Traditional food recognition using convolutional neural networks

CSA301 Deep Learning
Bachelor of Science in Computer Science
(Year IV, Semester I)

RESEARCHER(S)

12190002 Choki Dorji
12190019 Samten Gyaltshen
12190025 Sonam Chophel
12190039 Yeshi Choden

GUIDED BY

Yonten Jamtsho

Gyalpozhing College of Information Technology
Gyalpozhing : Mongar

1 Abstract

With the development of machine learning, as a branch of machine learning, deep learning has been applied in many fields such as image recognition, image segmentation, video segmentation, and so on through Convolutional Neural Network (CNN). The project on Traditional Food Recognition using Convolutional Neural Network mainly focuses on the food item in Bhutan that has been consumed from generation to generation. This project educates the younger generation about various food and supports culture and tradition by showing them food varieties that has existed since ancient times and it will also help the tourist to know the name of the food with the image of particular food. The purpose of the project is to develop a system to recognize traditional foods that are available in Bhutan. The scope of the project is on Bhutanese traditional foods, which includes unique food of Bhutan, Ema Datshi, Khurlay, Puta, and Hoentay. The project will involve using supervised image classification techniques to train a model to identify the traditional foods using computer vision. This model will then be deployed to a live website where tourists can upload the image to detect the traditional foods. Convolutional neural networks (CNN), a class of deep neural network, are majorly used for the process of image recognition. CNN consists of some basic layers like hidden layers and fully connected layers where hidden layers are used to extract and learn the features of training images and fully connected layers are used for classification of the image. The study concludes that the model perform well given the number of dataset and given algorithm. In the result analysis, the accuracy of the training dataset of images obtained is about 99%. The task of recognition can be made better by taking out noise from the dataset and the project can also be run on a larger dataset with more classes and images. Thus that's the future improvement.

Keyword: Deep Learning; Convolution Neural Network; Traditional Food Recognition; Image Augmentation.

2 Introduction

Bhutan being considered as the last Shangri La and is rich in culture and heritage, many people wish to visit Bhutan. Tourism has a substantial impact on Bhutan's GDP (Gross Domestic Product), which the nation depends on to maintain its unique way of life and unaltered customs and culture. Since our great lord, Zhabdrung Ngawang Namgyal, devoured traditional food, it is regarded as one of the most alluring lifestyles that Bhutanese people have and this makes traditional food in Bhutan blessed to be consumed. Tourists, on the other hand, find our traditional foods unappealing to them because they have no clue about what the names of the foods are and have trouble deciding what kind of food they want to eat. Consequently, the risk that visitors will lose interest in visiting our nation is therefore very high. Furthermore, future generations in our nation might not be aware of the background of traditional foods. Our cultural history will certainly disappear over time, and in the near future, younger generations will likely consume more meals associated with western lifestyles. In order to reduce the likelihood of such hazards, this project educates the younger generation about various foods and promotes culture and tradition by teaching them about the many food varieties that have existed since ancient times through the identification of foods. Thus, to make Bhutanese traditional food popular all over the world, this project will assist everyone to recognize the traditional foods in Bhutan with the help of deep learning techniques.

3 Related Work

A modern computer-based food recognition system for reliable food was developed and is now available in mobile devices and rich Cloud services, according to the paper "Study for Food Recognition System Using Deep Learning" [1]. It investigated numerous methods for identifying food and assessed their efficiency depending on a number of different characteristics. The system classifies each group of foods using a variety of food types, including Indian food, Food 101, and Food 11, and a number of data analysis techniques, including partial least squares (PLS), artificial neural networks (ANN), support vector machines (SVM), random forests, and k-nearest neighbor (KNN). The accuracy obtained by the researcher utilizing these methodology and techniques ranges from 70 percent to 100 percent .

Food Image Recognition and Food Safety Detection Method Based on Deep Learning [2], In this study, the detection task is divided into location and classification tasks for the practical application of restaurants, and the convolutional neural network is used to solve each task. This demonstrates how the experimental findings from the CNFood-252 dataset contribute to increasing the recognition accuracy. The picture matching approach is used to identify, and the dataset with less samples is generated, since the measuring method requires gathering a large number of training samples for presentation. Although the FewFood-50 dataset's testing findings indicate that this method's best accuracy is only 45.75 percent, the absence of the necessity for manual sample labeling demonstrates that it has strong development potential for practical popularization or application.

