Image Processing Model to Estimate Nutritional Values in Raw and Cooked Vegetables

Tan Jo Yen
NielsenIQ
Kuala Lumpur, Malaysia
joyen.tan@nielseniq.com

Sivakumar Vengusamy
Asia Pacific University of Technology and Innovation
Kuala Lumpur, Malaysia
dr.sivakumar@staffemail.apu.edu.my

Fabio Caraffini
Swansea University
Swansea, UK
fabio.caraffini@swansea.ac.uk

Stefan Kuhn
University of Tartu
Tartu, Estonia
stefan.kuhn@ut.ee

Simon Colreavy-Donnelly
University of Limerick
Limerick, Ireland
simon.colreavy@ul.ie

Abstract—**The availability of high-calorie foods with contentious nutritional content has led to a worldwide increase in chronic disease. Therefore, monitoring of eating habits and practising healthy eating habits is recommended. Clinical diet assessment methods and mobile calorie tracking apps can be used to record daily food consumption but are often not user-friendly. Convenient image-based assessment models are currently available to recognise and estimate the nutritional value of foods directly from food images, but they do not consider how nutritional value changes after cooking. Consequently, *VegeNet*, a multi-output InceptionV3-based convolutional neural network model has been developed, which estimates the nutritional values of cooked and uncooked vegetables. The explicit use of the cooking state is the main contribution of this work. This deep learning model successfully classifies the food images at 97% accuracy and estimates the nutritional values at 15.30% mean relative error, making it suitable as a visual-based added food assessment solution. This can help users save time and avoid under-reporting problems.**

I. Introduction

Food is a cornerstone of a healthy lifestyle. It is the energy source of living things that is needed to sustain and perform various activities. Today, humans are exposed to heavily processed foods that are more flavorful and convenient but less healthy [1]. Overconsumption of these foods leads to chronic diseases such as obesity, diabetes, and cardiovascular disease that have become more common and are now prevalent among young people [2, p. 81]. According to the World Health Organisation (WHO), worldwide obesity has increased by a factor of three each year since 1975: more than 650 million adults were diagnosed as obese in 2016, while 38 million children under the age of 5 were overweight or obese in 2019 [3]. Therefore, the WHO recommends practising healthy diets to overcome the mentioned problems and prevent diseases [4].

In order to improve physical health, dietitians need a better understanding of a person's dietary behaviour. For this, they use dietary assessment methods such as real-time recording, 24-hour nutritional recall, dietary history, and food frequency questionnaires [5]. Recently, calorie-tracking apps [6] have been developed for smartphone users to record foods eaten in meal-by-meal form. This allows them to calculate and track their food consumption patterns and provides personalised feedback in real-time and suggests diet goals to help users achieve their dieting goals better.

Some advantages of calorie-tracking apps are their systematic and data-centred structure. However, they are often counterintuitive to use, specifically for compulsive eaters, and are subject to under-reporting problems, as they mostly require the users to enter the foods eaten manually by the correct names and volumes. Users must find the closest items and weigh each ingredient to obtain the most accurate calorie and nutritional values. They may also tend to select suggestions with lower calories, leading to under-reporting and under-estimation of nutrient intake [7], [8]. To alleviate the problems mentioned above, this article introduces an image-based dietary assessment model that estimates the nutritional values of cooked versus uncooked vegetables, directly from the input image. It does this through deep learning techniques to perform food type and cooking method classification, as well as food weight estimation, by which the predicted outputs are taken to estimate nutritional values based on official food data. The explicit distinction of the cooking state of food was, to our knowledge, not previously done.

This work focuses on image analysis to differentiate between raw vegetables and vegetables cooked in different ways, because the weights, water content, and other nutritional values of the natural products change when cooked [9]. We focus on vegetables in this work since there is a wide range of cooking techniques for them, including uncooked consumption. Therefore, the main research question of this paper is if and to what degree consideration of the cooking state of vegetables can improve nutritional value estimation. The developed model adds value to existing visual-based food assessment technology, considering the nature of vegetables in both raw and cooked conditions, and can be deployed in mobile applications for more efficient and accurate diet monitoring. As a more convenient dietary assessment method, this can encourage people to track the nutrients consumed consistently, helping them better understand their own diet habits and choose healthier foods or cooking methods to maintain body health.

The remainder of this article is organised as follows:

- Section II reviews related works on image-based food recognition and calorie estimation;
- Section III explains the methods and datasets used;
- Section IV discusses the implementation steps and evaluation results;
- Section V discusses project limitation and suggestions for future work;
- Section VI concludes the article.

II. Related Work

Data scientists and researchers have developed image-based dietary assessment models using machine learning and deep learning approaches, of which the model functions can be categorised into two types, food recognition and estimation of food nutritional values.

