, 76.00% for 5 food-item image
[23]	2015	FooDD [41]	Canny edge detector for plate detection, Food segmentation	88.50%
[41]	2015	FooDD [41]	Graph Cut, Color-Texture based Segmentation, DNN,	76.00% for mix-food
[86]	2016	In-house	GoCARB	75.40%
[49]	2017	UNIMI NB2016	A combination of color, saturation, JSEG, and noise removal	Precision: 96.56%
[78]	2018	PFID+In-house, 60000	Manually drawn bounding box, GrabCut	NA
[87]	2018	FooDD	K(=3)-mean clustering, Otsu threshold, Canny Edge	Calorie estimation error ranges from 0.78% to 12.75%
[88]	2020	1250(9)	FCN, SegNet, ENet, DeepLabV3+, and Mask RCNN algorithms	IoU: 0.70, Sensitivity: 0.81, Specificity: 0.99
[75]	2021	566 pairs of before and after eaten food images	Superpixels feature extraction, Pixel based segmentation, saliency fusion, and contrast map generation	Weighted harmonic mean: 0.741, AUC: 0.954
[27]	2021	In-house, 256178	GrabCut	75.00%, RMSE: 9.40
[89]	2022	UEC-Food256, Food-101	Few-shot and Zero-shot segmentation	90.00% (Food-101, Zero-shot), 91.11% (UEC-Food256, Few-shot)
[90]	2022	UNIMIB2016, UECFOODPIXComplete, Food-201	Bayesian based DeepLabv3+, GourmetNet, MC-Dropout	99.00%, IoU: 0.81
[91]	2022	UNIMIB 2016	GAN	Calorie Estimation Acc: 95.21%
[92]	2022	In-house: 1646(6)	Mask R-CNN, ResNeXt-101-FPN	IoU: 0.90, F1 score: 0.994, Precision: 91.47%

modeling techniques used: 1) Traditional ML methods, and 2) Deep Learning methods. In traditional ML techniques, the selection of features from the food images plays a very significant role in the performance of the systems. In DL models, architecture plays an important part. For both types of classification techniques, the importance of the quality and the quantity of food image data available for training the models is significant.

1) TRADITIONAL ML METHODS

In our observed studies, the traditional ML techniques used in food classification are given in Table 6. Kong et al. [60] have extracted SIFT features and K-mean clustering of Visual words from their in-house food image datasets. These features are then applied to the KNN algorithm for training. They have achieved a high food classification accuracy of 92%. Their in-house dataset consists of 5 food classes collected using smartphone cameras and web scraping. There is only one food item in each of these training images. In [38], the researchers have used SIFT and one of its variants CSIFT, HOG, Gabor texture, and color in their extracted feature set. A framework of Multiple Kernel Learning-SVM (MKL-SVM) is used for classifying food types. They have created an in-house food image dataset for their study. The accuracy of their model is only 21%. As mentioned before, their main contribution is the successful use of the DPM model for segmentation. In [94], they have applied multiple features as visual descriptors both individually and together in the food classification framework. Their model has achieved an accuracy of 53% and 46% for SIFT and LBP features, respectively. However, they got an accuracy of 68.3% when they used these two features together along with the color features and Gabor texture. Thus, this study shows that combining visual descriptors in the traditional framework can increase the performance of the classification framework. Similarly, in [14], the researchers have used a combination of SIFT, LBP, color, HoG, and MR8 Filter as the features. They have achieved a food classification accuracy of 77.4%. Their results show that different extracted features of the same visual descriptors can increase the overall classification accuracy. In the food classification experiment by Beijbom et al. [14], they have developed a SVM food classifier. For their in-house dataset Menu Match, they have achieved an accuracy of 51.2% and for the dataset in [94], they have achieved an accuracy of 77.4%.

Table 7 lists the features that are used with the various ML techniques in the reviewed papers. We observe that the SIFT, Color, Gabor filter, Histogram, and LBP are the most popular features. All the features except fisher vector are used with SVM. All other ML techniques have their preferred feature set such as DCSM [26], super pixel [75], Visual Words (K-mean clustering) [60], RootHOG [4], PRICoLBP [48], recursive Bayesian estimation [96], etc.

We observe from Table 6 and Table 7 that SVM models or variants of SVM model, such as SVR are the most widely used models. Among the reviewed papers that have used traditional ML methods, half of them have used SVM models for food image classification. Among the rest of the papers, about half of them have used derivatives of SVM models such as, Radius-margin-based SVM with LogDet regularization (L-SVM) [36], MKL-SVM [38], etc. Among the research works using SVM models, Pouladzadeh et al. [5] have achieved the highest accuracy of 95%. They have used GraphCut for food image segmentation and have used visual descriptors like color, size, shape, and texture to identify the food classes. They have further experimented with food classification and developed a cloud-based SVM model [41]. They have extracted Gabor texture and color from the food images and then trained their cloud-based SVM to identify the food classes. They have achieved an accuracy of 94.5%, which matches the performance in their previous study. The study by Anthimopoulos et al. [36] has created a *Bag of Features* using color and SIFT features. This *Bag of Features* is then used for training the L-SVM classification model to identify the food among 11 food classes and achieved an accuracy of 78%. Chen et al. [94] have proposed a multi-class SVM with AdaBoost to classify the food from 50 food classes. They have extracted SIFT, LBP, color, histograms, and Gabor Texture from the food images to train the model and achieved an accuracy of 68.3%. However, when they have deployed the SVM model without the AdaBoost, they have received a lower accuracy of 62.7%. Another research work by Kong et al. [67] has used multi-class SVM models on SIFT and Gaussian Region Detection as the image features from the PFID dataset. They have achieved an accuracy of 84% in the extended dataset of PFID. Sudo et al. [25] have proposed an SVR model and applied histogram, SIFT and GMIM as the features for training the model. In the study presented by Zhu et al. [84], for the same dataset with the same extracted features, KNN algorithm outperforms the SVM model by 13%. In this study, their KNN model and SVM model have achieved accuracies of 70% and 57%, respectively.

We have noticed that most researchers have used SVM [35], [41] or a variation of SVM [36] as the ML techniques and Color, Texture, SIFT, and Histogram as the input features for food classification. The selection of extracted features has a considerable impact on the performance of the food classification models. These features have to be selected manually and they can also be dataset-specific. Thus, a large amount of time needs to be dedicated for identifying the correct features for training the models. Also, for poorly selected features, the traditional ML models may not perform adequately for large food classes. These are limitations of the traditional ML models. On the other hand, Deep Learning techniques can extract generalized contextual information from the image data without extracting features manually. Deep Learning technique eliminates the need for manual feature selection and user intervention for food classification. Therefore, Deep Learning methods may be more suitable for a fully automated food nutrition system from food images than traditional ML techniques.

2) DEEP LEARNING METHODS

Deep Learning (DL) is a sub-field of ML methods. These models are based on Artificial Neural Networks (ANN) and representation learning. In Deep Learning approaches, the researchers do not need to construct hand-made customized feature sets to identify the food classes as these approaches are built to extract features from the food images directly. Deep Learning can utilize structured, unstructured or in-between data for training. Since our review is limited to food nutrition framework using food image data, we only considered the Deep Learning methods that take images as input. In our investigation between 2011 and 2023 time periods, we observe a rise in using Deep Learning methods from 2014 for food identification and segmentation due to their exceptional classification capability compared to traditional ML methods. Convolutional Neural Network (CNN) is a widely preferred method in computer vision applications, including image classification, because of its ability to extract contextual information and classify large amounts of visual data. The reviewed articles given in Table 8 also used different variations of established CNN architectures to classify food images. We have observed that Alexnet, a variation of CNN architecture, is the most used DL technique for food image classification.

It is observed that Deep Learning methods such as CNN outperform traditional ML methods in the benchmark datasets like Food-101, UEC256, etc by large margin.

TABLE 6. Food classification based on traditional ML techniques. Dataset column: total number of food images followed by number of food classes in the parentheses.

Ref	Year	Dataset	Features	Method	Accuracy
[67]	2011	PFID+In-house	SIFT, Gaussian Region Detection	Multi-class SVM	84.00%
[61]	2011	In-house	Color, SIFT, Predominant Color divided into Local and Global features, Gabor, Entropy, Tamura, Haar Wavelet, Steerable, DAISY	SVM	86.10%
[60]	2012	In-house	SIFT, Visual words	KNN	92.00%
[38]	2012	In-house	Color, SIFT, CSIFT, Gabor Texture, HOG	MKL-SVM	21.00%
[94]	2012	In-house 5000 (50)	SIFT, Local Binary Pattern, color, histograms, Gabor Texture	SVM, Multi-class AdaBoost	68.30%
[12]	2012	Before-after eaten	RGB, HSV, YCbCr	SVM, nutrient database	NA
[95]	2013	In-house (250+)	SIFT, Gabor	K-means for clustering	60.70%
[24]	2013	1453(96)+143(1)	SCD, CSD, DCD, CLD, GOSDM, EFD, GFD	4 food classes, majority vote of nearest neighbors	34.00%
[62]	2013	In-house	SURF, color	SVM	81.60%
[63]	2014	In-house	BoF, SFTA, color	SVM	70.00%
[68]	2014	In-house	SIFT, DCD, MDSFIT, SCD	KNN	64.50%
[4]	2014	UEC-Food256	Color, RootHOG, Fisher Vector	One vs Rest ML Algorithm	50.01%
[25]	2014	In-house (2500)	Histogram, STFT, GMIM	SVR	NA
[5]	2014	[83]	GraphCut, color, size, shape, texture	SVM	95.00%
[84]	2014	3000(19)	SIFT, color, texture	KNN	70.00%
[84]	2014	3000(19)	SIFT, color, texture	SVM	57.00%
[14]	2015	Menu Match	SIFT, LBP, Color, HoG, MR8 Filter, LLC, K-means clustering, K(=3)NN	SVM	51.20%
[14]	2015	5000 (50) [94]	SIFT, LBP, color, HoG, MR8 Filter, LLC, K-means clustering, K(=3)NN	SVM	77.40%
[36]	2015	In-house	SIFT, color, Bag of features	L-SVM	78.00%
[41]	2015	In-house	Gabor, color	Cloud based SVM	94.50%
[28]	2015	In-house [23]	LBP, color	SVM	82.20%
[64]	2015	UEC-FOOD100+UEC-FOOD256	HOG, color	Fisher Vector	52.90%
[48]	2016	UNICT-FD1200	SIFT, PRICoLBP, Bag of textons (visual words distributions)	SVM	75.74%

TABLE 7. Modeling techniques and features mapping.