In the article Traditional food knowledge of Indonesia: a new high-quality food dataset and automatic recognition system [3], data collection and automated food recognition for Indonesian traditional meals were carried out in this study. Images of the dish were taken in a professional little studio. The food image data were taken with identical lighting conditions, camera settings, and shooting distances. With the use of a light intensity metre, movable illumination, and a laser distance measuring tool, the parameters were carefully measured and set up. The data gathering technique was effective in obtaining data for 1644 traditional food photos. 34 different traditional meal varieties were represented by these photographs, and 30–50 images were collected for each one. The raw food picture data was 53 GB in size. To facilitate training, testing, and validation, the data were separated into sets. To categorise Indonesian traditional cuisine, an automated identification system was created. Several different convolutional neural network (CNN) models, including Densenet121, Resnet50, InceptionV3, and Nasnetmobile, were used for training. The evaluation's findings showed that the automated recognition system could achieve a reasonable area under the receiver operating characteristics (AUROC) and high accuracy, precision, and recall values of more than 0.95 when working with high quality datasets.

In the article Mobile Multi-Food Recognition Using Deep Learning [4] they proposed a mobile food recognition system that uses the picture of the food, taken by the user's mobile device, to recognize multiple food items in the same meal, such as steak and potatoes on the same plate, to estimate the calorie and nutrition of the meal. The advantage 1 of recognizing items, instead of the whole meal, is that the system can be trained with only single item food images. Their experiments, conducted with the Food dataset, show an average recall rate of 90.98 %, precision rate of 93.05 % , and accuracy of 94.11 %.

By creating nutrition tracking applications, the Convolutional Neural Networks (CNN) for Food Image Recognition [5] study promotes health awareness. Preparing data sets, selecting CNN architecture, and CNN optimization methodologies in the context of food classification were the most difficult aspects of designing such applications. After conducting the survey, they came to the following conclusions: larger networks like Xception should be used, a minimum of 300 images per category is sufficient, image augmentation techniques that do not alter shapes are more beneficial, a dataset balancing strategy might not be necessary to address class imbalance in food datasets below an imbalance ratio of 7, and higher native image resolution during training is beneficial to classifier performance, especially for networks requiring larger input size.