A. Food recognition

Prior to the popular adoption of deep learning algorithms for image classification tasks, in particular, convolutional neural networks (CNNs), Support Vector Machines (SVMs) had been a promising machine learning approach used in the field of image data analysis to perform image classification and predictions. A group of researchers [10] used SVM as a classifier to identify food and fruits for calories estimations using the colour, size, and shape properties of the images, while [11] improved the SVM model by adding a Gabor filter to take texture segmentation as input features for more accurate food recognition. However, the SVM performs well only when the dataset is small [12], while the deep learning approach outperforms the SVMs due to its ability to automatically select features [13] and analyse a large image dataset as observed in [14]. At the present time, CNNs have been widely used for image classification tasks. CNNs, which started with AlexNet [15], are currently the de-facto standard in computer vision and image processing. ZFNet [16], VGGNet [17], GoogLeNet [18], and ResNet [19] are some examples of successful CNN architectures. InceptionV3 [20] is based on GoogLeNet. MobileNets [21] are optimized to give good results with low complexity.

A new CNN model for image recognition requires very large datasets for training, and this can take a long time to accumulate. Furthermore, these models are highly affected by overfitting issues [22], [23]. These problems can be solved using pre-trained networks [24] as the feature learning layer, with custom classification layers for specific image recognition tasks or, similarly, fine-tuning the pre-trained model, for both feature extraction and classification tasks. The researchers in [25] showed that a fine-tuned pre-trained Deep CNN (DCNN) model along with the SVM classifier performed better than the conventional SVM model and the DCNN model without fine-tuning. Similarly, [26] fine-tuned ResNet-50 for food classification in the Food-475 dataset, while [22] fine-tuned classification layers in the pre-trained InceptionV3 network for the image datasets ETH Food-101, UEC-Food 100, and UEC-Food 256. Moreover, [27] fine-tuned the MobileNet architecture [21] by replacing the average pooling layer and fully connected layer with global average pooling and batch normalisation layers to resolve the overfitting problem in MobileNet.

Initial models only recognised single food items in an image. To overcome this limitation, the Selective Search algorithm and Map Reduce were used by [28] to segment food regions before being processed with CNN. The researchers in [29] developed a CNN model that could create bounding boxes around each food item in the image and estimate the calories of the respective foods, by adding pseudo-bounding boxes to the dataset of the annotated image with calories while training the model. In particular, [30] developed a model that performs pixel-level classification rather than image classification with fully convolutional networks (FCN), to identify multiple foods in an image, but this model was trained with fake food image datasets, which is less representative of real food images. Rather than inter-class food recognition, [31] developed a CNN model that recognises intra-class dishes through the ingredients and cooking methods classifications. WISER [32] is a deep neural network with two branches that can learn more features from vertical layers of food images to recognise variances within food classes, while [33] integrated a superpixel-based mid-level feature extraction approach and a DCNN to extract image features, then adopted SVM as a classifier for food recognition within a class.

B. Food nutritional values estimation

Estimation of food nutritional values is based on food recognition, discussed in II-A, but goes beyond it and is crucial to improve the practicality of visual-based dietary assessment models. In [34], [35], CNN models were trained using food image datasets obtained from a school lunch blog and online recipes, in which calorie values are provided. This method assumes that the serving sizes of each labelled photo are for one person, which restricts its use and is subject to over- or underestimation, as the portion of foods eaten differs among individuals. As such, estimating food volume is critical for accurate estimation of food nutritional value, but this can be very challenging. Some researchers used reference objects for calibration to estimate the volume of food. The work in [11] took the user's thumb for calibration to compute food volumes based on area and depth information collected. Variously, in [36] preregistered reference objects of known size are used, while [37] used rice corns as a reference. [38] used a 5×5 PVC chequerboard card as a size reference to estimate food volume. However, this method requires reference objects to be always available whenever a food image needs to be captured and analysed. Segmentation techniques were adopted to segment food regions to help estimate nutritional values. K-means segmentation technique and GrabCut algorithms were used in [36] to segment "food", "dish", and "background", while [39] developed a CNN model for food region segmentation to produce border maps with pixel values ranging from 0 to 1, where values closer to 1 were recognised as pixels closer to the border. [40] demonstrated image thresholding and K-means++ clustering as segmentation techniques to extract food regions from the images. In contrast, [41] directly estimates the weights of the food and the nutritional values, with the help of a divided food tray provided by hospitals. The divided food tray allows the algorithm to easily identify the food area and hence estimate food weights.

CNNs have been widely adopted to perform food classification and calorie estimations on a variety of food image datasets and have been achieving promising results. However, it is observed that most articles do not consider the cooking status of the identified foods, whereas this concept is addressed in this article, which distinguishes between raw and cooked vegetables, a fundamental concept in calorie and nutrition estimation.

III. Materials and Methods

To identify raw and cooked vegetables and estimate their nutritional value, this paper proposes a multi-output CNN model, VegeNet which classifies the nutrition information data from the food type, the cooking methods, and estimates the weights of the foods in the images. Subsequently, the predicted values are taken to calculate the estimated nutritional values based on the US Department of Agriculture (USDA).