Techniques	Image Features						
	Color	Texture	SIFT	Histogram	Gabor Filter	Fisher Vector	LBP
SVM	[41], [5], [94], [28], [63], [62], [61], [67]	[41], [5]	[36], [38], [94], [37], [61]	[38], [5], [94]	[38], [94], [61]		[23], [94], [28]
KNN			[60], [68]				
One vs Rest	[4]			[4]		[4]	
L-SVM			[36]				
K-Means			[95]		[95]		
MKL-SVM			[38]	[38]	[38]		
Fisher Vector	[4], [64]			[4], [64]			

Kagaya et al. [97] used an in-house dataset and applied CNN model to their food classification framework. They have obtained an accuracy of 73.70%. Studies such as [7], [35], [39], [64], [98], and [99] have used a benchmark food image dataset *Food-101* in their food classification experiments. Among these studies, we find that Tan and Le [7] have achieved the highest accuracy of 93%. They have achieved this accuracy by implementing EfficientNet for food classification. Bossard et al. [35] have implemented a CNN food classification framework based on the ImageNet architecture and they have achieved an accuracy of 56.4% after 450000 iterations on the ETHZ Food-101 dataset. In [100], researchers used the Inception V3 architecture on the ETHZ Food-101 dataset and achieved an accuracy of 88.3%. Inception V3 is a CNN architecture by Google and part of the Inception architecture family. They also have applied the Inception V3 model on the UEC-FOOD100 and UEC-FOOD256 datasets, and achieved an accuracy of

81.45% and 76.17%, respectively. This study also proves that a model can achieve high accuracy by fine-tuning a model based on the dataset. Liu et al. [101] have proposed a Deep Convolutional Neural Network (DCNN) named DeepFood which is a variation of Inception CNN architecture. Their model has achieved accuracies of 77.4%, 76.3%, and 54.7% for ETHZ Food–101, UEC-FOOD–100, and UEC-FOOD–256 datasets, respectively.

We observe that some studies have used both traditional ML and Deep Learning techniques in the same experiments. The researchers first use the DL techniques to extract the contextual information from the image data instead of handcrafting the feature set. Then they utilize the extracted features for training the traditional ML techniques. In [102], the researchers have used CNN for image feature extraction and then fed the extracted features to train the SVM classification model. In [4], they have proposed a framework that uses DCNN to extract features from food images and classifies the food images by using Fisher Vector. They have achieved an accuracy of 72.3% on the UECFood-101 dataset. Akhi et al. [108] also have implemented the same framework. However, they have used pre-trained CNN architecture for feature extraction from food images. They have achieved accuracy of 99.13% and 95.79% for Bar-Food101 and PFID datasets, respectively. Thus, we can see that researchers can achieve good performance by deploying a framework built on both traditional ML and Deep Learning techniques.

In recent years, researchers are using transfer learning, which is basically off-the-shelf DCNN models such as, AlexNet [64], GoogleNet [39], [47], [98], and Inception V3 [3], more instead of building or training a DL model from scratch. It takes a lot of food image data to train a Deep Learning framework. Transfer learning can use prior knowledge of the domain. Therefore, researchers can obtain better performance without training the model with a large image dataset. Yanai et al. [64] have used a pre-trained DCNN with 1000 ImageNet food categories. They have fine-tuned their model by training it on 3 different food datasets: Food–101, UEC-FOOD–100 and UEC-FOOD–256. They have achieved accuracies of 78.77%, 67.57%, and 70.4% for the UEC-FOOD100, UEC-FOOD256, and ETHZ Food–101 datasets, respectively. They have proved that fine-tuning the pre-trained DL models can improve classification accuracy. In [39], researchers have used a pre-trained DCNN model, GoogleNet, and fine-tuned the model on the ETHZ Food-101 dataset and achieved an accuracy of 79%. We also observe that for the same food dataset, transfer learning with GoogleNet performs better (79%) [39] than transfer learning with Alexnet (70.41%) [64]. However, we observe that the Food-101 dataset gives better results when AlexNet is used for transfer learning [64] instead of training the AlexNet from scratch [35]. Since Alexnet is a Deep Learning model with many parameters, it requires a large amount of data to train from scratch. Therefore, better performance in Alexnet with transfer learning is perceivable.

Apart from utilizing transfer learning with pre-trained DCNN models, researchers have also exploited the effectiveness of ensembling various DCNNs. Pandey et al. [106] have ensembled three DCNN architectures: AlexNet, GoogleNet, and ResNet. They have achieved an accuracy of 72.12% for ETHZ Food-101 dataset and an accuracy of 73.5% for their in-house dataset. In recent years, especially in 2022, we have observed some studies conducted by [9] and [115] where the researchers attempted to increase the efficiency of their frameworks by utilizing optimization techniques such as Particle Swarm Optimization, Genetic Algorithm, Bayesian Fuzzy Clustering, etc. In both 2022 and 2023, we still observe the utilization of deep CNN-based frameworks such as DCNN [118], transfer learning CNN [116], ResNet50 [117], MobileNet V2 [119], deep CNN-based Progressive Region Enhancement Network [59], etc. for recognizing different food classes. These CNN-based food classification frameworks produce high performance (> 90%) in identifying food classes. This means researchers can investigate the relation and similarity among the good performing deep CNN based frameworks. The findings can then guide us toward enhancing the performance of estimating the volume and calories from food images. However, for the food nutrition estimation system, we also need to estimate the volume or weight of food from food images, which is discussed in the next section.

V. VOLUME OR WEIGHT ESTIMATION

Once the food has been segmented and classified, the researchers need to compute the volume or weight from food images to estimate the nutritional values. However, automated estimation of the food volume or weights from image data is a difficult task. Most of the food images are constructed in two dimensions. Two dimensional images do not have real life information such as volume, size, or portion of the foods that are used to estimate food value. Hence, as observed from Table 9, researchers use different approaches such as 3-D food images, shape templates, multiple-view food images, etc. to estimate volume or weight of the food items from the image data.

A. 3-D FOOD IMAGES

Since it is difficult to extract relevant information related to volume or weight of the food from 2D food images, many researchers have opted for using 3D food images to calculate the food volume from the image. Few researchers [39], [94] have used a special depth camera to capture the 3D composition of the food images. Depth camera acts as a 3D camera and is able to judge the width, height, area, volume, etc. when the object is placed within the frame. The 3D images from these cameras enable the researchers to estimate the volume of the food images. Similar results can also be achieved by attaching a laser device to the smartphone camera to estimate the volume of the food [121]. Although these methods have achieved promising results, in real life scenarios, the additional device can limit the user experience.

TABLE 8. Food classification based on Deep Learning techniques. In-house means collected by the research team. Dataset column: total number of images followed by the number of food classes in the parentheses. mean Average Precision (mAP).

Ref	Year	Dataset	Method	Accuracy(%)
[97]	2014	In-house (NA)	CNN	73.70
[35]	2014	Food-101	AlexNet	56.40
[36]	2014	Diabetes	ANNnh	75.00
[4]	2014	UECFood–100	DCNN + Fisher Vector	72.30
[64]	2015	UECFood–100	DCNN-Food	78.80
[102]	2015	FoodLog	CNN, SVM	NA
[64]	2015	UECFood–256	DCNN-Food	67.60
[64]	2015	Food-101	Transfer Learning (AlexNet)	70.41
[39]	2015	Food-101	Transfer Learning (GoogleNet)	79.00
[28]	2015	In-house (NA)	Patch-wise CNN	84.90
[100]	2016	Food–101	Inception V3	88.30
[103]	2016	VIREO	Arch-D	82.10
[101]	2016	Food-101	LeNet-5, AlexNet and Pre-trained GoogleNet	77.40
[103]	2016	UEC-Food–100	Arch-D	82.10
[26]	2016	UECFOOD-100	DCSM, VGG-16	Average Precision: 30
[47]	2016	Food-5K, Food-11, IFD	Pre-trained GoogleNet	83.60
[104]	2016	In-house (NA)	Deep NN	99.00
[100]	2016	UECFood–100	Inception V3	81.50
[101]	2016	UECFood–256	DeepFood	54.70
[100]	2016	UECFood–256	Inception V3	76.20
[105]	2017	In-house (NA)	CNN	86.72
[98]	2017	UEC256	Transfer Learning (GoogleNet)	76.30
[106]	2017	Food-101	Ensemble DCNN Net (GoogleNet, AlexNet, ResNet)	72.12
[44]	2017	UNIMI NB2016	VGG	78.30
[106]	2017	In-house (NA)	Ensemble DCNN Net (GoogleNet, AlexNet, ResNet)	73.50
[98]	2018	Food–101	DCNN + edge computing	77.00
[107]	2018	Food-101+100	Inception-v3, Inception-v4	85.00
[65]	2018	In-house: 5800(20), Food-101	CNN	92.23
[108]	2018	BarFood101, PFID	Pre-trained CNN + SVM	99.13 (BarFood101), 95.79 (PFID)
[87]	2018	FooDD	CNN	NA
[109]	2018	In-house: 1400	DenseNet169, ResNet50	80.60
[110]	2018	Food-101	Wide-Slice ResNet	90.27
[99]	2018	Food-101	Deep layer Aggregation	90.00
[56]	2019	Food-101	PAR-Net	90.40
[7]	2019	Food-101	EfficientNet	93.00
[54]	2019	MAFood-121	Regularized Uncertainty based Multi-Task Learning (RUMTL) Model	83.82
[3]	2019	Food-24	Transfer Learning (InceptionV3)	88.00

