How to Cite:
Bahadur, S., Kushwaha, R., Sharma, S., & Anuradha, S. G. (2022). Using deep learning techniques, a framework for estimating the nutritional value of food in real time. *International Journal of Health Sciences*, *6*(S2), 7635–7648. https://doi.org/10.53730/ijhs.v6nS2.6877

Using deep learning techniques, a framework for estimating the nutritional value of food in real time

Ms. Shefali Bahadur
Research Scholar, Amity Business School, Amity University, Lucknow

Dr. Rohit Kushwaha
Director, Amity Business School, Amity University, Lucknow

Dr. Suruchi Sharma
Associate Professor, School of Management, Graphic Era Hill University Dehradun

Dr. Anuradha S G
Professor, Department of CSE, RYMEC, affiliated to VTU, Karnataka, India

***Abstract*---**Dietary disorders have increased dramatically in recent decades as a result of poor eating habits. Mobile-based dietary evaluation systems that can take real-time pictures of meals and assess their nutritional content might be very helpful in changing dietary habits and, as a result, leading to a healthier lifestyle. By categorizing the input image of food, this article offers a unique approach for automatically estimating food properties such as components and nutritional value. For reliable food recognition, we use a variety of deep learning models. Aside from image analysis, features and components are computed using semantically relevant phrases extracted from a vast corpus of literature acquired over the Internet. We performed testing on a dataset of 100 classes with an average of 1000 images per class, and we were able to get a top 1 classification rate of up to 85%. As an extension of the benchmark dataset Food-101, a sub-continental food dataset has been created. According to the findings, our proposed technique is equally effective on the core Food-101 dataset and its expansion for sub-continental foods. The proposed solution is implemented as a mobile app for the healthcare business.

***Keywords*---**convolutional neural networks, vector embedding, attribute prediction, food identification.

Manuscript submitted: 27 Feb 2022, Manuscript revised: 09 March 2022, Accepted for publication: 18 April 2022

Introduction

High calorie food consumption can be detrimental and lead to obesity, a preventable medical disease characterized by abnormal fat buildup in the body. Obesity, diabetes, high cholesterol, heart attacks, high blood pressure, breast, colon, and prostate cancers [1], as well as other diet-related diseases, are all possibilities. People are more inclined to change their diet plans by paying more attention to the sorts of food they consume in order to deal with such difficulties. Dietary management is a major problem for people of all ages. Maintaining a balance between what one eats and how one limits food consumption, on the other hand, is a critical aspect of diet management. The vast rise in diseases such as high cholesterol, blood pressure, and strokes has created a need for nutritional and diet control, which has led to the use of costly nutrition treatments. Energy balance is well-known for its importance in sustaining a healthy weight and lifestyle. The above-mentioned diseases and allergies can be reduced if people become more aware of their food intake and nutritional value. The objective of this project is to develop a smartphone app that can take real-time photographs of meals and assess their nutritional value, allowing people to modify their eating habits and live a better lifestyle.

The majority of existing solutions, which are typically implemented as smart phone applications (for example, MyFitnessPal [2], SHealth [3]), let users keep track of their food consumption. Users may utilize these systems to help them achieve dietary objectives including weight gain/loss, allergy control, and eating a balanced diet. They do, however, require consumers to manually enter meal specifics and portion amounts. This may be highly time consuming and tiresome, causing consumers to avoid using these programs for lengthy periods of time.

Attributes		Ingredients		Nutrition	
Grill	1	Beef	1	Fat	1
Spicy	0.75	Cheese	0.9	Sodium	0.55
Juicy	0.64	Lettuce	0.8	Protein	0.44
Fry	0.54	Tomato	0.67	Cholesterol	0.24
Tasty	0.53	Potato	0.58	Carbs	0.7
Healthy	0.33	Soya	0.53	Zinc	0.3
		Egg	0.52	Manganese	0.15
		Onion	0.50		

Attributes		Ingredients		Nutrition	
Fry	1	Rice	1	Carbs	1
Spicy	0.69	Chicken	0.73	Fat	0.8
Grill	0.55	Saffron	0.63	Starch	0.49
Healthy	0.5	Chilli	0.57	Fiber	0.38
Hot	0.43	Ghee	0.55	Sodium	0.36
Healthy	0.33	Soya	0.4	Manganese	0.35
Juicy	0.39	Tomato	0.3	Zinc	0.23
Sour	0.21				

Figure 1. The suggested method estimates several food characteristics. The system also estimates food components and nutritional content. (a) Burger. (b) Biryani.