A. The VegeNet network

Transfer learning approaches are the key to solving many drawbacks in deep learning applications [42], [43], and are currently being studied and developed continuously [44]. Since we have a dataset with 7754 images, we also adopt transfer learning to avoid overfitting issues. It should be noted that in other work, e.g. [26] hundreds of thousands of images are common. *VegeNet* utilises pretrained InceptionV3 as the feature learning layer, whereby the InceptionV3 output is taken as input for the three branches to predict the type of food, the cooking method, and the weight of the food images (see Figure 1). The network consisting of the InceptionV3 network and the three branches can be described as a *InceptionV3 multi-output CNN* and is called VegeNet hereinafter. Generally, each of the three branches is made up of a flattening layer, a dense layer (128 neurons, RELU activation), a batch normalisation layer, and a dense layer as the final output layer. However, dropout layers are included in the food type and weight branch before the output dense layer, but not in the cooking method branch, because we found this to work best. This approach is taken to regularise the model whilst maintaining high prediction accuracy. In the output dense layers, 'Softmax' activation is used for food type and cooking method classification whereas 'linear' activation is used for weight estimation. The training parameters are summarised in Table I.

The network was designed by combining InceptionV3, a state-of-the-art pretrained network with a reasonable footprint, with individual branches for the three desired outputs. Those branches use standard elements of dense layers in CNN design. The CNN is not specific to food classification but is a proven architecture for similar tasks. The suitability of the overall approach is shown by the results which we will report later.

TABLE I. *VegeNet* training parameters

Parameter	Value
Adam optimiser	learning rate: 0.0001 decay: 0.000001
Training epochs	100
Batch size	64 (training) or 32 (validation)
Image data generator batch size	32 or 16
Training loss	Food type: *categorical crossentropy* Cooking method: *categorical crossentropy* Weight: *Mean Square Error* (MSE)
Loss weights	Food type: 1 Cooking method: 1.5 Weight: 4
Metrics	Food type: *Accuracy* Cooking method: *Accuracy* Weight: *Mean Absolute Error*

These predicted food types, cooking methods, and weight values produced by the CNN model are concatenated and linked to the nutrition information dataset to estimate the calories and macro-nutrients of vegetables. As the nutrition information data presented in the USDA dataset are per 100g of food weight, the nutritional values are calculated using $100g$ as the base weight and are multiplied by the estimated food weight, as illustrated in Equation 1.

$$N_i = \frac{W_{i-\text{estimated}} \cdot N_{\text{USDA}}}{100} \tag{1}$$

where N_i is the estimated nutritional value of the captured vegetable (calories and the macro-nutrients); $W_{i-\text{estimated}}$ is the estimated weight of the vegetables in the analysed image; and N_{USDA} indicates the nutritional value of the vegetable in the base weight $100g$ found in the USDA dataset.

B. Dataset

In this work, a food image dataset and a nutrition information dataset are required. The food image dataset is made up of 8 vegetable categories and 1 non-vegetable category. Vegetable images are collected primarily by preparing 8 types of vegetables (broccoli, red and green cabbages, carrots, cauliflower, corn, cucumber, lettuce), which are cut into multiple shapes and cooked using 5 methods (uncooked, cooked - no fat, roasted - no fat, roasted - with fat, cooked - with fat). The number of images for the cooking types varies for the different vegetables. This due to cooking habits and availability of images (e.g. roasting cabbage is not common, hence is not included). There are more images of uncooked vegetables since the main objective of the model is to distinguish between raw and cooked. The vegetables are prepared in different appearances, weighed, and captured from multiple angles and heights, which are then preprocessed through a 3-fold Canny edge detection and image cropping technique [45]. This newly created gallery of images is one of the contributions of this investigation and is made available in [46]. Furthermore, to help the model recognise non-vegetable foods, a non-vegetable

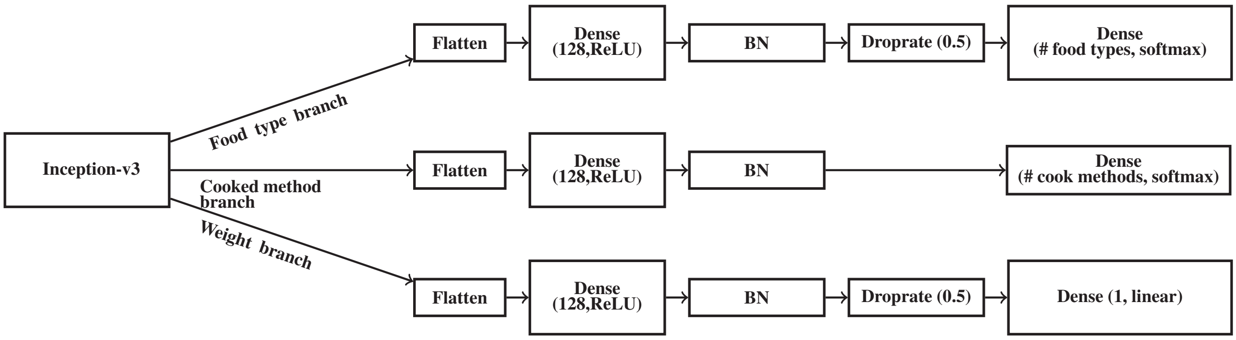


Fig. 1. Network architecture of *VegeNet*

category is added to the image dataset by sample images from hospital food images used in [41]. Overall, there are 5558 vegetable images and 1112 non-vegetable images. After preprocessing, there were 6754 images overall.

The weights of all foods are recorded as an attribute for model training. In addition, data augmentation is done to randomly generate 1000 images. The commands used from the Python Augmentor library are displayed in Figure 2.