4 Methodology

4.1 System Overview

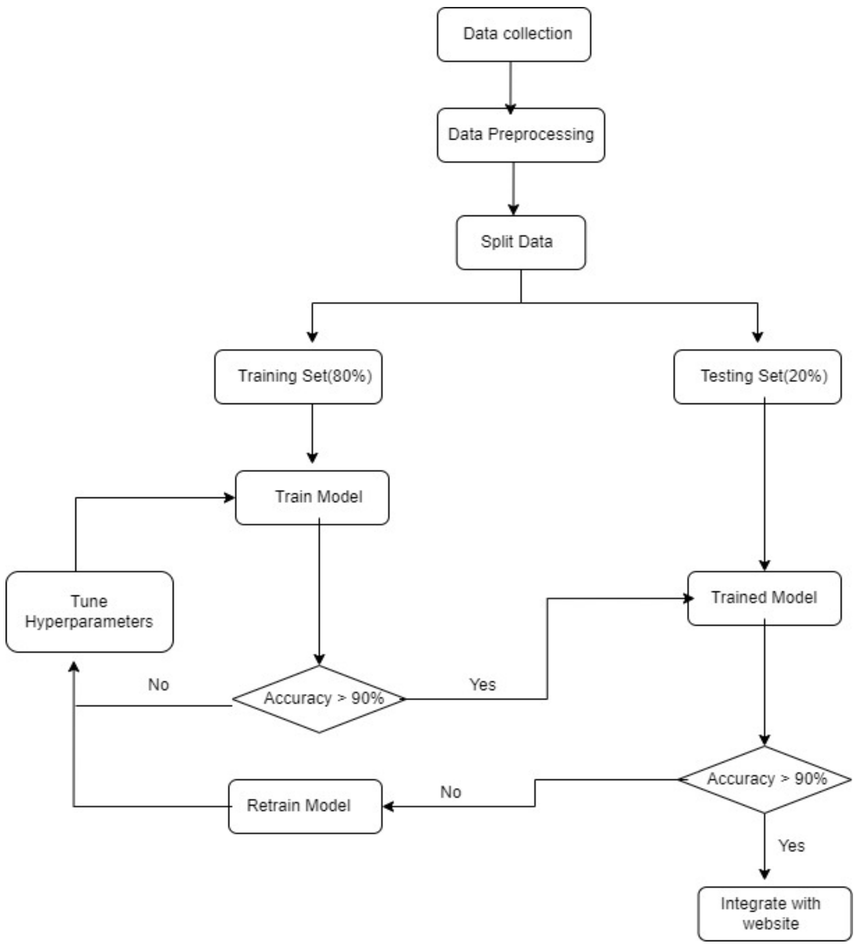

Figure 1: Flow chart for the system

Firstly, we will be collecting the data of unique food of Bhutan. After we are done with the collection of the data, we are going to do data preprocessing. Then the data will be splitted into training and testing set.

After that, the model will be trained using the training set, and its accuracy will be evaluated. Using the testing set, we will test the model if we achieve high accuracy. If the model's accuracy is poor, it will be retrained after the hyper-parameters are adjusted until the desired accuracy is attained.

Lastly, we will come to final decision seeing how well our model performs and it will be integrated with the website.

4.2 Algorithm

Convolutional Neural Network

A CNN architecture-based approach is recommended for recognizing traditional Bhutanese foods. In order to distinguish traditional Bhutanese foods, a CNN with three convolution layers is developed, as shown in Figure: 2. Image classification, image recognition, and computer vision (CV) were used in CNN applications; they are also highly helpful since they produce incredibly precise results, especially when a lot of data is involved. CNN keeps concentrating on the object's characteristics as it moves through the object's many levels. This rapid learning reduces the need for manual feature extraction. The input layer, hidden layers, and output layer are the stages in the classification process for images.

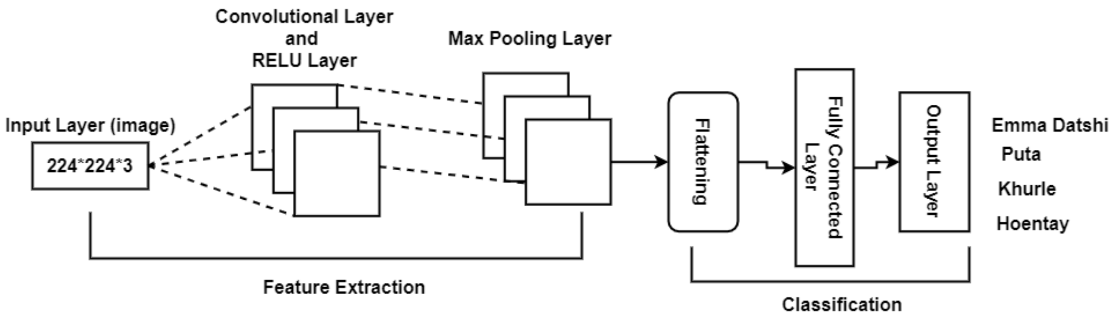

Figure 2: CNN architecture

Table 1: Summary of the model

Layer(type)	Output Shape	Param
Conv2d(Conv2D)	(None,222,222,32)	896
maxpooling2d (MaxPooling2D)	(None,111,111,32)	0
conv2d1(Conv2D)	(None,109,109,64)	18496
maxpooling2d1 (MaxPooling2D)	(None,54,54,64)	0
Conv2d2(Conv2D)	(None,52,52,128)	73856
maxpooling2d2 (MaxPooling2D)	(None,26,26,128)	0
Conv2d3(Conv2D)	(None,24,24,256)	295168
maxpooling2d3 (MaxPooling2D)	(None,12,12,256)	0
flatten (Flatten)	(None, 36864)	0
dense (Dense)	(None,64)	2359360
dense1 (Dense)	(None, 5)	325

This project will employ three convolutional layers with a relu activation function and three filter sizes on each. Between each convolutional layer, there will be a maxpooling layer to stop the model from overfitting. The flattening layer that follows has its output serve as the fully connected layer's input. A softmax function will be the output layer's last layer because this is a multi-class classification problem.

4.3 Dataset

For this project, our group will be collecting data(images) from websites and people through social media platforms. There are a total of 5 classes for our project:

1. EmaDatshi
2. Hoentey
3. Khurle
4. Puta
5. unknown

We will be collecting a total of 1000 images from each class. The train size, validation size and test size will be 70 percent, 20 percent and 10 percent respectively.