Furthermore, inexperienced users rely on self-reported calorie consumption, which are frequently inaccurate. Similarly, [4] analyses photos every day with the help of specialist dietitians. There are methods for automatically recognizing food using cell phone cameras [5]–[12]. However, the process of measuring dietary characteristics is not completed. Recognition of food is followed by food volume estimate [11] and then calorie computation in additional approaches [13]–[15]. The volume measuring technique, on the other hand, is time-consuming and prone to mistakes. Second, basic food products are subjected to such algorithms. The algorithm is also used for nutritional analysis of food products [16], which makes it expensive and prevents it from being widely used in everyday life.

The proposed technique is meant to be a first step in increasing public awareness about health and fitness concerns so that people may eat and live healthier lives. The proposed method assists in automatically analyzing the nutritional content of food by allowing a person to learn about what food may include and how healthy it may be. The basic concept is to detect food items from a photograph of a plate and then estimate food attributes such as calcium, iron, and other nutrients, as well as the dish's components. Our method delivers nutrition information that are identical to those seen on packaged foods. Hospitals and the health-care business can adopt the recommended technique. Patients who are aware of the nutritional value of foods will be more motivated to avoid items that are hazardous to their health. Convolutional Neural Networks (CNN) [17], [18] have shown tremendous promise in visual object identification tasks, and CNNs are now being used to recognize food items in images [9], [10], [12]. We use CNNs in this study to get top 1 recognition accuracy rates of 85 percent. Another difficulty is accurately calculating the food's components and nutritional value [19]. Our objective is to reduce human input as much as feasible and automate this process as much as possible. For evaluating the components and food characteristics, we use deep neural networks. Our attention is drawn not just to characteristics such as protein, calcium, and vitamins, but also to the components included in the foods. Our technology calculates food characteristics, ingredients, and nutritional content automatically.

The suggested system's result is a smartphone app called "Rate Your Plate," which takes a photograph of food, recognizes it, and automatically shows the components and characteristics. Figure 1 depicts the proposed algorithm's automated estimation of food characteristics. The proposed system is divided into two sections. CNNs are used in the first component to identify the food item in a picture (Section III-B). The second component determines food features by scraping components and nutrient counts from nutritional and recipe websites, as well as collecting content from internet archives. Using this data, which was trained on a two-layer neural network, we can compute the probability of existing components in a particular food item. A standard serving size for any food item may be used to calculate calories and nutritional value. Deep learning techniques, a server with a trained model to detect food photos and estimate their characteristics as well as ingredients, and a traditional mobile application are all used in the system. For training purposes, the classifier requires a big dataset with numerous pictures for each food item category. Publicly available datasets such as the Food-101 dataset [7] and Image-net were used to help with this. We included sub-continental cuisine to Food-101 since we couldn't locate a dataset with sub-continental meals.

The following is a list of the work's main contributions.

- Convolutional Neural Networks were used to train a food identification engine.
- A sub-continental cuisine-specific expansion of the Food-101 dataset with well-defined training and validation classes.
- Using vector space embedding, a real-time estimation of food characteristics [20]. This module is built on data collected from the internet archives of several nutritional and recipe websites.

The rest of the paper is organized as follows: In Section II, we give a summary of the pertinent research. We describe our technique in Section III, and the results of our trials are shown in Section IV. The paper comes to a close with Section V.

Literature Review

Mobile devices are evolving at a fast pace. Each season, a new generation of mobile devices is introduced, each one more capable and computationally powerful than the previous. The fast growth of wireless internet technologies that promise high data rates and ubiquitous device connectivity, as well as mobile multimedia services and apps, have the potential to transform the healthcare sector. Numerous research on the influence of mobile applications on healthcare procedures have been done [21]–[23]. Similarly, research has been conducted on the use of social media for health-related reasons [24], [25]. A mobile revolution in health care is also being driven by personal health applications. We'll go through the various ways for monitoring food consumption in this section.