Together with the 6754 original images, we have a dataset comprising 7754 images. Meanwhile, a duplicated set of food images is prepared by adding low-res mask filtering to all 7754 in order to study the effectiveness of added image filtering on the model classification accuracy. The nutritional information is obtained from [47], of which the nutrition information of the vegetables used is extracted. Data preprocessing, such as removing unwanted attributes and adding cooking method classes through average ratio calculations, is performed to match the food classes to the image dataset. The final USDA nutrition information dataset consists of 5 attributes (food_type, cook_method, energy (kcal), protein (g), carbohydrate (g), and fat (g)) and 32 actual categories out of 9 food types and 5 cooking methods (Table II).

For the sake of reproducibility, we not only made all images used in this study available in the repository [46], but we also deposited the source code necessary to reproduce the results produced by this piece of research. Note that of the 7754 images in the dataset, 6754 are randomly selected to train the model, and 1000 images are used for testing. Both sets will contain some of the 1000 images generated by data augmentation.

C. Implementation steps

Images are resized to 299×299 and fed into the InceptionV3 multi-output CNN model for model training and prediction. In addition to the original images, the duplicated dataset with added low-res mask filtering is also used for model training and testing. The model training and validation processes are performed in the Keras environment. The parameter settings for model training are provided in Table I, and the Keras EarlyStopping function is called to stop the model training process when the models do not improve for 10 epochs.

D. Graphical user interface (GUI)

To demonstrate the practicality of the model, a user interface is developed that allows users to select the images to be predicted and print the estimated nutritional values by loading the CNN model as a predictor and the nutrition information database for the retrieval of nutritional values. Figure 3 shows the process flow of the nutritional value estimation application, and Figure 4 shows the functions of this GUI. The GUI is intended to deliver a good user experience. Partly this is due to the fast execution of the model for prediction purposes, which our model shares with AI models in general, which normally execute quickly once trained.

IV. Results

The proposed model involves three predictions, food type classification, cooking method classification, and weight estimation, which are used to estimate the nutritional value of the food. Therefore, the validation and testing stage of the model involves both evaluation metrics for classification and estimation performances.

The results of the model evaluation are shown in Figures 5a to 5c and Tables III and IV.

The evaluation results show that VegeNet achieves 100% and 97% accuracy for food type and cooking method classifications, respectively (see Table III). Furthermore, the model estimates the weights of the foods and the nutritional values of the foods at around 16% mean relative errors. Specifically, Table IV shows that VegeNet achieves 13.86g of mean absolute error and 15.60% mean relative error in weight prediction. As nutritional values are calculated using USDA nutrition information at 100g base weight, a low error in weight prediction leads to a low error in nutritional value estimation. It is also concluded that the outstanding performance of VegeNet is mainly due to the adoption of the InceptionV3 pretrained network as the base convolutional neural network, which substantially enhances the feature learning ability of the model, performing better than the VGG16 and ResNet-50 model used by [34], [41]. VegeNet trained and tested with (USM) low-res mask-filtered images does not show significantly different results. Considering the quality of the results, this is not necessarily a sign that such filtering is generally not helpful,

```
augment.rotate(probability=0.3,max_left_rotation=10,max_right_rotation=10)
augment.shear(probability=0.3,max_shear_left=10,max_shear_right=10)
augment.flip_left_right(probability=0.3)
augment.random_brightness(probability=0.3,min_factor=0.4,max_factor=0.9)
augment.skew(probability=0.3)
```

Fig. 2. Command used for the data augmentation process with the Python Augmentor library

[ht]

TABLE II. SAMPLE OF FOOD IMAGES COLLECTED. THERE ARE EIGHT VEGETABLES, WHICH ALL CAN OCCUR WHITH UP TO FIVE COOKING METHODS, AND NON-VEGETABLES, FOR WHICH WE DO NOT DETERMINE AN ACCURATE COOKING METHOD

Food Type	Cooking Method	Calories ($Kcal$)	Protein (g)	Carbs (g)	Fat (g)
Lettuce	uncooked	15	1.36	2.87	0.15
	cooked - no fat	15.47	1.4	2.96	0.15
	cooked - with fat	43	1.3	3.33	3.15
	roasted - no fat	21.43	1.94	4.1	0.21
	roasted - with fat	36.82	1.89	3.98	2.91
Broccoli	uncooked	34	2.82	6.64	0.37
Cauliflower	uncooked	25	1.92	4.97	0.28
Green cabbage	cooked - no fat	26	1.33	6.02	0.1
Red cabbage	uncooked	31	1.43	7.37	0.16
Corn	uncooked	86	3.27	1.87	1.35
Carrot	uncooked	41	0.93	9.58	0.24
Cucumber	uncooked	15	0.65	3.63	0.11
Non-vegetable	other	0	0	0	0

TABLE III. MODEL CLASSIFICATION PERFORMANCE

		VegeNet	VegeNet USM filtered
Food type	Accuracy	1.0	1.0
	Precision	1.0	1.0
	Recall	1.0	1.0
Cooking method	Accuracy	0.97	0.97
	Precision	0.97	0.97
	Recall	0.97	0.97

TABLE IV. WEIGHT AND NUTRITIONAL VALUES ESTIMATION PERFORMANCE

			VegeNet	VegeNet USM filtered
Weight	MAE (g)		13.86	15.80
	MRE (%)		15.60	18.11
Nutrional values	MAE	Cal	4.73	5.61
		Carb (g)	0.80	0.95
		Prot (g)	0.19	0.22
		Fot (g)	0.21	0.25
	MRE (%)		15.30	17.76

but an indication that the results do not leave room for improvements.