TABLE 8. *(Continued.)* **Food classification based on Deep Learning techniques. In-house means collected by the research team. Dataset column: total number of images followed by the number of food classes in the parentheses. mean Average Precision (mAP).**

[111]	2019	ETH Food-101, VireoFood-172, ChineseFoodNet	MSMVFA (VGG, ResNet, DenseNet)	90.59
[6]	2020	UEC-FOOD100, UEC-FOOD256, FOOD101	Deep-CNN, RPN derived from Faster R-CNN	Accuracy: 71.70, Average Precision: 25.50 (UEC-FOOD100), 18.30 (UEC-FOOD256)
[112]	2020	CF-108, 100800(108)	Depth-wise separable convolution Mask R-DSCNN	Average Precision: 0.782
[88]	2020	In-house (NA)	FCN, ENet, SegNet, DeepLabV3+, and Mask RCNN	IoU: 0.70
[113]	2021	In-house (crowdsourced): 15,908(281)	CNN, Calorie Mama API	75.10
[114]	2021	23000 (23)	Inception V3, CNN	NA
[27]	2021	In-house (256178)	YOLOv4-tiny	80.90
[9]	2022	UEC FOOD-100	CNN Optimized with PSO and GA	82.30
[115]	2022	In-house: 390 (8)	RCNN	98.90
[116]	2022	In-house: 8520	Transfer Learning with pre-trained CNN models	93.59
[117]	2022	In-house: 2400(6)	CNN with 4 C-Layers, ResNet50	98.67 (CNN), 96.67 (ResNet50)
[118]	2022	In-house	CNN, OpenCV	95.30
[119]	2022	EgocentricFood [120], UECFood-256, In-house: 1800(4)	MobileNetV2	mAP: 0.90, F1-score: 0.86
[59]	2023	In-house: 1M (2000)	CNN based Deep Progressive Region Enhancement Network	98.74

B. FOOD SPECIFIC SHAPE TEMPLATES

He et al. [24] have estimated the food volume from the 2D food image by reconstructing the 3D food image by using a food-specific shape template. This technique works comparatively better for beverage food items. Because beverage containers are usually of cylindrical shape, and by using a cylinder shape template 3D images can be constructed. Similarly, Chae et al. [82] have also reconstructed 3D food images from the input 2D images by using shape specific templates. They have used a shape template for bread and a different shape template for drinks. Their study shows that by using food specific shape templates, the overall relative error for volume estimation is 11% for 17 drinks and 8% for bread slices. Another study [125] that also used a shape-based approach has collected a total number of 100 food samples of Western and Asian cuisine using a wearable camera. Using the automated method, they found that 85 food items out of 100 have less than 30% error. In [123], the researchers have used the shapes from silhouettes for food portion size estimation reconstructing multi-view 3D food images. They have achieved a mean error of 10% on a dataset with 4 food classes for multi-view volume estimation and a mean error of 17.9% on a dataset with 19 food classes for weight estimation. Although this method can easily estimate the relatively accurate volume of the foods from the 2D food images, this technique will not work for foods with irregular shapes or foods whose shape depends on the food preparation process [94].

C. MULTIPLE-VIEW FOOD IMAGES

Few studies [5], [12], [126] use side and top views of food images to estimate the food volume and weight. Dehais et al. [126] have proposed to reconstruct a 3D food image from 2D input data to estimate food volume from

TABLE 9. Volume or weight estimation performance comparison. In-house means collected by the research team. Dataset column: total number of images followed by number of food classes in the parentheses. Volume Estimation Error (VEE), Weight Estimation Error (WEE), Energy Estimation Error (EEE), Standard Deviation of Error (SDE).

Ref	Year	Dataset	Methods	Performance Description
[82]	2011	NA	Shape templates	VEE: 11% for 17 drinks, 8% for bread slices
[121]	2011	NA	Depth-capture laser device with smartphone	NA
[94]	2012	5000(50)	Depth camera	Preliminary Results, performance is NA
[12]	2012	Before-after eaten	Top & side images + index finger as reference	Preliminary Results, performance is NA, Error in the acceptable range, results vary with illumination and viewing angle
[122]	2012	6 fruit classes	Stereo food images	VEE: 7.7%
[60]	2012	In-house	3 images with a credit card or a short 120 degree video	SDE ±20%
[123]	2013	ETHZ Food-101	Shapes from silhouettes	VEE: 10% for 4 food classes, WEE: 17.9% for 19 food classes
[30]	2013	ETHZ Food-101	Pre-trained 3D model, Food information (shapes & orientation)	VEE: 10% for 5 food classes
[24]	2013	1453(96)+143(1)	Food (beverage)-specific shape template	VEE: 11%
[5]	2014	[83]	Top & side images, thumb as reference	VEE: 1% - 10% for 5 food classes
[25]	2014	2500	Histogram	Segmentation Error in Table 5, Food Classification accuracy in Table 6 and Calorie Estimation Error in Table 10
[39]	2015	ETHZ Food-101	Depth camera, CNN, and RANSAC	NA
[124]	2015	330 (19)	Reference object & container shape	EEE < 6%
[14]	2015	Menu Match	SVM, SIFT, HOG, LBP, Color, Gabor, MR8	Acc. 57.7%
[23]	2015	FooDD	2 image views with reference card	Segmentation Error in Table 5 and Calorie Estimation Error in Table 10
[125]	2015	100	Shape-based approach	VEE < 30% for 85% cases
[126]	2017	Meals-45, Angles-13, Plates-18, Meals-14	Two view images	MAPE: 8.2% - 9.8% (Two datasets with 45 dishes and 14 meals)
[87]	2018	FooDD	Forearm distance from food	NA
[65]	2018	5800(11), ETHZ Food-101	Food front edge detection, height & depth estimation, Stereo image analysis	VEE: 8.5% (4 food classes)
[13]	2018	PFID+In-house (60000)	GAN	EEE < 10.89%
[29]	2019	60	Smartphone motion sensor	VEE 16.65% for 10 food classes
[115]	2022	In-house (390(8))	Top and side views with coin as reference, pixel method, mask-based RCNN, pre-trained ResNet	Precision for food shapes: Amorphous (90.46%), Convex (90.9%), Regular square (98.5%), Regular circle (98.9%)

four different datasets: Meals-45, Angles-13, Plates-18, and Meals-14. In two distinct datasets, they attained a Mean Absolute Percentage Error (MAPE) ranging from 8.2% to 9.8% for 45 dishes in the 1st dataset and 14 dishes in the 2nd dataset.

D. REFERENCE OBJECTS

To estimate the food volume from the image data, it is vital to know additional information, such as the scale and rotation of the food in the image. The volume of the foods can be closely estimated if these additional parameters can be perceived. To extract this relevant information from the food images, researchers have placed reference objects with known size and scale in the images [124], [127]. These reference objects can be any object with known size and scale. In [127], the researchers have used a standard plate and container as the reference objects with a mean error of 3.41% for the 2 dimensions: length and width. In some studies, researchers use the user's index finger [12] or thumb [5] in the top view or both top view and side view to construct a 3D image from which the food volume of the target can be calculated. Some researchers have used reference cards to build a 3D Food Image for shape and size estimation of the food [60]. Villalobos et al. [12] have used the index finger as the reference object and captured top and side views of food images with the reference placed in the image. In the study conducted by Pouladzadeh et al. [5], they have similarly used the thumb as the reference object and captured top and side views of food images. Their study shows that the volume estimation errors range between 10% in the worst case and 1% in the best case for a non-mixed food dataset with five classes. Some studies have used reference cards with known size and scale [23], [60] to reconstruct 3D food images from 2D images. Sadeq et al. [87] use the user's forearm as reference length for food volume estimation of the images. This technique has achieved low standard error for some of the food classes. For food classes with high irregularity, such as apple, mango, etc., this method gives low performance. In some recent frameworks for volume estimation systems in 2022, such as the one proposed by Kadam et al. [115], the method of utilizing reference objects like coins is still prevalent. Kadam et al. [115] have employed a fixed-dimension coin for volume estimation in their framework. This coin provides a Pixel per Metric (PPM) ratio that is utilized to determine the height and diameter of the container. However, their assumption that the volume of amorphous food items is equivalent to the volume of the container is often not the case. In many instances, there may be discrepancies in height and width between the actual volume of the amorphous food and the container.

E. MOTION SENSOR, CROWDSOURCING

Alternatively, Yang et al. [29] have proposed a fiducial-marker free technique that uses smartphone motion sensor data to detect camera orientation for volume estimate from 2D food images. Their volume estimation framework has achieved an absolute error of 16.65% for 10 food classes. In [14] and [128], the researchers have opted for crowd sourcing their food volume and the nutritional information where individual users evaluate the foods. These kinds of approaches are not automated and produce very error prone results. Therefore, these methods are not suitable for any food nutrition estimation system.

F. STEREO FOOD IMAGES

Subhi et al. [65] have proposed front edge detection of food items for height and depth estimation in stereo image analysis on ETHZ Food-101 dataset. They have extended the dataset by adding extra 5800 food images from 11 food classes. They have achieved a Mean Error (ME) of 8.5% with four food classes. Similarly, Rahman et al. [122] have also used stereo food images to reconstruct 3D food images to estimate the volume of the foods from six fruit classes, where they have achieved a mean error of 7.7%.