Manual dietary evaluations [2], [3], [26] to automated sensing methods are all used to measure food consumption. The automated imaging-based techniques are briefly discussed in this section. Pouladzadeh et al. [13] presented a method that involves taking a picture of the meal and processing it through a series of preset stages that follow a pipe line design. Food picture segmentation and food portion identification are two of these processes. Calorie counting is done with the use of nutritional information tables. The technology frequently fails to recognise different meal portions in mixed food, as well as segment them correctly. The suggested area measuring method is based on a depth estimate method. Their method, however, is based on a dataset that is far too basic, consisting of food items arrayed on flat white plates. Using a depth camera like the Kinect, Chen et al. [27] estimate the quantity of food for calorie evaluation. The algorithm, however, is unsuited for everyday use because it is based on Kinect. [28] Proposes a model-based approach to determining the size of meal portions. The approach consists of three steps: base plane localisation, food segmentation, and volume estimate. A 2D-3D model to picture registration approach is used to estimate volume. When it comes to shadows, reflections, complicated cuisine, ingredients, and motion blurring, the algorithm fails miserably. Similarly, Fang et al. [29] estimate the meal portion size using specific fiducially markers put in the scenes. Im2calorie [30] calculates food categories, ingredients, volume, and calories for specific servings. The calorie-annotated dataset utilised, however, is insufficient [31]. In the approaches listed above, the basic strategy to calorie estimating is to first recognise the food type, then estimate food portion sizes, and then estimate calories using standard nutritional fact tables.

Other methods for directly estimating calories from food pictures exist [31], [32]. Ege and Yanai [31] estimate food calories directly from images of food by learning about meal categories, components, and preparation instructions at the same time. They claim that learning categories, components, and calories at the same time would improve performance since there is a link between them. Various techniques for food recognition alone have also been presented. Two approaches for recognizing food are proposed by Ahmed and Ozeki [8]. Speed Up Robust Features (SURF) and Spatial Pyramid Matching (SPM). The first method (SURF) uses a dictionary of code words to generate histograms, which are then compared to those code words using a linear kernel classification algorithm. The latter (SPM) offers spatial information by splitting and subdividing the given image into increasingly smaller sub regions and creating histograms in each. Kawano and Yanai [6] describe a real-time mobile food identification system that continuously captures image frames from the camera device. Food recognition takes place within the boxes created by the user around the food items on the screen. The graph cut based segmentation algorithm Grab- Cut is used for precise food segmentation. The recognitions are carried out using the linear kernel SVM (support vector machine). The camera position and viewing orientation must be maintained in order to provide more reliable SVM classifications. Convolutional neural networks have also been employed to enhance recognition accuracy [30], [31], [33], and [35].

For machine learning-based food identification systems, having a big dataset is critical. Food-101 [7] is a big collection of food pictures that is freely distributed to the public. It has 1000 pictures for each of the 101 food categories. Similarly, UEC Food 100 [36] has 100 food picture categories. VIREO Food-172 [37] is a collection of 110,241 food pictures organized into 172 categories. Each meal picture has 300 ingredients individually labelled on it. [31] and [38] are two calorie-annotated datasets. There is no publicly accessible dataset that includes sub continental cuisine. As a result, we generated a new dataset that includes sub-continental as well as other prevalent cuisines. We utilize an entirely different technique for food attribute estimation, utilizing Word2Vec to represent words in vector space from a huge dataset. We acquired a significant quantity of text data from the internet, primarily from culinary, nutritional, and recipe websites, to achieve accurate and relevant findings from vector space embedding of words. When compared to milk and apple, semantically similar terms such as milk and yoghurt will appear adjacent in the vector space embedding. The aim is to compute food characteristics using a distance measure in vector space.

Methodology

The components that make up the proposed system are described in this section. The suggested system is made up of two key components:

- Food Recognition: Recognizing food items from pictures

- Attribute Estimation: Using a textual corpus, estimating food characteristics of the detected food item

This article also includes a description of the dataset utilized in this study.

