Fruits and Vegetables Recognition System in Dzongkha Using Visual Geometry Group Network

Karma Wangchuk[1] , Tsheten Dorji[2] , Parshu Ram Dhungyel[3], Pema Galey[4]
Information and Technology Department, College of Science and Technology, Royal University of Bhutan,

E-mail: karma.cst@rub.edu.bt, tshetendorji.cst@rub.edu.bt, parshuram.cst@rub.edu.bt, pemagaley.cst@rub.edu.bt

Abstract

English speaking among the youth is gaining popularity in Bhutan. The Ministry of Education and Dzongkha Development Commission have been trying to promote the national language through different strategies. Furthermore, developed interesting learning materials in Dzongkha. However, youth find it difficult even to name common fruits and vegetables in Dzongkha, and this is an increasing concern. The purpose of this study was to develop an automatic fruit recognition system in Dzongkha using machine learning techniques. Ten classes of fruits and 17 classes of vegetables that are found in Bhutan are considered for the study. The fruits and vegetable datasets were downloaded from Kaggle and Websites. Furthermore, images were augmented and rotated every 15 degrees. The dataset comprised 405000 images. The model was trained and deployed using customized VGGNet and Gradio. The training and validation accuracy of the proposed model was 97.76% and 97.80% respectively.

Key Words: *Fruits recognition, VGGNet, CNN, deep learning, and vegetable classification*

1. INTRODUCTION

Language defines the essential trait of nationality (Pott, 1856, as cited in Van Driem, 2014). Furthermore, Emil Cioran 1987 as cited in (Van Driem, 2014) stated that the people do not reside in the country but rather in the language. The government of Bhutan has been trying to promote the national language of Bhutan in both written and spoken expression. In addition, there are vivid executive orders and even the Royal decree stating that Dzongkha should be used in all official correspondences. However, the English language has been the popular language used amongst the youth in Bhutan and official correspondences. Dzongkha Development Commission's survey conducted in 2017 revealed that of the 43 government offices surveyed, only 10% used Dzongkha, while the rest used English for official correspondence.

The Ministry of Education and Dzongkha Development Commission have been developing exciting courses and reading materials in Dzongkha. However, the recommendation that was given by the National Council to teach history and other subjects in Dzongkha had to be dropped due to a lack of interest from students and insufficient Dzongkha teachers (Kuensel, 2017). Despite having ample reading materials in Dzongkha, students are interested in English. As a result, youth are not able to name common fruits and vegetables found in Bhutan in Dzongkha which is an increasing concern for both parents and the government of Bhutan.

The main objective of this study was to train and deploy a Machine Learning model for the detection and recognition of common fruits and vegetables found in Bhutan in the Dzongkha. This paper is organized as follows. Section 2 discusses related works followed by methodology in section 3. The results and discussion are explained in section 4 followed by a conclusion in section 5.

2. RELATED WORK

Computer Vision and Deep Learning have been used in various fields for visual inspection (Gomes & Leta, 2012). In agriculture, these technologies are used for disease and weeds detection, fruits recognition (Muresan & Oltean, 2018) and vegetable classification (Zeng, 2017). Furthermore, researchers and scholars are using these algorithms for bruise (Du et al., 2020), and freshness detection (Valentino et al., 2021), classification of fruits based on sizes, pest control, analysis of shapes and colour of cereals. There are numerous robotics applications implemented in the agriculture domain from these state-of-the-art algorithms.

Robotic platforms and technologies have been facilitating the agricultural sector in automation such as automating plantation, harvesting, weeds detection (Jin et al., 2022), estimation of yield, and automatic counting (Song et al., 2014). Using deep neural networks Sa et al. (2016) built an accurate, fast, and reliable fruit detection system. Similarly, a robotic vision system was used with Faster R-CNN for multi-class fruit detection (Wan & Goudos, 2020). Koirela et. al. (2019) implemented and compared different deep-learning algorithms to detect mango fruits. They have compared Faster R-CNN based on VGG and ZF architectures. Furthermore, YOLOv3 and YOLOv2 were trained using an image resolution of 512 x 512 pixels. The authors have proposed MangoYOLO architecture based on attributes of YOLOv2 and YOLOv3 that outperformed other deep learning algorithms.