G. HISTOGRAM, PRE-TRAINED 3D MODEL, GAN, DEEP CNN

Sudo et al. [25] have used histograms to detect food volume from 2D food images from a dataset of 2500 images. Their method of utilizing regression analysis with label histogram yielded better results than using predictor image features directly. This method has obtained mean errors from 31.8% to 40.6% in nutrition prediction. Hence, their method may not be applicable for reliable nutritional estimation in real life. Xu et al. [30] have used a pre-trained 3D model of various food shapes with food orientation information for 3D reconstruction of the food images. This method of food volume estimation has attained a mean error of 10% for the ETHZ Food-101 dataset with 5 food classes. In a study conducted by Fang et al. [78], the researchers have used Generative Adversarial Networks (GAN) to map the energy distribution in food images and attained an error rate of less than 10.89% for energy estimation. They have conducted the experiment on PFID dataset extended by their in-house dataset of 60000 new food images. Therefore, the researchers may investigate the characteristics of these CNN based techniques for better performance in volume estimation in future. In 2022, Kadam et al. [115] have at first utilized a fixed-dimension coin as a reference object for volume estimation and subsequently applied a RCNN-based food segmentation model as a volume estimator. The Deep Learning (DL) model was developed by fine tuning a pretrained ResNet model and trained using a dataset of Indian breakfast food images that included eight different classes of food in various shapes.

We have observed that despite the recent development of volume or weight estimation frameworks, it is still a challenging task to estimate volume from a single image without reference objects. Historically, shape templates [82], [125], silhouettes shapes [123] approaches are utilized by the researchers to estimate food volume. This technique is not applicable to food with irregular shapes. In real life, food images do not contain reference objects in the frames. Thus, the food volume estimation frameworks with reference objects [87], [124] will not have good performance in everyday life. Few works [5], [23] have used top and side views of meals for volume estimation. Yet, this technique still shows weakness for irregularly shaped food items. Most of these volume estimations from food images methods are conducted in controlled environments. In real life, most of these methods may not be applicable for reliable nutrition estimations. This is because although volume estimation is an important part of the food nutrition estimation system, the nutritional value of the food also depends on the food preparation methods. The volume estimation may not be able to differentiate the volume of the same foods that are prepared with different methods.

VI. NUTRITION ESTIMATION

Nutrition estimation from food includes calorie estimation, carbohydrate estimation, protein estimation, etc. In this paper, all kinds of food value estimations are considered nutrition estimations of food. The overarching aim of our review is to get a general understanding of food nutrition estimation systems using food images. The performance of the automated food nutrition estimation systems from food images depends on all of its sub-tasks, including the quality and quantity of food images in the datasets, accurate segmentation of the food images, proper food classification, estimation of the volume of identified food items, and finally retrieval of corresponding nutritional values of the food. Since the nutrition estimation of the food depends on the performance of the previous steps, the nutrition or calorie estimation may remain error prone in the long run. With error prone results, the nutritional value of the food may overestimate or underestimate. Some of the reviewed articles' primary focus is to estimate calorie intake from food images as input data without user intervention. Some other articles have used a semi-automatic approach for nutritional value estimation where they need feedback from users. Therefore, for counting approximate calories from food images, researchers have taken the following two distinct approaches: 1) automated retrieval of the nutritional information from food nutrition databases [27], [34], and

2) manual user input such as crowd sourcing using web platforms, smartphone apps, etc. [16], [103]. In the automated nutrition estimation system using food images, researchers have a food nutrition database where the nutritional values of all the foods' classes are given in standard measurement. To estimate the nutritional values of foods, the researchers use those nutritional values for the identified food classes and the estimated volumes of the food items. The ground truth nutritional values of the food classes can be collected from different sources. Some of the techniques used by the researchers to collect these data are given below.

US Department of Agriculture (USDA): USDA has a list of food items and descriptions with their caloric information. This calorie information is considered the standard value of the food. In some studies like the one conducted by Williamson et al. [129], the researchers have used the caloric values collected from USDA for their food value estimation frameworks.

Menu from Restaurants: Some health conscious restaurants provide the calorie, food preparation process, ingredients, etc. with the food menus. In the Menu Match dataset [14], the nutritional values along with the weight of the food items are obtained from the menu of the restaurants.

Input from Experts: Researchers can also collect calorie value of foods from nutrition experts in their controlled lab environment as presented in the study conducted by Meyers et al. [39]. This way, they can get the closest calorie values for each food item in the image.

Crowd-sourced: In this method, researchers use a web application that takes the eye estimated nutrition values of the food images from users around the world. This method may work for developing a large dataset such as UEC-FOOD256 dataset [16]. However, this technique is mostly error-prone. In the semi-automated nutrition estimation from food images, users manually provide nutritional value or other value from at least one of the sub-tasks including food class identification, food volume estimation, drawing bounding boxes for food segmentation, etc., to obtain the nutritional information of the target food image. Noronha et al. [128] proposed a framework where the nutritional information of the food image is crowd-sourced. Individual users have eye-estimated the nutritional values of the food images.

Table 10 displays the reviewed articles on calorie estimation from food images. Researchers, for example, [5], [60] have used a nutrition table and a density table. The nutrition table contains the weight and energy (calories) of the food, and the density table contains information on the density of the food. After food item segmentation, classification, and volume estimation, the estimated calorie is computed by the equation 7.

$$C_p = \frac{C_t \times V_e \times \rho_t}{M_t} \tag{7}$$

where, C_p is the estimated calorie of the target food item, C_t is the calorie of the identified food class, V_e is the estimated food volume, ρ_t is the standard density of the food item, and M_t is the standard mass of the food item.

Pouladzadeh et al. [5] have obtained an average accuracy of 86% in calorie estimation. Later, they improved the performance of their calorie estimation in [104] by proposing two different approaches to calculate the dimensions of the food items in the image: 1) utilizing finger as reference object, and 2) using distance and angle between the mobile and the food, and user's height. Both processes show a small range of standard error. In [63], the authors have grouped the food images by the range of calories for each of these food classes. For instance, Grilled pork with rice was in the range of 450-600 calories. Thus, if their framework could estimate the calorie within this range, the framework considered it as correct estimation. They first identified the food class and then predicted the caloric value of the identified food by using the predefined data about the amount of calories of each food-class. The accuracy and the false positive value for calorie estimation for each of the classes is in the range of 34%-54%. Chen et al. [94] have presented a calorie estimation framework that uses an identification function and an estimation function. The identification function finds five (5) top candidate food classes that most closely match the food items in the image. Interactively in the app, the user needs to select the correct food item and then the estimation function in the framework measures the quantity or amount of the food items. In [39], the authors have proposed a mobile framework that classifies the food items of the image in real time and uses the predicted class to look up the nutritional information of the food items. They have received -25.35 ± 26.37 and 152.95 ± 15.61 for mean error and mean absolute error, respectively, on the Menu Match dataset. In [114], the authors have developed a web application where a user uploads a target food image. The application identifies the food class and then calculates the caloric value in real time. In this study, the authors have computed the confidence level of the food classification model and the caloric value of each food item. Anthimopoulos et al. [23] have developed a Carbohydrate (CHO) estimation framework that does the food item segmentation, food class classification, volume estimation, and uses the USDA nutritional database to calculate the approximate CHO value of the food image. Most of the calorie estimation frameworks [23], [39] retrieve the nutritional value from the USDA nutrition database.

In the recent years of 2022 and 2023, the researchers are utilizing CNN based Deep Learning techniques to improve the performance of their nutrition estimation frameworks. Among the Deep Learning techniques, researchers are currently widely using the Mask R-CNN model for nutrition estimation [91]. Jaswanthi et al. [91] have achieved a good performance of mean Average Precision (mAP) of 85.43% for estimating caloric values using only Mask R-CNN. Other deep CNN based food calorie estimation techniques [91],

TABLE 10. Calorie estimation performance comparison. In-house means collected by the research team. Dataset column: total number of images followed by number of food classes in the parentheses, False Positive Rate (FPR), Macro Average Accuracy (MAA), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Accuracy (Acc).

Ref	Year	Dataset	Method	Calorie Estimation
[128]	2011	In-house	Calories Crowdsourced	ME: 41.0 ± 18.8, MAPE: 11.5%
[12]	2012	Before-after eaten	Detect salient objects' calorie and nutrients	NA
[60]	2012	In-house	Food classes, volumes, and calorie densities	NA, See volume estimation in Table 9
[94]	2012	5000 (50)	Identification and Estimation functions	NA, See food Classification in Table 6
[5]	2014	3000	Food segmentation, classification, volume estimation, & caloric equation	Acc.: 86%
[25]	2014	2500	Regression analysis & semantic segmentation	ME: 31.8% - 33.6%
[63]	2014	In-house	Cluster food images by calorie	Acc.: 46.4%, FPR: 38%
[39]	2015	ETHZ Food-101, Menu Match	Food classification, nutrition lookup	ME: -25.35 ± 26.37, MAE: 152.95 ± 15.61
[23]	2015	FooDD	CHO estimation: food identification, segmentation, volume estimation and USDA database	MAE: 12.28 ± 9.56 gm
[104]	2016	10000	Finger Based Calorie Measurement	Standard Error: 0.16
[87]	2018	FooDD	Volume estimation, calorie equation	Standard Error: 7.46%, Relative Standard Error: 4.5%
[27]	2021	256178	CNN (YOLOv4-tiny) & GrabCut	CHO estimation RMSE: 9.4
[114]	2021	23000 (23)	Food Classification: Inception V3 & CNN, Calorie from database	NA
[10]	2022	UNIMIB2016	Bayesian Fuzzy Clustering, Imperialist Competitive Algorithm based Deep Belief Network (IpCA-DBN)	MAA: 0.9643, Standard Accuracy: 0.9877, RMSE: 42.625
[115]	2022	In-house: 390(8)	Food shape: Amorphous, Convex, Regular (square & circle) with reference coin	NA, See volume estimation in Table 9
[130]	2022	In-house: (10)	Deep CNN with transfer learning, mean-shift segmentation, visual saliency	Acc: 98.51
[118]	2022	In-house	CNN, OpenCV	error variation of ± 10 calories
[131]	2022	In-house: (25)	Near Infrared Spectroscopy (NIRS) technology	15.32% better than baseline calorie estimation
[91]	2022	UNIMIB 2016	Deep CNN	Acc: 95.21%
[132]	2022	ChinaMartFood-109: 10,921 (18)	Inception V3, normalizing arithmetic and harmonic mean	Acc: 94 %
[92]	2022	In-house: 1646(6)	Mask R-CNN, ResNeXt-101-FPN	MAE: 5.26, R^2: 0.804
[133]	2022	Indian Food Nutrient Dataset (IFND): (255)	CNN	Acc: 97%

[118], [130], [132], [133] also achieve high performance for calorie estimation, between the accuracies of 94% [132] and 98.5% [130] and an error variation of ± 10 calories [118]. We also observe that additional techniques are used to improve the performance of the food nutrition estimation systems, such as mean shift segmentation, visual saliency in [130], OpenCV in [118], normalizing arithmetic mean and harmonic mean in [132]. Lately, in a study by Hu et al. [131], Near Infrared Spectroscopy (NIRS) technology has been used to estimate the calories of food images. Their method performs 15.32% better than baseline calorie estimation by CNN frameworks.