Dataset

Our objective is to create a dataset that includes common culinary products as well as cuisines from the subcontinent. We began by playing with the freely accessible Food-101 [7] collection of food pictures. It has 1000 pictures for each of the 101 food categories. Food-101 was created with multi-class categorization in mind. There are additional datasets that have been used for food recognition in the past, including Food-5k. Food-5k has 5000 pictures, with 2500 photographs of food and 2500 images of non-food. This dataset, however, is inappropriate for our task because it can only be utilised for binary classification to separate food from non-food items. Furthermore, Food- 101 ignores food items and classes from sub-continental cuisine, which represents for a significant portion of the subcontinent's food consumption. Because certain sub-continental meals have little inter-class variance and are quite similar to one another, gathering high-quality data for correct categorization is a major issue. The results for textual search searches for food photographs returned by Google are very relevant and include very little noise. Our new dataset is created by conducting a Google search against each of our dataset's labels, depending on the results of the Google search engine. In the newly generated dataset, Food-101 courses are widely used and eaten. As a consequence, the final dataset contains all of the food items from Cuisine-101 as well as 100 new sub-continental food classes. The images in the dataset are separated into two groups: training and validation. Each lesson contains about 800 training photographs and 200 validation photos.

Food Recognition

Food Recognition is the process of recognizing a food item from a picture. Due to the exceptional performance of CNNs, we attempted to train our dataset using transfer learning using top performing pre-trained models. Based on their success in other domains, we employed pre-trained CNN models such as VGG-16, VGG-19 [18], Inception-v3 [39], Inception-v4 [40], and ResNet [41]. These models are pre-trained using the Image net dataset. Transfer learning is used to train these models on our dataset. Dropout, Re LU activations, and softmax layers are added after the previous completely connected layer is removed. Fine tuning the model on our dataset took about 15 hours on a single Titan X GPU with 12GB of RAM. The rest of the study's identification engine was based on the Inception-v3 and Inception-v4 models, which performed better.

Fine Tuning

Based on their performance on our dataset, we picked Inception-v3 and Inception-v4 as the appropriate models, and after initial filtering, we continued to enhance the model's validation accuracy on our dataset. This was accomplished using a number of techniques. To begin, substantial data augmentation is carried out to guarantee that the model is as robust to affine changes as possible and that the images are trained as rapidly as possible. To create copies that are changed from the original image, various transformations are done to each photo in the dataset in each epoch, using random values given by parametric range. Among these are translations, rotations, shearing, zooming, and flipping. Using the data augmentation parameters given in Table 1 yields the best results. Images are also rescaled to cater for the model's need to analyse large RGB coefficient values. The RGB value was multiplied by 1/255, resulting in normalized target values that range from 0 to 1. This improves the image's resistance to light fluctuations and speeds up data processing.

Batch normalization and regularization, as well as multi-crop assessment, are further actions that may be taken to improve accuracy. In multi-crop assessment, several crops of an image are obtained from different parts of the picture during testing, and each crop is assessed separately. In the list of anticipated labels, the most often occurring class is taken into account. Four crops plus a central crop are generated in this work by evenly splitting the image into four squares. After that, the picture is flipped and the same procedure is followed. As a result, each image receives ten crops in total. In addition, learning rates and decay values are adjusted for more precise outcomes. Early stopping saves time since it ends training if the model's validation accuracy does not increase after a certain number of continuous epochs. The final two convolution blocks of the Inception model, as well as the final fully connected layer, are also made trainable, allowing us to learn more high-level characteristics relevant to our dataset.