Furthermore, machine vision was used for multiple fruits and vegetables detection and grading (Varghese et al., 2021). They have taken real-time fruit or vegetable images and extracted features for data processing. The system detected the chemical ripening of the fruits or vegetables and also predicted the shelf life of the fruits. Moreover, the identification and detection of fruits are vital prerequisite tasks for automatic yield prediction (Mai et al., 2020).

CNN-based methods were used to classify and identify vegetable leaf disease (Jaiswal et al., 2021). The new dataset was created using an open database of Plant Village. The customized dataset comprised five classes of diseased and healthy leaves. The accuracy obtained by sequential and GoogLeNet models were 98.48% and 97.47% respectively. In addition, the Visual Geometry Group (VGG) model was used for image classification. VGG model was based on the Convolutional Neural Network (CNN) proposed by Simonyan & Zisserman (2014). The VGG-16 was the runner-up in the ILSVRC challenge 2014. The model obtained 92.7% top-5 test accuracy in ImageNet. In this study, our proposed model was based on the VGG model. However, we used 8 layers (VGG-8).

3. METHODOLOGY

Figure 1 shows the overview of the study conducted. The literature studies were conducted to select state-of-the-art computer vision algorithms. Furthermore, the best practices for image acquisition and image pre-processing methods were studied during the literature review. The Dzongkha fruits and vegetable detection model was trained and deployed using VGGNet and Gradio respectively. In the following section, each phase is discussed in detail.

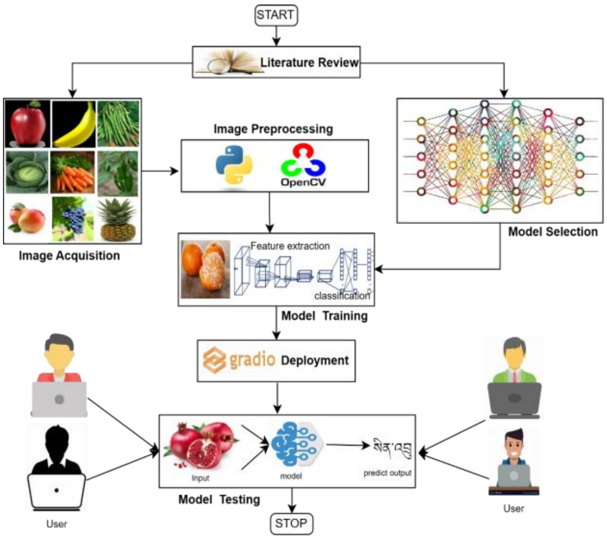


Fig. 1 Overview of the proposed study

3.1 Data Acquisition

Data collection is one of the most important phases of machine learning end-to-end projects. In object classification, images have been collected in different lighting conditions, distances, and angles to train robust models. In addition, the resolution of the images affects accuracy. However, dataset downloaded from *Kaggle* and *websites* do not have these variation.

Fig. 2 Ten classes of fruits and 17 classes of vegetables

In this study, the fruit and vegetable images were downloaded from *Kaggle* (Seth, 2020). Lettuce and spinach have similar features. As a result, spinach was removed. Similarly, the radish was excluded from training the model. Furthermore, images were downloaded from *websites* using Fatkun bulk downloading extension tools. The dataset consisted of 10 classes of fruits: apple (ཨེ་པཱལ), banana (ངང་ལག), grapes (ཆུ་རྒུན་འབྲུམ), kiwi (ཕོ་བེཀྣི), mango (ཨམ་ཙུ་ཀུ་ལི), orange (ཚལ་ལུ), pear (སླི), pineapple (ཀྲུ་ནག་གོང་ཅེ),watermelon (ཁ་རི་ཨུ་ཟ), pomegranate (སེན་འབྲུ), and 17 classes of vegetables: cabbage (འདབ་མ་ཀོ་པི), bean (སྲནམ་ཙུམ), cauliflower (མེ་ཏོག་ཀོ་པི), carrot (ལ་ཕུག་དམརཔོ), chili (ཨེ་མ), garlic (སྐྱུ་སྒོགཔ), ginger (ས་སྒ), lettuce (ཏོ་སྐྱེས་པད་ཚོད), onion (སྒོགཔ),

tomato (ལླམ་བན་ད), turnip (གཡུང་རྗོག), potato (ཀེ་བ), pea (སོད་སྲནམ), eggplant (ཇོ་ལོམ), cucumber (གོན), corn (ཨ་ཤོམ), lemon (ཧུམ་ཙུང) as shown in figure 2.