In recent years, the performance of nutrition estimation frameworks are improving drastically. From our observation, one of the main factors of this improvement is the utilization of DL-based algorithms [92], [132], [133] for calorie estimation. Traditionally, a nutritional look up table and calorie equation have been used for calorie estimation [5], [23]. These kinds of approaches heavily rely on the good performances of the previous steps such as food segmentation, classification, and volume estimation [23], [39]. Calorie estimation directly from the food images reduces this performance dependency [92], [132]. However, the insufficient food datasets with calorie values have made the training for DL-based nutrition estimation frameworks challenging. Moreover, a less diverse dataset domain hinders the growth of DL-based nutrition estimation frameworks by creating a domain-dependent system.

VII. DISCUSSION

This study provides a systematic review of the existing frameworks for the complete workflow of nutrition estimation systems from food images. Our findings categorize the nutrition estimation system into three different groups: Food Classification, Volume or Weight Estimation, and Nutrition Estimation. Additionally, our work encompasses other aspects of the nutrition estimation frameworks, such as methods of food image acquisition, description of food datasets used in the nutrition estimation frameworks, and the widely used input features for food classification and segmentation methods. Our review explores and compares the performance of the existing dietary related frameworks to comprehend the ongoing advancement in the field of image-based food nutrition estimation systems. Our research finds that in recent years, researchers are preferring utilizing Deep Learning techniques on all the steps of dietary assessment frameworks. Food segmentation methods have evolved from thresholding [23] and shape-based [82] Graph Cut [15] algorithms using food image features to Deep Learning techniques such as GAN [91], Mask R-CNN [92], DeepLanV3+ [90], Zero-shot segmentation [89], etc. Similar trends can also be seen in volume estimation frameworks [78], [115] and nutrition estimation frameworks [91], [92], [132], [133]. The most noticeable trend in using Deep Learning methods such as RCNN [115], OpenCN CNN [118], MobileNetV2 [119], and so on can be seen in food classification systems from 2014 to 2023. Researchers are enthusiastic about utilizing DL techniques because of their capability to learn directly from food images. However, the DL techniques are black-boxed and the network's internal logic is difficult to explain. This black-box method in training creates difficulties for researchers to comprehend why their framework is behaving in a certain way.

A. CHALLENGES

Though numerous research works have been observed in our study, there remain some challenges and limitations in the field of nutrition estimation from food images.

1) FOOD IMAGE SEGMENTATION AND CLASSIFICATION

It is difficult to classify food items in meal images consisting of multiple food classes. Hence, image segmentation is done before food classification. Food image segmentation is challenging as many foods have irregular features such as irregular shapes and edges, non-uniform contours, etc. It can be more difficult when multiple food items are mixed together or placed on top of one another resulting in occlusion. Many studies use GraphCut, GrabCut, etc., in food image segmentation methods. But, these methods have difficulties in generalizing the food segmentation process.

2) FOOD VOLUME/WEIGHT ESTIMATION

The proposed models have difficulties in doing volume estimation without any reference objects in the two-dimensional food images. It is also challenging to estimate volume and calorie, and to segment images of foods with irregular shapes, edges, non-uniform contours, rotation, low resolution, occlusion, mixed food items, etc. Differences in food preparation can result in different colors, shapes, and textures for the same food. This also adds challenges in this research area. Volume estimation from a single two dimensional image is also a challenging task.

A large image dataset with many food classes, and many images representing each class, is needed to achieve better estimation performance for both traditional ML and DL classification techniques. No large food image datasets with good image quality and many food classes are publicly available. Most of the large food image datasets appear to use web scraping to collect the data. Hence, the quality of the images is not good. Food image datasets with large food classes are needed to implement more advanced Deep Learning algorithms.

B. RECOMMENDATION FOR FUTURE WORKS

We suggest the following future research directions based on the research gaps we have found in our review.

Large Standard Food Dataset Construction: It is inevitable to develop a standard large-scale food image dataset such as "Imagenet" [134] in the future for advanced food nutrition evaluation. Researchers can scrap social media and relevant websites to amass a large amount of food images. Later, the experts can provide manual annotation and food information, such as calories, nutrition, food items, etc., to the dataset. This way, a large-scale standard dataset of food images can be constructed. Moreover, researchers should also consider different cuisines around the world and the difference in food making due to geological differences. Therefore, it is necessary to have joint efforts from scientists from all over the world to construct these large standard food datasets.

Personal Dieting and In-Patient Care: Food computing for personal dieting and In-Patient care will be a promising field for researchers. It is growing rapidly as a promising field in the health domain. Many researchers such as [23] and [114] have estimated calories and nutritional value from the food images to aid diabetic patients in need. As more and more people become health conscious, the demand for computational help for maintaining a healthy lifestyle will increase. Hence, one of the important future directions of food nutrition estimation systems will be building personalized food computational modeling for health care.

Robust and Generalized Food Recognition System: The first priority for dietary assessment and nutritional management systems is to develop robust and generalized food recognition systems. In recent years, Deep Learning approaches to recognize food items from images such as [7] and [115] have provided researchers with great opportunities. One of the limitations we have observed is that most of the Deep Learning approaches are not tested for drastically different cuisines from all over the world. With the construction of a large standard food image dataset of different cuisines, the researchers also have to build frameworks that can recognize food items from many different cuisines. Therefore, this can be a very prominent research direction that can be explored by scientists all around the world.

Broad Subtask Learning for Food Computing: We notice the existence of different subtasks of food image segmentation, food classification, food volume estimation while the studies calculate the nutritional values. These subtasks help the food value calculating framework to achieve better performance. However, from the review, we are yet to notice any large improvement in the cases of food image segmentation or food volume estimation. There is also a lack of datasets that have been annotated by experts for food image segmentation. Hence, the researchers can focus on constructing a large standard food image dataset to train the Deep Learning approaches to segment food items. It is also a challenging task to estimate the volume of the food from two-dimensional images. In recent years, studies like [13], and [42] have used GAN and pre-trained 3D modeling to create 3D views of the food portions from 2D images for volume estimation. However, utilizing GANs and pre-trained 3D models are still in the development phases for food volume estimation methods. Thus, we conclude that the research towards food volume estimation using appropriate food image datasets is a very viable future direction.

Food Computing for Health Logs: Food logs are most critical for health care. Food computing can be used for recommending nutritional foods based on previously logged food information. In [88] and [114], researchers have presented frameworks for food nutrition estimation. But these frameworks do not preserve the food records for the users. In the future, researchers can develop frameworks that can store these logs of daily nutrition intake. Thus, the users can reflect upon their food habits and can maintain their health.

Apart from the mentioned future directions, there can be other emerging areas in the field of food and nutrition, like construction of cooking robots, recommendation systems of food, prediction of the probabilities of disease from the daily food intake, creation of new recipes based on the users' preferences, etc.

VIII. CONCLUSION

Instantaneous estimation of food nutrition value from the food images is critical for multiple classes of people including pre-diabetic and pre-obese people, specially who are at lifelong risk of diabetes and obesity, and elderly people who are at risk of malnutrition. For all of them, quality of life is at stake. Availability of a lot of data and popularity of machine learning methods, especially Deep Learning techniques, have attracted many researchers to this field. Yet, we do not find much effort in extensive reviews on food value estimation from food images. In this paper, we have conducted an extensive literature review of food nutrition value estimation only from the image dataset as input. We have provided a food value application domain taxonomy and based our review on that. We have discussed the high impact research articles on food segmentation, food item classification, volume or weight estimation, and finally nutrition estimation. We have presented the current benchmark datasets along with their acquisition methods. We have provided and analyzed the mapping between the traditional ML techniques and the handcrafted image features. We have noticed an increasing trend of using Deep Learning algorithms for food item classification from images. This upward trend matches with the rapid advancement of computer vision-based Deep Learning algorithms. We have identified the current challenges related to the food image segmentation, food classification, and food volume estimation steps. We have recommended the opportunities for future work. These lines of future directions need further research and joint collaboration of scientists from all over the world.

REFERENCES

[1] A. Fakhrou, J. Kunhoth, and S. Al Maadeed, "Smartphone-based food recognition system using multiple deep CNN models," *Multimedia Tools Appl.*, vol. 80, nos. 21–23, pp. 33011–33032, Sep. 2021.

[2] J. Gao, W. Tan, L. Ma, Y. Wang, and W. Tang, "MUSEFood: Multi-sensor-based food volume estimation on smartphones," in *Proc. IEEE SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, Aug. 2019, pp. 899–906.

[3] S. Kayikci, Y. Basol, and E. Dorter, "Classification of Turkish cuisine with deep learning on mobile platform," in *Proc. 4th Int. Conf. Comput. Sci. Eng. (UBMK)*, 2019, pp. 1–5.

[4] Y. Kawano and K. Yanai, "FoodCam-256: A large-scale real-time mobile food RecognitionSystem employing high-dimensional features and compression of classifier weights," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 761–762.

[5] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, "Measuring calorie and nutrition from food image," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 8, pp. 1947–1956, Aug. 2014.

[6] L. Jiang, B. Qiu, X. Liu, C. Huang, and K. Lin, "DeepFood: Food image analysis and dietary assessment via deep model," *IEEE Access*, vol. 8, pp. 47477–47489, 2020.