V. DISCUSSION

A. Contribution

To the best of our knowledge, this is the first visual-based dietary assessment model that distinguishes the nutritional content of cooked and uncooked vegetables. This work is essential, as vegetables are one of the main sources of nutrients for a healthy diet, but nutritional values differ when the vegetable is cooked in different ways. In addition to calorie values, the model also estimates the values of the macro-nutrients in vegetables - carbohydrate, protein, and fat. This is because calorie value is just the total energy consumed yet the macro-nutrients are more important to maintain a balanced diet. To estimate food nutritional values, this proposed model estimates the food portion by weight predictions, hence an image dataset that consists of 32 food categories and 7754 vegetable and non-vegetable images, with known weights, was prepared for model training and testing.

It should be noted that the various cooking techniques are not present for all vegetables in equal proportions, and some combinations are not existing at all. Furthermore, non-vegetables are not differentiated at all. This is in line with practical situations, where not all food and cooking types will be represented equally. Since there is still a wide variety of foods and we have an independent validation and test set, the results should still be reliable.

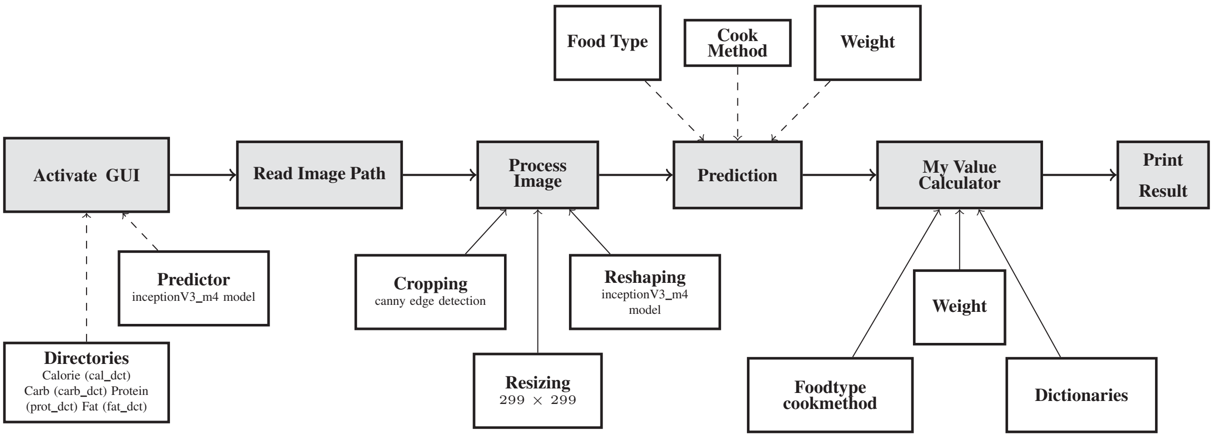


Fig. 3. Process flow of developed GUI

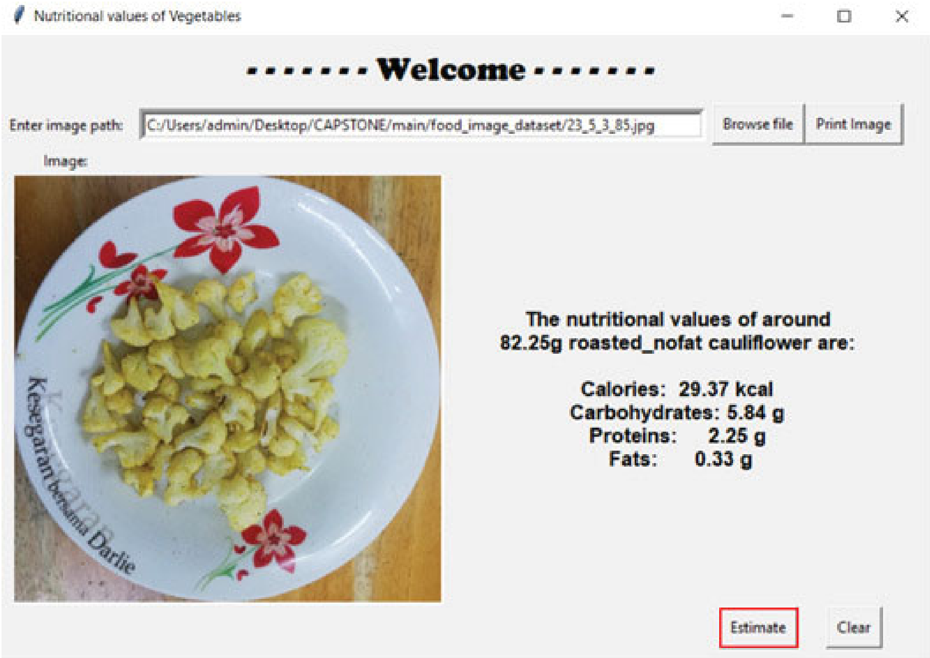


Fig. 4. GUI demonstration. The estimated nutritional values of the identified raw/cooked vegetable from the input image (displayed by clicking on the 'Print Image' button) are calculated and displayed by clicking on 'Estimate'