3.2 Image Preprocessing

Data cleaning is a paramount step in Machine Learning. Researchers spend quality time in data pre-processing. In this preprocessing phase, several images were created from a single image as illustrated in figure 3. Using the original image, 24 images were generated after every 15-degree angle. Furthermore, images were reduced to 64 x 64 resolutions. The reduction of image size helps to train models faster. However, the features of the images are compromised. Each class consisted of randomly selected 1500 images in the final dataset.

Machine Learning models take vectors to train the model. Consequently, images are converted into numbers using the python pickling library. As a result, training time was reduced. Feeding images to the model during the training tends to take more time. The dataset was divided into training and validation sets of 70% and 30% respectively.

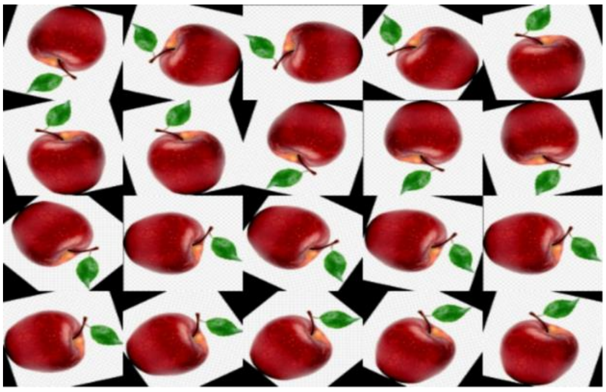


Fig. 3 Generation of images from a single image by rotation

3.3 Model Training and Deployment

The CNN-based model called VGG was proposed by Simonyan & Zisserman (2014). The authors proposed 16 layers to classify 1000 objects of image size 224 x 224. However, we used 8 layers (VGG-8) with an image size of 64 x 64. Figure 4 shows the proposed architecture adapted from VGG16 (Simonyan & Zisserman, 2014) model. There are six convolutional layers and three max-pooling layers for feature extractions followed by two dense layers for classification. In addition to these layers, batch normalization and dropout ratio were added. In each block, two convolutional layers were stacked followed by the batch normalization, max pooling, and dropout layers as shown in figure 4. The number of filters used for the first block was 64 each with a filter size of 5 x 5. However, in the second and third blocks, the filters used were 64 and 32 respectively with the same filter size of 3 x 3. The dropout ratio used was 0.25 in each block.

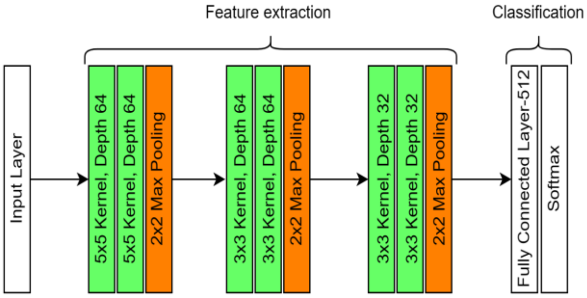


Fig. 4 The proposed architecture of the Bhutanese Fruits and Vegetables recognition Model

The filters analyze and extract features from the images. A different number of filters are used to extract various features from the images such as edges, colors, and vertical and horizontal lines that contribute to training the model for the classification task. However, batch normalization and dropout are applied in the neural network to improve the performance of the model by mitigating covariate shift and overfitting respectively. The max pooling reduces the dimension of the images by removing the pixels. The *relu* activation function was used in each of the convolutional layers. However, in the last layer of classification, the *softmax* activation function was used to distribute probability between 27 classes. The class with the highest probability would be predicted class by the model.

In the classification layer, the first dense layer consisted of 512 neurons to compute the weights of the classes. These weights are learned in forward and backward propagation. Using the last layer with the *softmax* activation function, the model predicts one of the classes. The model was trained for 50 epochs and using the *gradio* python library the model was deployed for testing.

4. RESULTS AND DISCUSSION

The proposed model was trained using Google Colab. A Tesla T4 graphic card was allocated for the training. T4 has 320 Turing Tensor cores, 2560 NVIDIA Cuda cores, and 16 GB GDDR6 memory capacity with 320+ GB/s bandwidth. However, Google Colab free allows 12 hours of training. As a result, the image size was reduced to 64 x 64, and only six convolutional layers were implemented.