[7] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[8] X. Chen, Y. Zhu, H. Zhou, L. Diao, and D. Wang, "ChineseFoodNet: A large-scale image dataset for Chinese food recognition," 2017, *arXiv:1705.02743*.

[9] M. Chopra and A. Purwar, "Food image recognition by optimizing CNN with PSO and GA," in *Proc. 14th Int. Conf. Contemp. Comput.*, New York, NY, USA, Aug. 2022, pp. 37–42.

[10] S. J. Minija and W. R. S. Emmanuel, "Imperialist competitive algorithm-based deep belief network for food recognition and calorie estimation," *Evol. Intell.*, vol. 15, no. 2, pp. 955–970, Jun. 2022.

[11] F. Zhou and Y. Lin, "Fine-grained image classification by exploring bipartite-graph labels," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1124–1133.

[12] G. Villalobos, R. Almaghrabi, P. Pouladzadeh, and S. Shirmohammadi, "An image processing approach for calorie intake measurement," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2012, pp. 1–5.

[13] S. Fang, Z. Shao, R. Mao, C. Fu, E. J. Delp, F. Zhu, D. A. Kerr, and C. J. Boushey, "Single-view food portion estimation: Learning image-to-energy mappings using generative adversarial networks," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 251–255.

[14] O. Beijbom, N. Joshi, D. Morris, S. Saponas, and S. Khullar, "Menu-Match: Restaurant-specific food logging from images," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2015, pp. 844–851.

[15] P. Pouladzadeh, A. Yassine, and S. Shirmohammadi, "FooDD: Food detection dataset for calorie measurement using food images," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 441–448.

[16] Y. Kawano and K. Yanai, "Automatic expansion of a food image dataset leveraging existing categories with domain adaptation," in *Proc. ECCV Workshop Transferring Adapting Source Knowl. Comput. Vis. (TASK-CV)*, 2014, pp. 3–17.

[17] W. Min, S. Jiang, L. Liu, Y. Rui, and R. Jain, "A survey on food computing," *ACM Comput. Surveys*, vol. 52, no. 5, pp. 1–36, Sep. 2020.
[18] M. A. Subhi, S. H. Ali, and M. A. Mohammed, "Vision-based approaches for automatic food recognition and dietary assessment: A survey," *IEEE Access*, vol. 7, pp. 35370–35381, 2019.
[19] M. Chopra and A. Purwar, "Recent studies on segmentation techniques for food recognition: A survey," *Arch. Comput. Methods Eng.*, vol. 29, no. 2, pp. 865–878, Mar. 2022.
[20] K. V. Dalakleidi, M. Papadelli, I. Kapolos, and K. Papadimitriou, "Applying image-based food-recognition systems on dietary assessment: A systematic review," *Adv. Nutrition*, vol. 13, no. 6, pp. 2590–2619, Nov. 2022.
[21] L. M. Amugongo, A. Kriebitz, A. Boch, and C. Lutge, "Mobile computer vision-based applications for food recognition and volume and calorific estimation: A systematic review," in *Healthcare*, vol. 11. Basel, Switzerland: Multidisciplinary Digital Publishing Institute, 2023, p. 59.
[22] L. M. König, M. Van Emmenis, J. Nurmi, A. Kassavou, and S. Sutton, "Characteristics of smartphone-based dietary assessment tools: A systematic review," *Health Psychol. Rev.*, vol. 16, no. 4, pp. 526–550, Oct. 2022.
[23] M. Anthimopoulos, J. Dehais, S. Shevchik, B. H. Ransford, D. Duke, P. Diem, and S. Mougiakakou, "Computer vision-based carbohydrate estimation for type 1 patients with diabetes using smartphones," *J. Diabetes Sci. Technol.*, vol. 9, no. 3, pp. 507–515, May 2015.
[24] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Food image analysis: Segmentation, identification and weight estimation," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2013, pp. 1–6.
[25] K. Sudo, K. Murasaki, J. Shimamura, and Y. Taniguchi, "Estimating nutritional value from food images based on semantic segmentation," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., Adjunct Publication*, New York, NY, USA, Sep. 2014, pp. 571–576.
[26] W. Shimoda and K. Yanai, "Foodness proposal for multiple food detection by training of single food images," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manag.*, Oct. 2016, pp. 13–21.
[27] P. Chotwanvirat, N. Hnoohom, N. Rojroongwasinkul, and W. Kriengsinyos, "Feasibility study of an automated carbohydrate estimation system using Thai food images in comparison with estimation by dietitians," *Frontiers Nutrition*, vol. 8, Oct. 2021, Art. no. 732449.
[28] S. Christodoulidis, M. Anthimopoulos, and S. Mougiakakou, "Food recognition for dietary assessment using deep convolutional neural networks," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 458–465.
[29] Y. Yang, W. Jia, T. Bucher, H. Zhang, and M. Sun, "Image-based food portion size estimation using a smartphone without a fiducial marker," *Public Health Nutrition*, vol. 22, no. 7, pp. 1180–1192, May 2019.
[30] C. Xu, Y. He, N. Khannan, A. Parra, C. Boushey, and E. Delp, "Image-based food volume estimation," in *Proc. 5th Int. Workshop Multimedia Cooking Eating Activities*, 2013, pp. 75–80.
[31] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.
[32] I. Gallo, A. Calefati, S. Nawaz, and M. K. Janjua, "Image and encoded text fusion for multi-modal classification," in *Proc. Digit. Image Comput., Techn. Appl. (DICTA)*, Dec. 2018, pp. 1–7.
[33] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," in *Proc. IEEE 30th Can. Conf. Electr. Comput. Eng. (CCECE)*, Apr. 2017, pp. 1–4.
[34] J. Harashima, Y. Someya, and Y. Kikuta, "Cookpad image dataset: An image collection as infrastructure for food research," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 1229–1232.
[35] L. Bossard, M. Guillaumin, and L. V. Gool, "Food-101—Mining discriminative components with random forests," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 446–461.
[36] M. M. Anthimopoulos, L. Gianola, L. Scarnato, P. Diem, and S. G. Mougiakakou, "A food recognition system for diabetic patients based on an optimized bag-of-features model," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1261–1271, Jul. 2014.
[37] G. M. Farinella, D. Allegra, and F. Stanco, "A benchmark dataset to study the representation of food images," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 584–599.
[38] Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2012, pp. 25–30.
[39] A. Myers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. Murphy, "Im2Calories: Towards an automated mobile vision food diary," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1233–1241.
[40] X. Wang, D. Kumar, N. Thome, M. Cord, and F. Precioso, "Recipe recognition with large multimodal food dataset," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun. 2015, pp. 1–6.
[41] P. Pouladzadeh, S. Shirmohammadi, A. Bakirov, A. Bulut, and A. Yassine, "Cloud-based SVM for food categorization," *Multimedia. Tools Appl.*, vol. 74, no. 14, pp. 5243–5260, Jul. 2015.
[42] R. Xu, L. Herranz, S. Jiang, S. Wang, X. Song, and R. Jain, "Geolocalized modeling for dish recognition," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1187–1199, Aug. 2015.
[43] G. Ciocca, P. Napoletano, and R. Schettini, "Food recognition and leftover estimation for daily diet monitoring," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 334–341.
[44] G. Ciocca, P. Napoletano, and R. Schettini, "Food recognition: A new dataset, experiments, and results," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 588–598, May 2017.
[45] M. Merler, H. Wu, R. Uceda-Sosa, Q.-B. Nguyen, and J. R. Smith, "Snap, Eat, RepEat: A food recognition engine for dietary logging," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manag.*, Oct. 2016, pp. 31–40.
[46] J. Rich, H. Haddadi, and T. M. Hospedales, "Towards bottom-up analysis of social food," in *Proc. 6th Int. Conf. Digit. Health Conf.*, Apr. 2016, pp. 111–120.
[47] A. Singla, L. Yuan, and T. Ebrahimi, "Food/non-food image classification and food categorization using pre-trained GoogLeNet model," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manag.*, Oct. 2016, pp. 3–11.
[48] G. M. Farinella, D. Allegra, M. Moltisanti, F. Stanco, and S. Battiato, "Retrieval and classification of food images," *Comput. Biol. Med.*, vol. 77, pp. 23–39, Oct. 2016.
[49] G. Ciocca, P. Napoletano, and R. Schettini, "Learning CNN-based features for retrieval of food images," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2017, pp. 426–434.
[50] S. Hou, Y. Feng, and Z. Wang, "VegFru: A domain-specific dataset for fine-grained visual categorization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 541–549.
[51] G. Waltner, M. Schwarz, S. Ladstatter, A. Weber, P. Luley, M. Lindschinger, I. Schmid, W. Scheitz, H. Bischof, and L. Paletta, "Personalized dietary self-management using mobile vision-based assistance," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2017, pp. 385–393.
[52] H. Muresan and M. Oltean, "Fruit recognition from images using deep learning," *Acta Universitatis Sapientiae, Inf.*, vol. 10, no. 1, pp. 26–42, Aug. 2018.
[53] C. Gungor, F. Baltaci, A. Erdem, and E. Erdem, "Turkish cuisine: A benchmark dataset with Turkish meals for food recognition," in *Proc. 25th Signal Process. Commun. Appl. Conf. (SIU)*, May 2017, pp. 1–4.
[54] E. Aguilar, M. Bolaños, and P. Radeva, "Regularized uncertainty-based multi-task learning model for food analysis," *J. Vis. Commun. Image Represent.*, vol. 60, pp. 360–370, Apr. 2019.
[55] Q. Cai, J. Li, H. Li, and Y. Weng, "BTBUFood-60: Dataset for object detection in food field," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Feb. 2019, pp. 1–4.
[56] J. Qiu, F. P.-W. Lo, Y. Sun, S. Wang, and B. Lo, "Mining discriminative food regions for accurate food recognition," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2019, pp. 1–11.
[57] W. Min, L. Liu, Z. Luo, and S. Jiang, "Ingredient-guided cascaded multi-attention network for food recognition," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 1331–1339.
[58] W. Min, L. Liu, Z. Wang, Z. Luo, X. Wei, X. Wei, and S. Jiang, "ISIA Food-500: A dataset for large-scale food recognition via stacked global-local attention network," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 393–401.
[59] W. Min, Z. Wang, Y. Liu, M. Luo, L. Kang, X. Wei, X. Wei, and S. Jiang, "Large scale visual food recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jan. 18, 2023.