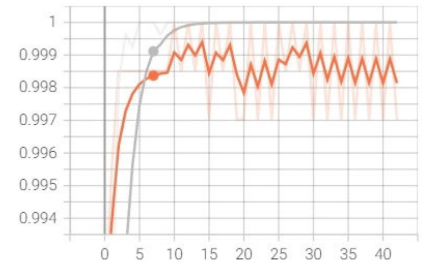
When comparing our results with related works, Table V shows that VegeNet outperforms them in both classification and nutritional value estimation tasks. VegeNet's classification performance is the highest at 97% and its nutritional value estimation is the second most accurate at only 15.30% mean relative error. Nevertheless, all models are trained and tested with different datasets: these values are only the general guidelines to assess the model performance in this domain.

Looking at our initial question if the cooking state can be used for nutritional value estimation, we can clearly give a positive answer. Our results show the model trained with the cooking state outperforms conventional models. Furthermore, the architecture of VegeNet is suitable for this task. Specifically, the results show that the network is able to classify the cooking state very accurately, showing that this task alone can realistically be left to AI.

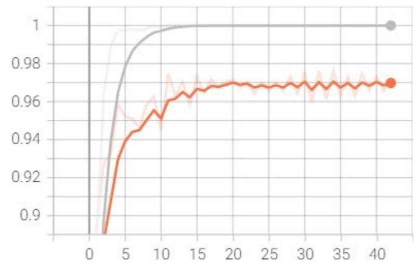
B. Limitation and future work suggestion

Despite the high model accuracy obtained in this project, several limitations are observed:

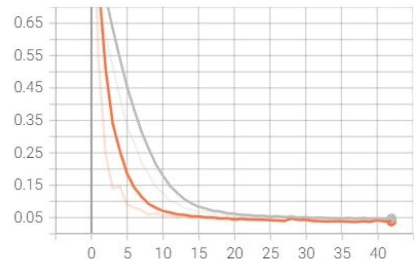
1) The limitation of datasets. To estimate the nutritional value through food weight prediction, an image dataset with known food weights is required; yet, this type of dataset is not easily obtained, especially the vegetable images. Thus, the images collected are limited by vegetable type and cooking method, as well as container, location, background, capturing device, and ambient



(a) Accuracy of food type classification.



(b) Cooking method accuracy.



(c) Weight (MAE).

Fig. 5. Model training output over the number of epochs. The x-axis represents epochs and the y-axis accuracy respectively MAE. The red curves represent validation and the gray curves represent training. The faint lines are original data, the solid lines are smoothened.

TABLE V. Model performances of the proposed model and related work. [1]Mean relative error

Model	Image dataset	Classification Accuracy	Nutritional value estimation[1]
VegeNet (this project)	Self-captured vegetable images + Hospital food images (32 categories, 7,754 images)	97.00 %	15.30 %
Faster R-CNN + ResNet50 [41]	Hospital food images (40 categories, 20,084 images)	73.35 %	16.97 %
Multi-task CNN [35]	Recipe dataset (36 categories, 7361 images)	81,20 %	27.40 %
Faster R-CNN [34]	School lunch dataset (21 categories, 4,877 images)	90.7 %	21.40 %
SVM + Gabor filter [28]	Self-captured images (15 categories, 3,000 images)	90.41 %	14.00 %

lighting, which make the prediction of the model less general. In addition, the model is subject to overfitting due to the small image dataset utilised. Furthermore, the nutrition information extracted from the USDA dataset is not always exact, for example "cooked with added oil" does not clearly specify the amount of oil added, also potentially leading to a deviation between the predicted value and the actual value. To overcome this limitation, a collection of large image datasets with known food weights and the correct food nutritional values shall be created for future work, so that the models developed are more accurate.

2) It is known that in real life, the images inputted for predictions are captured under various environmental conditions, like low or high lighting, for example, and would affect the image quality; hence influencing the model prediction accuracy. Other than low-res mask filtering, there are many conventional image pre-processing techniques like variational denoising methods and transform techniques for image denoising. Moreover, deep CNN-based image denoising models are found to advance the image preprocessing performance. The addition of deep learning models in the image pre-processing step shall be done in future work.
3) This model only considers the energy and macro-nutrient content of vegetables. A more accurate visual-based diet assessment application should involve more types of food and micronutrient information and should be built in collaboration with professional nutritionists.

VI. Conclusion

In the last decade, conventional machine learning and deep learning models, in particular CNNs, have been developed to recognise food through image classification and estimate its calories; yet these models mainly focus on prediction for cooked foods and did not consider the nutritional value difference between cooked and uncooked foods. This paper proposes an InceptionV3-based multi-output CNN model, which estimates the nutritional values of cooked and uncooked vegetables, to address this issue. The explicit use of the cooking state is a novel contribution of this paper. This model can predict the type of food and the cooking method with near 100% accuracy and, combined with its weight estimate,

estimate the nutritional value at around 16% mean relative error, clearly outperforming existing models. In particular, the ability to predict the cooking state accurately is a novel achievement. The separate output of cooking state, weight, and food type also allows further processing of the results and is a step towards explainable AI.