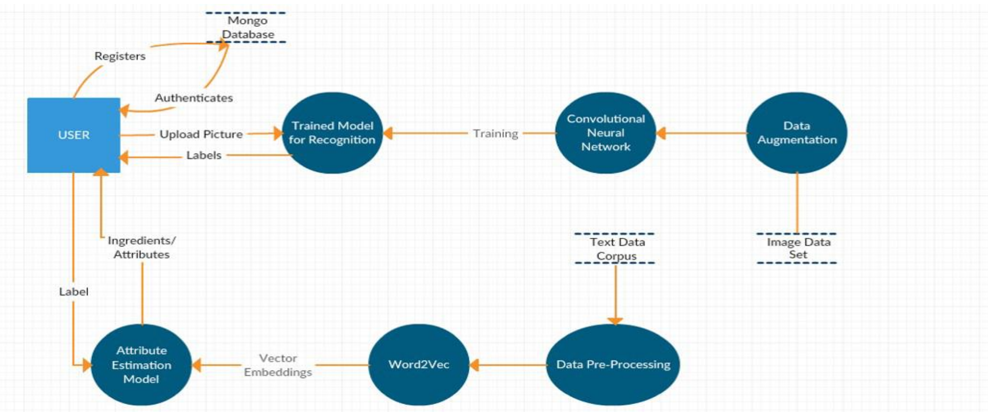

Figure 2. The suggested system's System Flow

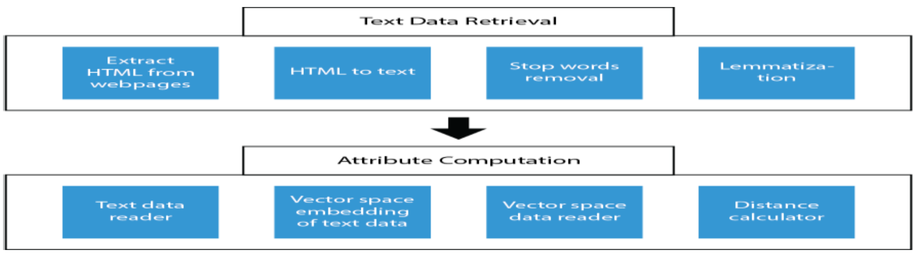

Figure 3. Ingredient Attribute Estimation Using Vector Space Embedding of Textual Data Retrieved from the Internet

Attribute Estimation

Following food recognition, the next step is to compute food properties such as ingredients and nutritional value. As indicated in Section II, there are two techniques that are commonly utilized. In the first technique, the meal portion size is calculated, and then the characteristics are computed using conventional nutritional tables [13], [27], [30]. The food characteristics are learnt directly from the picture in the second technique [31], [32]. We use a completely new method and represent words in vector space using data from a huge dataset [20]. In order to obtain accurate and meaningful conclusions from vector space embedding of words, we collected a large amount of text data from the internet, particularly from food and nutrition sites. The data is then trained using Word2Vec [20], which produces word embedding. Words that are synthetically and semantically similar are near or have a shorter distance in the vector space than words that are unrelated. To detect food features, the aim is to compute the distance between the food item and the constituents in the learned vector space. A food item's closeness to an attribute in the original text suggests a high chance of that characteristic being present in the food item. The entire flow is depicted in Figure 3. is more relevant since it returns sites that correspond to precise labels, but Common-content Crawl's is more generic.

The HTML format is used to store the raw text data acquired from Google and Common Crawl. HTML tags, CSS, JavaScript code, and comments have been removed during the preprocessing process. To extract individual words, such meta-data is stripped from the text and tokenized. This makes the data readable and allows for the processing of individual words. The data must be devoid of auxiliary and unnecessary terms like 'this,' 'are,' and 'an' for semantic-based text analysis. These terms have the greatest recurrence probability and should thus be eliminated from the text data so that learning of important data is not hampered. Stop words like these are deleted. Text can be written in a variety of ways, including past tense, future tense, plural, and so on. The likelihood of semantically identical terms is divided into numerous variations as a result of this. To remedy this issue, the text has been lemmatized. Lemmatization is the process of matching a phrase to its base word in a dictionary. For example, the words 'eating' and 'tomatoes' are lemmatized to 'eat' and 'tomato.' The word corpus that results is subsequently used for training purposes.

Table 1. For data augmentation, the parameters utilized and their values

Parameter	Value
Width Shift	0.2
Height Shift	0.2
Rotation	90
Rescale	1/255
Shear	0.2
Zoom	0.2
Horizontal Flip	True
Vertical Flip	True

Table 2. Word2Vec configuration for word embedding training

Parameter	Value
Window Size	8
Vector Dimensionality	200
Sample	1e-4
Negative	25
Iterations	15

Table 3. Food identification method accuracies on our dataset and the Food101 dataset