[60] F. Kong and J. Tan, "DietCam: Automatic dietary assessment with mobile camera phones," *Pervas. Mob. Comput.*, vol. 8, no. 1, pp. 147–163, Feb. 2012.

[61] M. Bosch, F. Zhu, N. Khanna, C. J. Boushey, and E. J. Delp, "Combining global and local features for food identification in dietary assessment," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1789–1792.

[62] Y. Kawano and K. Yanai, "Real-time mobile food recognition system," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 1–7.

[63] N. Tammachat and N. Pantuwong, "Calories analysis of food intake using image recognition," in *Proc. 6th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Oct. 2014, pp. 1–4.

[64] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun. 2015, pp. 1–6.

[65] M. A. Subhi and S. Md. Ali, "A deep convolutional neural network for food detection and recognition," in *Proc. IEEE-EMBS Conf. Biomed. Eng. Sci. (IECBES)*, Dec. 2018, pp. 284–287.

[66] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Feb. 1999, pp. 1150–1157.

[67] F. Kong and J. Tan, "DietCam: Regular shape food recognition with a camera phone," in *Proc. Int. Conf. Body Sensor Netw.*, May 2011, pp. 127–132.

[68] Y. He, C. Xu, N. Khanna, C. J. Boushey, and E. J. Delp, "Analysis of food images: Features and classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 2744–2748.

[69] A. E. Abdel-Hakim and A. A. Farag, "CSIFT: A SIFT descriptor with color invariant characteristics," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 1978–1983.

[70] B. Kartikeyan and A. Sarkar, "An identification approach for 2-D autoregressive models in describing textures," *CVGIP, Graph. Models Image Process.*, vol. 53, no. 2, pp. 121–131, Mar. 1991.

[71] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[72] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[73] R. Arandjelovic and A. Zisserman, "Three things everyone should know to improve object retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2911–2918.

[74] T. Ojala, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Aug. 2002.

[75] S. K. Yarlagadda, D. M. Montserrat, D. Guera, C. J. Boushey, D. A. Kerr, and F. Zhu, "Saliency-aware class-agnostic food image segmentation," *ACM Trans. Comput. Healthcare*, vol. 2, no. 3, pp. 1–17, Jul. 2021.

[76] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. CVPR*, Jun. 2011, pp. 2097–2104.

[77] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, vol. 463. New York, NY, USA: ACM Press, 1999.

[78] S. Fang, C. Liu, K. Tahboub, F. Zhu, E. J. Delp, and C. J. Boushey, "CTADA: The design of a crowdsourcing tool for online food image identification and segmentation," in *Proc. IEEE Southwest Symp. Image Anal. Interpretation (SSIAI)*, Apr. 2018, pp. 25–28.

[79] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, "Food recognition using statistics of pairwise local features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2249–2256.

[80] C. Rother, V. Kolmogorov, and A. Blake, "'GrabCut': Interactive foreground extraction using iterated graph cuts," in *Proc. ACM SIGGRAPH Papers*, New York, NY, USA, 2004, pp. 309–314.

[81] W. Shimoda and K. Yanai, "CNN-based food image segmentation without pixel-wise annotation," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 449–457.

[82] J. Chae, I. Woo, S. Kim, R. Maciejewski, F. Zhu, E. J. Delp, C. J. Boushey, and D. S. Ebert, "Volume estimation using food specific shape templates in mobile image-based dietary assessment," in *Proc. SPIE*, vol. 7873, Feb. 2011, Art. no. 78730K.

[83] W. Wu and J. Yang, "Fast food recognition from videos of eating for calorie estimation," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jun. 2009, pp. 1210–1213.

[84] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey, and E. J. Delp, "Multiple hypotheses image segmentation and classification with application to dietary assessment," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 1, pp. 377–388, Jan. 2015.

[85] F. Kong, H. He, H. A. Raynor, and J. Tan, "DietCam: Multi-view regular shape food recognition with a camera phone," *Pervasive Mobile Comput.*, vol. 19, pp. 108–121, May 2015.

[86] D. Rhyner, H. Loher, J. Dehais, M. Anthimopoulos, S. Shevchik, R. H. Botwey, D. Duke, C. Stettler, P. Diem, and S. Mougiakakou, "Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: A comparative study," *J. Med. Internet Res.*, vol. 18, no. 5, p. e101, May 2016.

[87] N. Sadeq, F. R. Rahat, A. Rahman, S. I. Ahamed, and M. K. Hasan, "Smartphone-based calorie estimation from food image using distance information," in *Proc. 5th Int. Conf. Netw., Syst. Secur. (NSysS)*, Dec. 2018, pp. 1–8.

[88] C. N. C. Freitas, F. R. Cordeiro, and V. Macario, "MyFood: A food segmentation and classification system to aid nutritional monitoring," in *Proc. 33rd SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Nov. 2020, pp. 234–239.

[89] Y. Honbu and K. Yanai, "Unseen food segmentation," in *Proc. Int. Conf. Multimedia Retr.*, Jun. 2022, pp. 19–23.

[90] E. Aguilar, B. Nagarajan, B. Remeseiro, and P. Radeva, "Bayesian deep learning for semantic segmentation of food images," *Comput. Electr. Eng.*, vol. 103, Oct. 2022, Art. no. 108380.

[91] R. Jaswanthi, E. Amruthatulasi, C. Bhavyasree, and A. Satapathy, "A hybrid network based on GAN and CNN for food segmentation and calorie estimation," in *Proc. Int. Conf. Sustain. Comput. Data Commun. Syst. (ICSCDS)*, Apr. 2022, pp. 436–441.

[92] N. Aditama and R. Munir, "Indonesian street food calorie estimation using mask R-CNN and multiple linear regression," in *Proc. 2nd Int. Conf. Power, Control Comput. Technol. (ICPCT)*, Mar. 2022, pp. 1–6.

[93] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 800–810, Aug. 2001.

[94] M.-Y. Chen, Y.-H. Yang, C.-J. Ho, S.-H. Wang, S.-M. Liu, E. Chang, C.-H. Yeh, and M. Ouhyoung, "Automatic Chinese food identification and quantity estimation," in *Proc. SIGGRAPH Asia Tech. Briefs*, Nov. 2012, pp. 1–4.

[95] P. Duan, W. Wang, W. Zhang, F. Gong, P. Zhang, and Y. Rao, "Food image recognition using pervasive cloud computing," in *Proc. IEEE Int. Conf. Green Comput. Commun. IEEE Internet Things IEEE Cyber, Phys. Social Comput.*, Aug. 2013, pp. 1631–1637.

[96] Y. Wang, Y. He, F. Zhu, C. Boushey, and E. Delp, "The use of temporal information in food image analysis," in *Proc. Int. Conf. Image Anal. Process.* Cham, Switzerland: Springer, 2015, pp. 317–325.

[97] H. Kagaya, K. Aizawa, and M. Ogawa, "Food detection and recognition using convolutional neural network," in *Proc. 22nd ACM Int. Conf. Multimedia*, New York, NY, USA, Nov. 2014, pp. 1085–1088.

[98] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, M. Yunsheng, S. Chen, and P. Hou, "A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure," *IEEE Trans. Services Comput.*, vol. 11, no. 2, pp. 249–261, Jan. 2018.

[99] F. Yu, D. Wang, E. Shelhamer, and T. Darrell, "Deep layer aggregation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2403–2412.

[100] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *Proc. 2nd Int. Workshop Multimedia Assist. Dietary Manag.*, Oct. 2016, pp. 41–49.

[101] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, and Y. Ma, "DeepFood: Deep learning-based food image recognition for computer-aided dietary assessment," in *Proc. Int. Conf. Smart Homes Health Telematics*. Cham, Switzerland: Springer, 2016, pp. 37–48.

[102] K. Waki, K. Aizawa, S. Kato, H. Fujita, H. Lee, H. Kobayashi, M. Ogawa, K. Mouri, T. Kadowaki, and K. Ohe, "DialBetics with a multimedia food recording tool, FoodLog: Smartphone-based self-management for type 2 diabetes," *J. Diabetes Sci. Technol.*, vol. 9, no. 3, pp. 534–540, May 2015.

[103] J. Chen and C.-W. Ngo, "Deep-based ingredient recognition for cooking recipe retrieval," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 32–41.

[104] P. Pouladzadeh, P. Kuhad, S. V. B. Peddi, A. Yassine, and S. Shirmohammadi, "Food calorie measurement using deep learning neural network," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2016, pp. 1–6.

[105] S. Mezgec and B. K. Seljak, "NutriNet: A deep learning food and drink image recognition system for dietary assessment," *Nutrients*, vol. 9, no. 7, p. 657, Jun. 2017.

[106] P. Pandey, A. Deepthi, B. Mandal, and N. B. Puhan, "FoodNet: Recognizing foods using ensemble of deep networks," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1758–1762, Dec. 2017.

[107] R. Yunus, O. Arif, H. Afzal, M. F. Amjad, H. Abbas, H. N. Bokhari, S. T. Haider, N. Zafar, and R. Nawaz, "A framework to estimate the nutritional value of food in real time using deep learning techniques," *IEEE Access*, vol. 7, pp. 2643–2652, 2019.

[108] A. B. Akhi, F. Akter, T. Khatun, and M. S. Uddin, "Recognition and classification of fast food images," *Global J. Comput. Sci. Technol.*, vol. 18, no. 1, pp. 7–13, 2018.

[109] N. F. P. Setyono, D. Chahyati, and M. I. Fanany, "Betawi traditional food image detection using ResNet and DenseNet," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Oct. 2018, pp. 441–445.