4.4 Evaluation Metrics (confusion matrix)

A confusion matrix is a matrix representation that demonstrates how accurately the trained model predicts each target class in relation to counts. Confusion Matrix is mostly utilized for supervised learning's classification algorithms. It can be used to calculate performance metrics like accuracy, precision, recall, and F1-score in order to assess the effectiveness of a classification model. It is helpful in calculating recall, precision, accuracy, and the f1-score

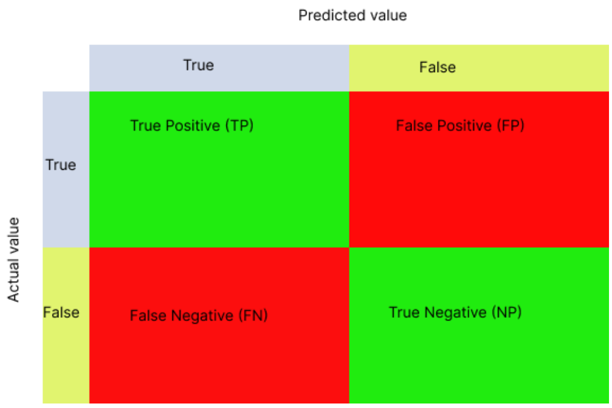

Figure 3: confusion matrix

- **True Positive (TP):** when the trained model correctly predicts the positive class.
- **True Negative (TN):** when the model correctly predicts the negative class.
- **False Positive (FP):** when the model incorrectly predicts the positive class.
- **False Negative (FN):** when the model incorrectly predicts the negative class.

Precision

Precision is the ratio between the correctly predicted observations to the total predicted positive observations. It is calculated as;

$$Precision = \frac{TP}{TP + FP}$$

Recall

Recall is a ratio between the correctly predicted positive observations to the observations in the actual class. It is calculated as;

$$Recall = \frac{TP}{TP + FN}$$

Accuracy

It is the ratio of correctly predicted class to the total instances.The datasets with an equal number of examples in each class are the ones where this method of determining accuracy is most well-known. The following formula is used to determine accuracy:;

$$Accuracy = \frac{TP}{TP + FP + FN + TN}$$

F1-Score
When output estimate with recall and precision can no longer be possible, the F1-Score is a feasible solution. The formula for averaging precision and recall is as follows;

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall}$$

5 Results and Discussions

This section of the paper presents the results of the model. Paper will be discussing the evaluation in terms of accuracy, precision, recall, f1-score, and confusion matrix. Accuracy is monitored on the training, validation and test datasets; this metric represents the mean percentage of correctly classified classes on a dataset.

After selecting the best loss and optimizer for the model, the model is then trained using that loss function and optimizer. The model is then evaluated to measure the performance or the accuracy on the training and testing dataset using the score and losses. Following plots of loss and accuracy on the training and validation sets were being plotted. The plot helped to check the existence of 'overfitting' and 'underfitting' by plotting the differ- ence between training and validation accuracy. Besides, it helped monitor if there is an improvement in the training accuracy.