		Accuracy							
		Top1		Top2		Top3		Top5	
Models		Own DS	Food101 DS	Own DS	Food101 DS	Own DS	Food101 DS	Own DS	Food101 DS
Inception-v3	Single Crop	79.8	80	87.9	80.9	91.6	84.8	95	89.6
Inception-v4	Multi Crop	81.81	72.25	-	-	-	-	95.51	89.53
Inception-v3	Single Crop	83.8	78.3	89.8	85.2	92.4	87.9	94.7	90.5
Inception-v4	Multi Crop	85	79.22	-	-	-	-	95.67	90.63

Training and Vector Embeddings

Word2Vec effectively calculates vector representations of words in a multi-dimensional space using deep learning algorithms. The log-linear classifier learns words that exist within a specific range of the input, taking into account the fact that words that appear further away from the input may have a lower likelihood of being semantically related. Each word in the text corpus is used as an input by the log-linear classifier, which learns words that exist within a specific range of the input, taking into account the fact that words that appear further away from the input may have a lower likelihood of being semantically related. It employs two architectures for generating vector representations from corpora: Continuous Bag of Words (CBOW) and Skip-Gram Model (SGM). We've used the CBOW technique in our work. The proposed approach detects semantic links in the embedded vector space by using a cosine distance metric between word vectors. We utilised the Google implementation of Word2Vec in our research. Word2Vec training can take several forms, including the architecture used, the dimensions of vector space used, the size of the window used to learn word occurrences, the training method used, such as softmax or negative sampling, the threshold for down-sampling, and so on. After testing with various settings, the most relevant conclusions are obtained using the values presented in Table 2. Word2Vec takes in the whole text corpus, creates a vocabulary of words used in that text, learns the distances between those words, and outputs a binary file with learned vector embedding.

Attribute Extraction

To assess the chance of occurrence of a food attribute in a food item, we compute the distance between the food attribute and the food label in the vector space learned using Word2Vec. A user may be interested in ingredients, nutritional values, or features, depending on the application. A list of these properties is produced in a static format. In the lists, we attempted to include all conceivable and important qualities that can occur in a food item. Following the recognition of the food item, the predicted food label and food attributes are sent to Word2Vec, which calculates the cosine distance in vector space between the predicted label and the food attributes. The Word2Vec module's Similarity function, based on cosine distance, returns the likelihood of one word occurring inside the window size of another word. Following the discovery of each attribute's resemblance to the anticipated label, we calculate its estimated chance of being in the text other than the predicted label, i.e. its likelihood of existing in the food item. Text has multiple scales of ingredients, nutrition, and attributes. Ingredients have the highest frequency, thus their sample size is the largest, making them the most accurate. Because the text is lacking in nutrition values and characteristics appear seldom, these are less reliable. The probabilities of each of these categories are normalised by dividing them by the category with the highest probability. We can obtain a more realistic image thanks to this localization.

Results and Discussions

Food Recognition

As mentioned in Section III-B, CNN models based on Inception-v3 and Inception-v4 outperform the other models investigated. These models are fine-tuned on our dataset as well as Food-101 for comparison. The recognition accuracy results for our dataset and the Food-101 dataset are shown in Table 3. The collection contains 200 culinary categories, including sub-continental cuisine groupings. The model is trained on 68705 training photos and tested on 5284 validation images without taking into consideration data augment. CNN models based on Inception-v3 and Inception-v4 outperform the other models examined, as indicated in Section III-B. For comparison, these models are fine-tuned on our dataset as well as Food-101. Table 3 shows the recognition accuracy results for our dataset and the Food-101 dataset. There are 200 culinary categories in the collection, including sub-continental food groups. Without taking into account data augment, the model is trained on 68705 training photographs and evaluated on 5284 validation shots.

Table 3 from the Food-101 Dataset demonstrates the accuracy of our technique. Culinary-101 is a popular series of 101 culinary tutorials on the Food Spotting website. With no extra data, the models are trained on the 90900 training photos and tested on the 10100 validation images. After 40 epoch, the model accuracy levelled out, as seen in Figure 4.