[110] N. Martinel, G. L. Foresti, and C. Micheloni, "Wide-slice residual networks for food recognition," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 567–576.

[111] S. Jiang, W. Min, L. Liu, and Z. Luo, "Multi-scale multi-view deep feature aggregation for food recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 265–276, 2020.

[112] Y. Li, X. Xu, and C. Yuan, "Enhanced mask R-CNN for Chinese food image detection," *Math. Problems Eng.*, vol. 2020, pp. 1–8, Jul. 2020.

[113] X. Chen, E. Johnson, A. Kulkarni, C. Ding, N. Ranelli, Y. Chen, and R. Xu, "An exploratory approach to deriving nutrition information of restaurant food from crowdsourced food images: Case of hartford," *Nutrients*, vol. 13, no. 11, p. 4132, Nov. 2021.

[114] S. A. Ayon, C. Z. Mashrafi, A. B. Yousuf, F. Hossain, and M. I. Hossain, "FoodieCal: A convolutional neural network based food detection and calorie estimation system," in *Proc. Nat. Comput. Colleges Conf. (NCCC)*, Mar. 2021, pp. 1–6.

[115] P. Kadam, S. Pandya, S. Phansalkar, M. Sarangdhar, N. Petkar, K. Kotecha, and D. Garg, "FVEstimator: A novel food volume estimator wellness model for calorie measurement and healthy living," *Measurement*, vol. 198, Jul. 2022, Art. no. 111294.

[116] J. R. Rajayogi, G. Manjunath, and G. Shobha, "Indian food image classification with transfer learning," in *Proc. 4th Int. Conf. Comput. Syst. Inf. Technol. Sustain. Solution (CSITSS)*, Dec. 2019, pp. 91–100.

[117] A. N. M. Zulfikri, F. Y. A. Rahman, S. Shabuddin, and R. Mohamad, "Food recognition based on deep learning algorithms," in *Proc. IEEE Symp. Ind. Electron. Appl. (ISIEA)*, Jul. 2022, pp. 1–4.

[118] S. Sathish, S. Ashwin, M. A. Quadir, and L. K. Pavithra, "Analysis of convolutional neural networks on Indian food detection and estimation of calories," *Mater. Today, Proc.*, vol. 62, pp. 4665–4670, Jan. 2022.

[119] S. Elbassuoni, H. Ghattas, J. E. Ati, Z. Shmayssani, S. Katerji, Y. Zoughbi, A. Semaan, C. Akl, H. B. Gharbia, and S. Sassi, "DeepNOVA: A deep learning Nova classifier for food images," *IEEE Access*, vol. 10, pp. 128523–128535, 2022.

[120] M. Bolanos and P. Radeva, "Simultaneous food localization and recognition," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 3140–3145.

[121] J. Shang, M. Duong, E. Pepin, X. Zhang, K. Sandara-Rajan, A. Mamishev, and A. Kristal, "A mobile structured light system for food volume estimation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 100–101.

[122] M. H. Rahman, Q. Li, M. Pickering, M. Frater, D. Kerr, C. Bouchey, and E. Delp, "Food volume estimation in a mobile phone based dietary assessment system," in *Proc. 8th Int. Conf. Signal Image Technol. Internet Based Syst.*, Nov. 2012, pp. 988–995.

[123] C. Xu, Y. He, N. Khanna, C. J. Boushey, and E. J. Delp, "Model-based food volume estimation using 3D pose," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 2534–2538.

[124] S. Fang, C. Liu, F. Zhu, E. J. Delp, and C. J. Boushey, "Single-view food portion estimation based on geometric models," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2015, pp. 385–390.

[125] W. Jia, H.-C. Chen, Y. Yue, Z. Li, J. Fernstrom, Y. Bai, C. Li, and M. Sun, "Accuracy of food portion size estimation from digital pictures acquired by a chest-worn camera," *Public Health Nutrition*, vol. 17, no. 8, pp. 1671–1681, 2014.

[126] J. Dehais, M. Anthimopoulos, S. Shevchik, and S. Mougiakakou, "Two-view 3D reconstruction for food volume estimation," *IEEE Trans. Multimedia*, vol. 19, no. 5, pp. 1090–1099, May 2017.

[127] Y. Yue, W. Jia, and M. Sun, "Measurement of food volume based on single 2-D image without conventional camera calibration," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 2166–2169.

[128] J. Noronha, E. Hysen, H. Zhang, and K. Z. Gajos, "Platemate: Crowdsourcing nutritional analysis from food photographs," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2011, pp. 1–12.

[129] D. A. Williamson, H. R. Allen, P. D. Martin, A. Alfonso, B. Gerald, and A. Hunt, "Digital photography: A new method for estimating food intake in cafeteria settings," *Eating Weight Disorders-Stud. Anorexia, Bulimia Obesity*, vol. 9, no. 1, pp. 24–28, Mar. 2004.

[130] R. Hafiz, M. R. Haque, A. Rakshit, and M. S. Uddin, "Image-based soft drink type classification and dietary assessment system using deep convolutional neural network with transfer learning," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 5, pp. 1775–1784, May 2022.

[131] H. Hu, Q. Zhang, and Y. Chen, "NIRSCam: A mobile near-infrared sensing system for food calorie estimation," *IEEE Internet Things J.*, vol. 9, no. 19, pp. 18934–18945, Oct. 2022.

[132] P. Ma, C. P. Lau, N. Yu, A. Li, and J. Sheng, "Application of deep learning for image-based Chinese market food nutrients estimation," *Food Chem.*, vol. 373, Mar. 2022, Art. no. 130994.

[133] L. K. Gautam and V. S. Gulhane, "Food assessment model for Indian elderly persons using CNN and image processing techniques," in *Proc. 3rd Int. Conf. Commun., Comput. Electron. Syst.* Cham, Switzerland: Springer, 2022, pp. 1093–1104.

[134] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

JAMALIA SULTANA received the B.S. degree in computer science and engineering (CSE) from the Bangladesh University of Engineering and Technology (BUET), where she is currently pursuing the degree with the Department of CSE. She is also a Graduate Research Fellow with the Department of CSE, BUET. Her research interests include applied machine learning in interdisciplinary fields, mathematical models, security, and reliability.

BENZIR MD. AHMED received the B.S. degree in electrical engineering from the Bangladesh University of Engineering and Technology (BUET), and the M.S. degree in computer science from North Dakota State University (NDSU), USA, in 2005. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering (CSE), BUET. He is an Assistant Professor with the Department of CSE, United International University (UIU).

MOHAMMAD MEHEDY MASUD received the Ph.D. degree in computer science from the University of Texas at Dallas, USA, in December 2009. He is an Associate Professor with the College of Information Technology (CIT), United Arab Emirates University (UAEU). During his service at UAEU, he secured more than 7.2M Dirhams (US$2M) research grant from different external funding organizations, including Information Communication Technology (ICT), United Arab Emirates. He has published more than 60 research articles in impact factor journals, including IEEE Transactions on Knowledge and Data Engineering and *Knowledge and Information Systems* (KAIS) and high ranked conferences, including IEEE International Conference on Data Mining (ICDM). He is the lead author of the book *Data Mining Tools for Malware Detection* and the Principal Inventor of a U.S. patent. His research interests include big data mining, healthcare data analytics, healthcare decision support, data stream mining, machine learning, and intrusion detection. He has delivered keynote speeches and tutorials in numerous international conferences and workshops. He has received several awards for his scholarship and service contributions, including the Prestigious Academic Award for Research Excellence from UAE Department of Education and Knowledge (ADEK), in 2017; the UAE University Merit Excellence Award, in 2019 and 2020; the College Research Excellence Award from the College of IT, in 2018; and the College Service Excellence Award from the College of IT, in 2019. He served as the lead guest editor for journals, the program co-chair for conferences, and in technical program committees of many international conferences.

A. K. OBIDUL HUQ received the Ph.D. degree from the University of Malaysia, in 2017. He is a Professor with the Department of Food Technology and Nutritional Science, Mawlana Bhashani Science and Technology University (MBSTU), Bangladesh. He is also the Team Leader of the Functional Food Research Group and the President of the Pushtibid Development Foundation, Bangladesh. He has published more than 70 peer-reviewed articles in different journals and periodicals and a text book *Human Nutrition and Applied Dietetics*. His research interests include human nutrition, dietetics, functional foods, food chemistry and analysis, development of eco-friendly bio-adsorbents for the treatment of wastewater effluent, and the management of food industries' by-products.

MOHAMMED EUNUS ALI is a Professor with the Department of CSE, Bangladesh University of Engineering and Technology (BUET). He is the Group Leader of the Data Science and Engineering Research Laboratory (DataLab), CSE, BUET. His research papers have published in top ranking journals and conferences, such as the *VLDB Journal*, IEEE Transactions on Knowledge and Data Engineering, *DMKD*, *Information Systems Journal*, *WWWJ*, *DKE*, ICDE, CIKM, EDBT, PVLDB, and UbiComp. His research interests include database systems and information management, including spatial databases, practical machine learning, and social media analytics. He served as a Program Committee Member for many prestigious conferences, including SIGMOD, VLDB, AAAI, and SIGSPATIAL.

MAHMUDA NAZNIN (Member, IEEE) received the B.Sc. and M.Sc. degrees in CSE from Bangladesh University of Engineering and Technology (BUET), and the M.S. and Ph.D. degrees in CS from North Dakota State University (NDSU), USA. She is a Professor and the Head of the Department of CSE, BUET. She has published more than 40 research publications in different international peer reviewed journals and conferences, such as IEEE Photonic Technology Letters, Springer's *Wireless Networks*, *Journal of Graph Algorithms and Applications*, IEEE Globecom, IEEE WiMob, and IEEE Compsac. She is involved in different professional activities and serves as a reviewer for different internationally reputed journals, such as IEEE Transactions on Computational Social Systems, IEEE Transactions on Vehicular Communication, Springer's *Wireless Networks*, and *Peer to Peer Networks*. She has served as a TPC co-chair, a TPC member, and a panelist, in different international conferences and workshops.