The early stopping mechanism with the patience three was implemented. Training stopped at 50 epochs with training and validation accuracy of 97.76% and 97.80%% respectively. Figure 5 shows the accuracy and loss comparison of training and validation data over 50 epochs. Training accuracy rose gradually. However, validation accuracy rose at the beginning and experienced a fluctuating trend till 25 epochs. Similarly, the loss validation observed a fluctuating trend until the 37 epoch. However, training loss plummeted at the beginning epochs and remained constant after 30 epochs as shown in figure 5.

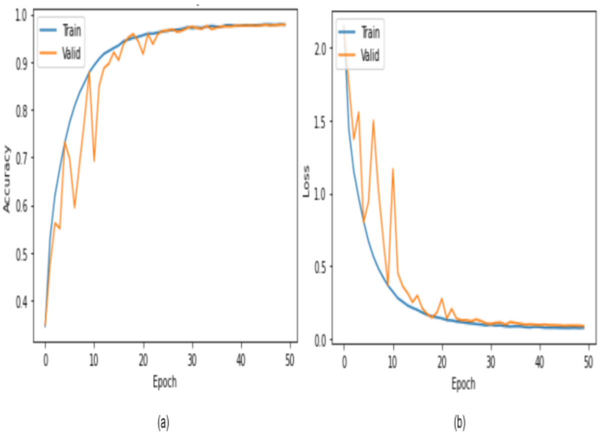
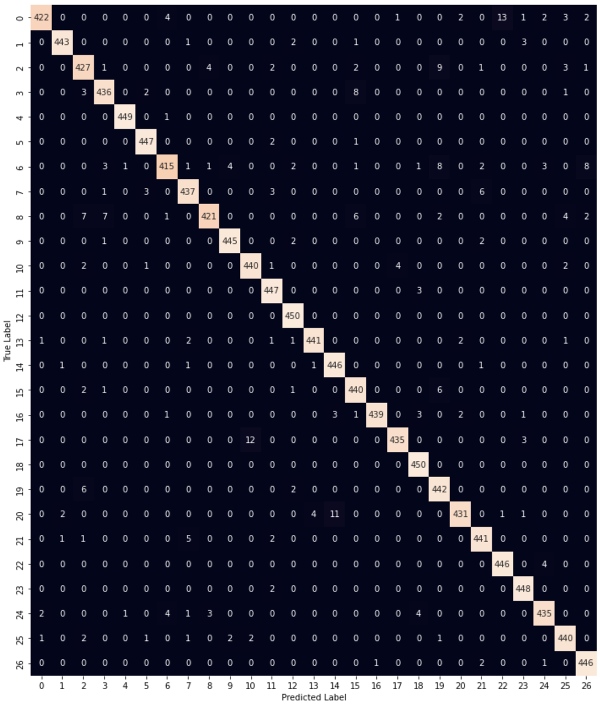


Fig. 5 Accuracy and loss of the model training: (a) accuracy vs. epoch and (b) loss vs. epoch



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
0	422	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	13	1	2	3	2
1	0	443	0	0	0	0	0	1	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	3	0	0	0
2	0	0	427	1	0	0	0	0	4	0	0	2	0	0	0	2	0	0	0	9	0	1	0	0	0	3	1
3	0	0	3	436	0	2	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	1	0
4	0	0	0	0	449	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	447	0	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	3	1	0	415	1	1	4	0	0	2	0	0	1	0	0	1	8	0	2	0	0	3	0	8
7	0	0	0	1	0	3	0	437	0	0	0	3	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0
8	0	0	7	7	0	0	1	0	421	0	0	0	0	0	0	6	0	0	0	2	0	0	0	0	0	4	2
9	0	0	0	1	0	0	0	0	0	445	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0
10	0	0	2	0	0	1	0	0	0	0	440	1	0	0	0	0	0	4	0	0	0	0	0	0	0	2	0
11	0	0	0	0	0	0	0	0	0	0	0	447	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	450	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	1	0	0	1	0	0	0	2	0	0	0	1	1	441	0	0	0	0	0	0	2	0	0	0	0	1	0
14	0	1	0	0	0	0	0	1	0	0	0	0	0	1	446	0	0	0	0	0	0	1	0	0	0	0	0
15	0	0	2	1	0	0	0	0	0	0	0	0	1	0	0	440	0	0	0	6	0	0	0	0	0	0	0
16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	1	439	0	3	0	2	0	0	1	0	0	0
17	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	435	0	0	0	0	0	3	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	450	0	0	0	0	0	0	0	0
19	0	0	6	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	442	0	0	0	0	0	0	0
20	0	2	0	0	0	0	0	0	0	0	0	0	0	4	11	0	0	0	0	0	431	0	1	1	0	0	0
21	0	1	1	0	0	0	0	5	0	0	0	2	0	0	0	0	0	0	0	0	0	441	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	446	0	4	0	0
23	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	448	0	0	0
24	2	0	0	0	1	0	4	1	3	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	435	0	0
25	1	0	2	0	0	1	0	1	0	2	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	440	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	1	0	446