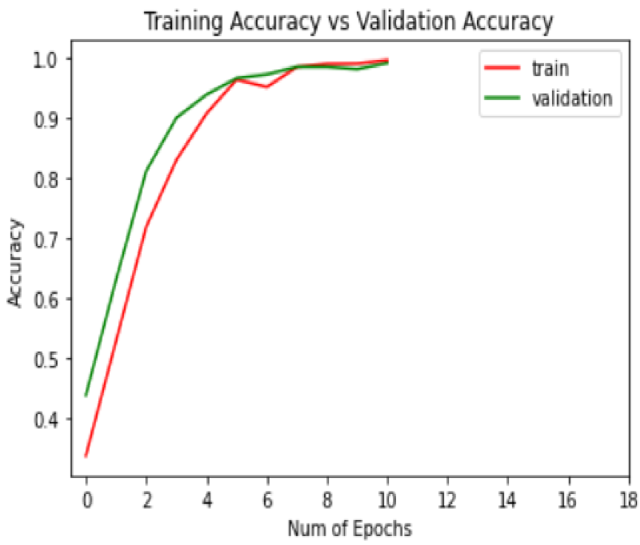

Figure 4:Training and Validation Accuracy Graph

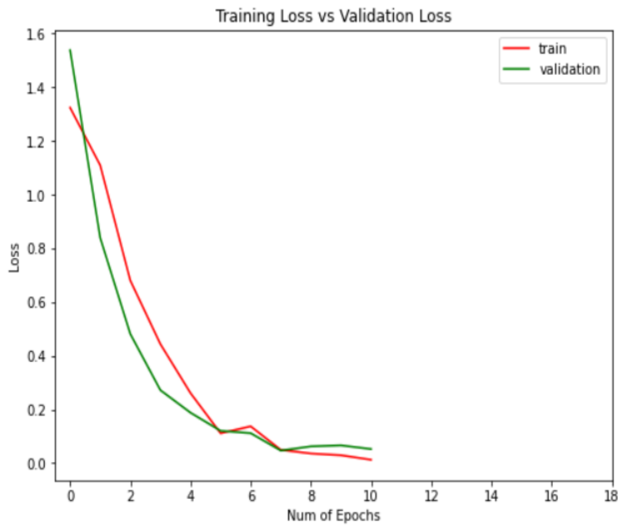

Figure 5: Training and Validation Loss Graph

The current food recognition model achieved 99.89% for training and 31% accuracy on the test dataset. The following figure is the report generated from predicting the test dataset which shows the precision, recall and f1-score, macro avg, weighted avg and the overall accuracy.

Table 1: Classification-report

	precision	**recall**	**f1-score**	**Support**
EmaDatshi	1.00	0.95	0.97	100
Hoentey	0.97	1.00	0.99	100
Khurle	1.00	0.99	0.99	100
Puta	1.00	0.94	0.97	100
Unknown	0.89	0.97	0.93	100
Accuracy			0.97	500
Macro avg	0.97	0.97	0.97	500
Weighted avg	0.97	0.97	0.97	500

Evaluating the model using confusion matrix: Evaluated the model by plotting a confusion matrix for predicted labels against the true label of test images. To visualize the number of images that got predicted correctly and the number which were mispredicted, the confusion matrix from matplotlib is being used. The matrix helps to see the number of test samples which were misclassified. From the matrix below, the images got predicted correctly although the model cannot be considered good.

Table 2: Confusion Matrix

Actual Classes \ Predicted Classes	EmaDatshi	Hoentey	Khurle	Puta	unknown
EmaDatshi	95	0	0	0	5
Hoentey	0	100	0	0	0
Khurle	0	0	99	0	1
Puta	0	0	0	94	6
Unknown	0	3	0	0	97

The figure given below shows the model after being integrated with the Traditional food recognition website.