REFERENCES

[1] C. A. Monteiro, R. B. Levy, R. M. Claro, I. R. R. de Castro, and G. Cannon, "Increasing consumption of ultra-processed foods and likely impact on human health: evidence from brazil," *Public Health Nutrition*, vol. 14, no. 1, p. 5–13, 2010.

[2] *Diet, nutrition, and the prevention of chronic diseases: report of a joint WHO/FAO expert consultation*, ser. WHO technical report series. World Health Organization, 2003, vol. 916.

[3] WHO, "Obesity and overweight," https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight, accessed: 2022-06-24.

[4] WHO, "#healthyathome: Healthy diet," https://www.who.int/campaigns/connecting-the-world-to-combat-coronavirus/healthyathome/healthyathome---healthy-diet, accessed: 2022-06-24.

[5] O. Shim Jee-Seon, K. Kyungwon, and H. Chang, "Dietary assessment methods in epidemiologic studies," *Epidemiol Health*, vol. 36, p. e2014009, 2014.

[6] A. H. Sarah Davis, "Best calorie counting apps of 2023," https://www.forbes.com/health/body/best-calorie-counting-apps/, May 2023.

[7] V. Kipnis, D. Midthune, L. S. Freedman, S. Bingham, A. Schatzkin, A. Subar, and R. J. Carroll, "Empirical evidence of correlated biases in dietary assessment instruments and its implications," *American Journal of Epidemiology*, vol. 153, no. 4, pp. 394–403, 2001. [Online]. Available: https://doi.org/10.1093/aje/153.4.394

[8] C. Pettitt, J. Liu, R. M. Kwasnicki, G.-Z. Yang, T. Preston, and G. Frost, "A pilot study to determine whether using a lightweight, wearable micro-camera improves dietary assessment accuracy and offers information on macronutrients and eating rate," *British Journal of Nutrition*, vol. 115, no. 1, p. 160–167, 2016.

[9] M. Filipic, "Chow line: Steam, roast vegetables to retain nutrients," https://cfaes.osu.edu/news/articles/chow-line-steam-roast-vegetables-retain-nutrients, accessed: 2023-05-26.

[10] G. Villalobos, R. Almaghrabi, P. Pouladzadeh, and S. Shirmohammadi, "An image processing approach for calorie intake measurement," in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings*. Budapest, Hungary: IEEE, 6 2012, pp. 1–5.

[11] P. Pouladzadeh, S. Shirmohammadi, and R. Al-Maghrabi, "Measuring calorie and nutrition from food image," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 8, pp. 1947–1956, 2014.

[12] D. Bhargava, S. Vyas, and A. Bansal, "7 - comparative analysis of classification techniques for brain magnetic resonance imaging images," in *Advances in Computational Techniques for Biomedical Image Analysis*, D. Koundal and S. Gupta, Eds. Dehradun, India: Academic Press, 2020, pp. 133–144. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128200247000074

[13] H. Hasan, H. Z. Shafri, and M. Habshi, "A comparison between support vector machine (SVM) and convolutional neural network (CNN) models for hyperspectral image classification," *IOP Conference Series: Earth and Environmental Science*, vol. 357, no. 1, p. 012035, nov 2019. [Online]. Available: https://doi.org/10.1088/1755-1315/357/1/012035

[14] H. Kagaya, K. Aizawa, and M. Ogawa, "Food detection and recognition using convolutional neural network," in *Proceedings of the 22nd ACM International Conference on Multimedia*, ser. MM '14. New York, NY, USA: Association for Computing Machinery, 11 2014, p. 1085–1088. [Online]. Available: https://doi.org/10.1145/2647868.2654970

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

[16] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," 2013. [Online]. Available: https://arxiv.org/abs/1311.2901

[17] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1409.1556

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," 2014. [Online]. Available: https://arxiv.org/abs/1409.4842

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385

[20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 6 2016, pp. 2818–2826.

[21] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017. [Online]. Available: https://arxiv.org/abs/1704.04861

[22] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *MADiMa '16: Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '16. New York, NY, USA: Association for Computing Machinery, 10 2016, p. 41–49. [Online]. Available: https://doi.org/10.1145/2986035.2986042

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA: IEEE, 6 2015, pp. 1–9.

[24] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, Y. Yao, A. Zhang, L. Zhang, W. Han, M. Huang, Q. Jin, Y. Lan, Y. Liu, Z. Liu, Z. Lu, X. Qiu, R. Song, J. Tang, J.-R. Wen, J. Yuan, W. X. Zhao, and J. Zhu, "Pre-trained models: Past, present and future," *AI Open*, vol. 2, pp. 225–250, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666651021000231

[25] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. Turin, Italy: IEEE, 7 2015, pp. 1–6.