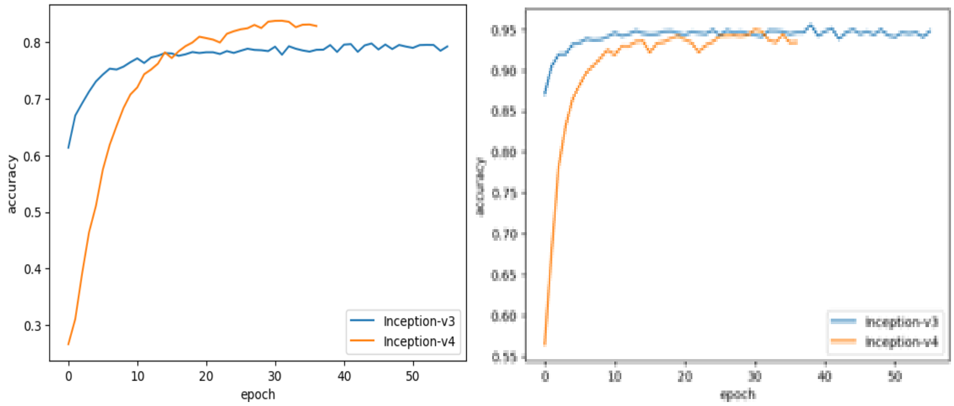

Figure 4. Accuracy vs. Epoch Graph on our dataset with Inception-v3 and Inception-v4. (a) Top-1. (b) Top-5.

Conclusions

This study describes a system that takes use of the widespread usage of mobile devices to deliver health information on the foods we consume. The software uses a photo of the meal to display estimated ingredients and nutritional values in the dish. A dataset containing common and sub continental food items is developed. To recognize food items, we use a fine-tuned Inception model and offer a technique for estimating the detected food item's characteristics. Data augmentation, multi-crop assessment, regularization, and other similar approaches are used to enhance the findings. Our dataset has an accuracy of 85 percent. Our proposed approach for estimating characteristics yielded promising results as well. Future endeavors in this sector might involve putting this work into practice and improving the android app with additional capabilities to make it a full guide for everyday meals.

References

1. M. F. Munsell, B. L. Sprague, D. A. Berry, G. Chisholm, and A. Trentham-Dietz, "Body mass index and breast cancer risk according to post-menopausal estrogen-progestin use and hormone receptor status," *Epidemiol. Rev.*, vol. 36, no. 1, pp. 114–136, 2014.
2. *Myfitnesspal.* Accessed: Oct. 1, 2018. [Online]. Available: https://www.myfitnesspal.com/
3. *Samsung Health.* Accessed: Oct. 1, 2018. [Online]. Available: https://health.apps.samsung.com/
4. *Rise*. Accessed: Oct. 1, 2018. [Online]. Available: https://www.rise.us/
5. M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney, "Recognition and volume estimation, of food intake using a mobile device," in *Proc. Workshop Appl. Comput. Vis. (WACV)*, 2009, pp. 1–8.
6. 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.
7. L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101—Mining discriminative components with random forests," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zürich, Switzerland: Springer, 2014, pp. 446–461.

8. A. Ahmed and T. Ozeki, "Food image recognition by using bag-of-surf features and hog features," in *Proc. 3rd Int. Conf. Hum.-Agent Interact.*, 2015, pp. 179–180.
9. 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 Man- age.*, 2016, pp. 41–49.
10. N. Martinel, G. L. Foresti, and C. Micheloni. (2016). "Wide- slice residual networks for food recognition." [Online]. Available: https://arxiv.org/abs/1612.06543
11. S. Fang, F. Zhu, C. Jiang, S. Zhang, C. J. Boushey, and E. J. Delp, "A comparison of food portion size estimation using geometric models and depth images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 26–30.
12. P. Pouladzadeh and S. Shirmohammadi, "Mobile multi-food recognition using deep learning," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 13, no. 3s, p. 36, 2017.
13. 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.
14. F. Zhu *et al.*, "The use of mobile devices in aiding dietary assessment and evaluation," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 4, pp. 756–766, Aug. 2010.
15. W. Zhang, Q. Yu, B. Siddiquie, A. Divakaran, and H. Sawhney, "snap- n-eat'food Recognit. nutrition estimation a smartphone," *J. Diabetes Sci. Technol.*, vol. 9, no. 3, pp. 525–533, 2015.
16. J. Noronha, E. Hysen, H. Zhang, and K. Z. Gajos, "Platemate: Crowd-sourcing nutritional analysis from food photographs," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, 2011, pp. 1–12.
17. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Pro- cess. Syst.*, 2012, pp. 1097–1105.
18. K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
19. F. Cordeiro, E. Bales, E. Cherry, and J. Fogarty, "Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based cap- ture," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, 2015, pp. 3207–3216.
20. T. Mikolov, K. Chen, G. Corrado, and J. Dean. (2013). "Efficient esti- mation of word representations in vector space." [Online]. Available: https://arxiv.org/abs/1301.3781
21. V. Vodopivec-Jamsek, T. de Jongh, I. Gurol-Urganci, R. Atun, and J. Car, "Mobile phone messaging for preventive health care," *Cochrane Database Syst. Rev.*, vol. 12, p. CD007457, Oct. 2012.
22. G. Nasi, M. Cucciniello, and C. Guerrazzi, "The role of mobile technolo- gies in health care processes: The case of cancer supportive care," *J. Med. Internet Res.*, vol. 17, no. 2, 2015.
23. M. S. Marcolino, J. A. Q. Oliveira, M. D'Agostino, A. L. Ribeiro, M. B. M. Alkmim, and D. Novillo-Ortiz, "The impact of mhealth interven- tions:

Systematic review of systematic reviews," *JMIR mHealth uHealth*, vol. 6, no. 1, 2018.

24. S. A. Moorhead, D. E. Hazlett, L. Harrison, J. K. Carroll, A. Irwin, and C. Hoving, "A new dimension of health care: Systematic review of the uses, benefits, and limitations of social media for health communication," *J. Med. Internet Res.*, vol. 15, no. 4, 2013.
25. E. Hagg, V. S. Dahinten, and L. M. Currie, "The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review," *Int. J. Med. Informat.*, vol. 115, pp. 92–105, Jul. 2018.
26. J.-E. Lee, S. Song, J. S. Ahn, Y. Kim, and J. E. Lee, "Use of a mobile application for self-monitoring dietary intake: Feasibility test and an intervention study," *Nutrients*, vol. 9, no. 7, p. 748, 2017.
27. M.-Y. Chen *et al.*, "Automatic Chinese food identification and quantity estimation," in *Proc. Tech. Briefs SIGGRAPH Asia*, 2012, p. 29.
28. H.-C. Chen *et al.*, "Model-based measurement of food portion size for image-based dietary assessment using 3D/2D registration," *Meas. Sci. Technol.*, vol. 24, no. 10, p. 105701, 2013.
29. S. Fang, F. Zhu, C. J. Boushey, and E. J. Delp, "The use of co-occurrence patterns in single image based food portion estimation," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 462–466.
30. A. Meyers *et al.*, "Im2Calories: Towards an automated mobile vision food diary," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1233–1241.
31. T. Ege and K. Yanai, "Image-based food calorie estimation using knowl- edge on food categories, ingredients and cooking directions," in *Proc. Thematic Workshops*, 2017, pp. 367–375.
32. T. Miyazaki, G. C. de Silva, and K. Aizawa, "Image-based calorie content estimation for dietary assessment," in *Proc. IEEE Int. Symp. Multime- dia (ISM)*, Dec. 2011, pp. 363–368.
33. K. Yanai and Y. Kawano, "Food image recognition using deep con- volutional network with pre-training and fine-tuning," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jun./Jul. 2015, pp. 1–6.
34. Y. Cui, F. Zhou, J. Wang, X. Liu, Y. Lin, and S. Belongie, "Kernel pooling for convolutional neural networks," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 3049–3058.
35. C. Liu *et al.*, "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, Mar. 2018.
36. Y. Matsuda, H. Hoashi, and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," in *Proc. IEEE Int. Conf. Multi- media Expo (ICME)*, Jul. 2012, pp. 25–30.
37. J. Chen and C.-W. Ngo, "Deep-based ingredient recognition for cooking recipe retrieval," in *Proc. ACM Multimedia Conf.*, 2016, pp. 32–41.
38. *Allrecipes*. Accessed: Oct. 1, 2018. [Online]. Available: http://allrecipes.com/
39. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
40. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI*,

2017, pp. 4278–4284.

41. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.