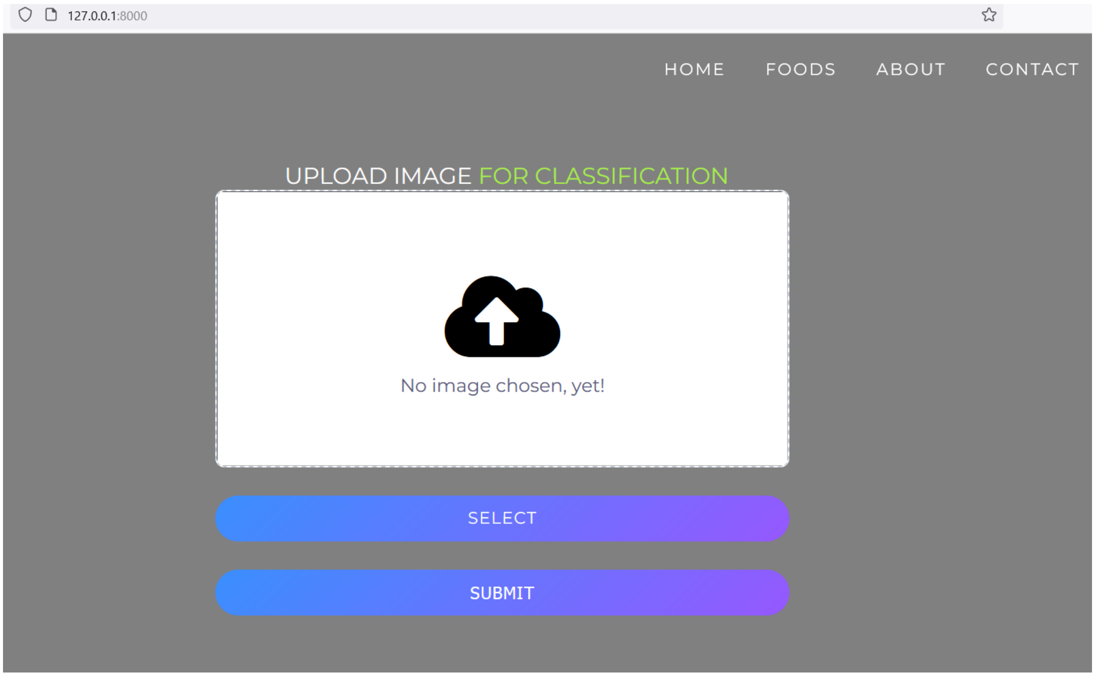

Figure 6: Home Page

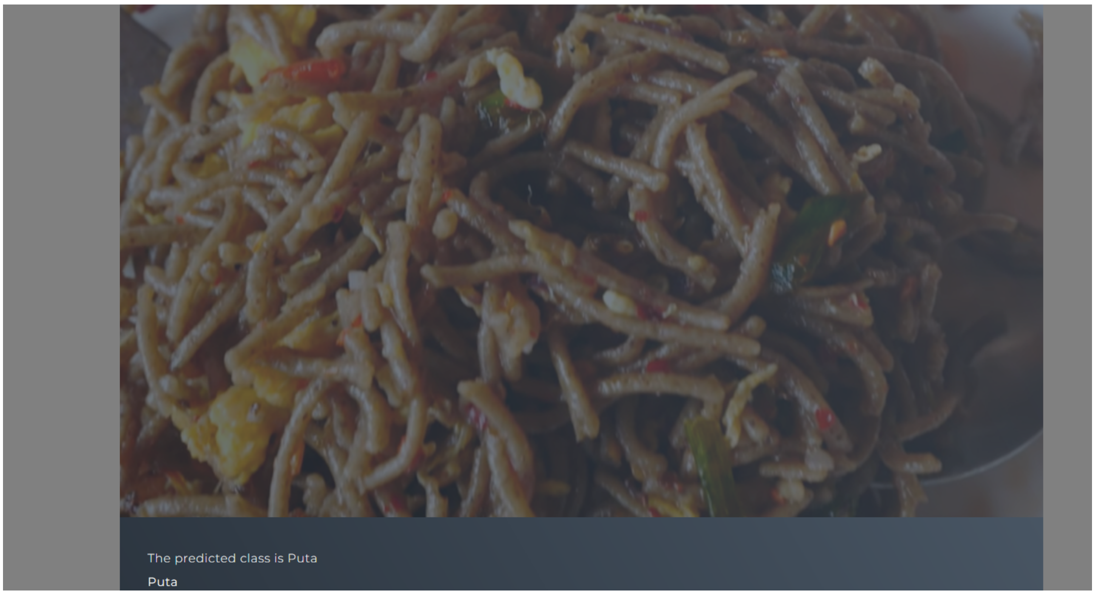

Figure 7: Predicted result

6 Conclusion

In this project, food images are recognized to appropriate classes using a deep learning approach called a convolution neural network. The task of recognition can be made better by taking out noise from the dataset, thus that's the future improvement. The project can also be run on a larger dataset with more classes and images in each class because doing so increases accuracy by learning more features and reduces the loss rate.

In the result analysis, the accuracy of the training dataset of images obtained is about 99%. As for the future, a sizable amount of data must be used to train the CNN, and even a well-trained network cannot achieve 100% segmentation accuracy, which will increase the error rate of recognition. In order to achieve better results, we can therefore create a larger dataset that contains various food images.

References

[1] N. O. Salim, S. R. Zeebaree, M. A. Sadeeq, A. Radie, H. M. Shukur, and Z. N. Rashid, “Study for food recognition system using deep learning,” in *Journal of Physics: Conference Series*, vol. 1963, no. 1. IOP Publishing, 2021, p. 012014.

[2] Y. Wang, J. Wu, H. Deng, and X. Zeng, “Food image recognition and food safety detection method based on deep learning,” *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[3] A. Wibisono, H. A. Wisesa, Z. P. Rahmadhani, P. K. Fahira, P. Mursanto, and W. Jatmiko, “Traditional food knowledge of indonesia: a new high-quality food dataset and automatic recognition system,” *Journal of Big Data*, vol. 7, no. 1, pp. 1–19, 2020.

[4] 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