[26] G. Ciocca, P. Napoletano, and R. Schettini, "Cnn-based features for retrieval and classification of food images," *Computer Vision and Image Understanding*, vol. 176-177, pp. 70–77, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1077314218302467

[27] S. Phiphiphatphaisit and O. Surinta, "Food image classification with improved mobilenet architecture and data augmentation," in *Proceedings of the 2020 The 3rd International Conference on Information Science and System*, ser. ICISS 2020. New York, NY, USA: Association for Computing Machinery, 3 2020, p. 51–56. [Online]. Available: https://doi.org/10.1145/3388176.3388179

[28] P. Pouladzadeh and S. Shirmohammadi, "Mobile multi-food recognition using deep learning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 13, no. 3s, aug 2017. [Online]. Available: https://doi.org/10.1145/3063592

[29] T. Ege and K. Yanai, "Multi-task learning of dish detection and calorie estimation," in *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*, ser. CEA/MADiMa '18. New York, NY, USA: Association for Computing Machinery, 7 2018, p. 53–58. [Online]. Available: https://doi.org/10.1145/3230519.3230594

[30] S. Mezgec, T. Eftimov, T. Bucher, and B. Koroušić Seljak, "Mixed deep learning and natural language processing method for fake-food image recognition and standardization to help automated dietary assessment," *Public Health Nutrition*, vol. 22, no. 7, p. 1193–1202, 2019.

[31] X.-J. Zhang, Y.-F. Lu, and S.-H. Zhang, "Multi-task learning for food identification and analysis with deep convolutional neural networks," *Journal of Computer Science and Technology*, vol. 31, no. 3, pp. 489–500, 2016.

[32] N. Martinel, G. L. Foresti, and C. Micheloni, "Wide-slice residual networks for food recognition," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Lake Tahoe, NV, USA: IEEE, 3 2018, pp. 567–576.

[33] J. Zheng, L. Zou, and Z. J. Wang, "Mid-level deep food part mining for food image recognition," *IET Computer Vision*, vol. 12, no. 3, pp. 298–304, 2018. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-cvi.2016.0335

[34] T. Ege and K. Yanai, "Estimating food calories for multiple-dish food photos," in *2017 4th IAPR Asian Conference on Pattern Recognition (ACPR)*. Nanjing, China: IEEE, 12 2017, pp. 646–651.

[35] ——, "Image-based food calorie estimation using recipe information," *IEICE Transactions on Information and Systems*, vol. E101.D, no. 5, pp. 1333–1341, 2018.

[36] K. Okamoto and K. Yanai, "An automatic calorie estimation system of food images on a smartphone," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '16. New York, NY, USA: Association for Computing Machinery, 10 2016, p. 63–70. [Online]. Available: https://doi.org/10.1145/2986035.2986040

[37] T. Ege, W. Shimoda, and K. Yanai, "A new large-scale food image segmentation dataset and its application to food calorie estimation based on grains of rice," in *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '19. New York, NY, USA: Association for Computing Machinery, 10 2019, p. 82–87. [Online]. Available: https://doi.org/10.1145/3347448.3357162

[38] H. Hassannejad, G. Matrella, P. Ciampolini, I. D. Munari, M. Mordonini, and S. Cagnoni, "A new approach to image-based estimation of food volume," *Algorithms*, vol. 10, no. 2, 2017. [Online]. Available: https://www.mdpi.com/1999-4893/10/2/66

[39] J. Dehais, M. Anthimopoulos, and S. Mougiakakou, "Food image segmentation for dietary assessment," in *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '16. New York, NY, USA: Association for Computing Machinery, 10 2016, p. 23–28. [Online]. Available: https://doi.org/10.1145/2986035.2986047

[40] Y. A. Sari, V. W. Saputra, A. Agustina, Y. A. Wani, and Y. G. Bihanda, "Comparison of image thresholding and clustering segmentation methods for understanding nutritional content of food images," in *Proceedings of the 5th International Conference on Sustainable Information Engineering and Technology*, ser. SIET '20. New York, NY, USA: Association for Computing Machinery, 11 2021, p. 124–129. [Online]. Available: https://doi.org/10.1145/3427423.3427441

[41] P. Ruenin, J. Bootkrajang, and J. Chawachat, "A system to estimate the amount and calories of food that elderly people in the hospital consume," in *Proceedings of the 11th International Conference on Advances in Information Technology*, ser. IAIT2020. New York, NY, USA: Association for Computing Machinery, 7 2020, pp. 1–7. [Online]. Available: https://doi.org/10.1145/3406601.3406613

[42] P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, and G. Prasad, "Tangent space features-based transfer learning classification model for two-class motor imagery brain–computer interface," *International journal of neural systems*, vol. 29, no. 10, p. 1950025, 2019.

[43] H. S. Nogay and H. Adeli, "Detection of epileptic seizure using pretrained deep convolutional neural network and transfer learning," *European neurology*, vol. 83, no. 6, pp. 602–614, 2020.

[44] H. Liu, F. Gu, and Z. Lin, "Auto-sharing parameters for transfer learning based on multi-objective optimization," *Integrated Computer-Aided Engineering*, vol. 28, no. 3, pp. 295–307, 2021.

[45] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[46] J. Y. Tan, "Vegenet - image datasets and codes," https://doi.org/10.5281/zenodo.7254508, Oct. 2022.

[47] "Fooddata central," https://fdc.nal.usda.gov/, accessed: 2022-06-24.