Fig. 6 Confusion matrix of fruits and vegetables

The confusion matrix's index 0-26 are the names of the fruits and vegetables as shown in Table 1. For example, indices 21 and 22 are pineapple and pomegranate respectively. The performance of the proposed model is shown in figure 6. The highest misclassification class was chili with 35 images followed by cucumber with 29 images of misclassification. However, all images of grapes and orange classes were correctly classified followed by carrot with one misclassified image as shown in figure 6. It was observed that 13 images of apples were classified as pomegranate, 12 images of onion as garlic, and 11 images of pear as lemon. We found that the probability of misclassification is higher with classes having similar features such as apple and pomegranate.

Table 1. Fruits and vegetable name in English and Dzongkha

Index	**English**	**Dzongkha**
0	Apple	ཨེ་པླལ
1	Banana	ངང་ལག
2	Bean	སྲནམ་ཙུམ
3	Cabbage	འདབ་མ་གོ་བི
4	Carrot	ལ་ཕུག་དམརཔོ
5	Cauliflower	མེ་ཏོག་གོ་བི
6	Chili	ཨེ་མ
7	Corn	ཨ་ཤོམ
8	Cucumber	གོན
9	Eggplant	ངོ་ཡོམ
10	Garlic	སྒྱུ་སྒོགཔ
11	Ginger	ས་སྒ
12	Grapes	ཆུ་རྒུན་འབྲུམ
13	Kiwi	ཡོ་བིཀྲི
14	Lemon	ཏུམ་ཙུང
15	Lettuce	ཏོ་སྐྱེས་པད་ཚོད
16	Mango	ཨམ་ཙུ་ཀུ་ལི
17	Onion	སྒོགཔ
18	Orange	ཚལ་ལུ
19	Pea	བོད་སྲནམ
20	Pear	སླི
21	Pineapple	རྒྱ་ནག་གོང་ཙེ
22	Pomegranate	སེན་འབྲུ
23	Potato	ཀེ་བ
24	Tomato	ལྷམ་བན་ད
25	Turnip	གཡུང་ཏོག
26	watermelon	ཁ་རི་སྦུ་ཟ

Table 2 displays the classification report of each class. The carrot has a distinct shape and colour. Consequently, the carrot class was observed with the highest percentage of classification. However, peas, apples, chili, and beans obtained the lowest precision, recall, and f1-score respectively. Nevertheless, the weighted average of all the classes was 98%.

Table 2. Classification report consisting of precision, recall, and f1-score

Class	Precision (%)	Recall (%)	F1-score (%)
apple	0.99	0.94	0.96
banana	0.99	0.98	0.99
bean	0.95	0.95	0.95
cabbage	0.97	0.97	0.97
carrot	1.0	1.0	1.0
cauliflower	0.98	0.99	0.99
chili	0.97	0.92	0.95
corn	0.97	0.97	0.97
cucumber	0.98	0.94	0.96
eggplant	0.99	0.99	0.99
garlic	0.97	0.98	0.97
ginger	0.97	0.99	0.98
grape	0.98	1.0	0.99
kiwi	0.99	0.98	0.98
lemon	0.97	0.99	0.98
lettuce	0.96	0.98	0.97
mango	1.0	0.98	0.99
onion	0.99	0.97	0.98
orange	0.98	1.0	0.99
pea	0.94	0.98	0.96
pear	0.99	0.96	0.97
pineapple	0.97	0.98	0.97
pomegranate	0.97	0.99	0.98
potato	0.98	1.0	0.99
tomato	0.98	0.97	0.97
turnip	0.97	0.98	0.97
watermelon	0.97	0.99	0.98
weighted avg	**0.98**	**0.98**	**0.98**

The graphical user interface was designed using the *gradio* library. Images can be uploaded or dragged and dropped to predict the class. The model predicts the class as shown in figure 7 and figure 8 on clicking the *submit* button. To test with different fruit and vegetable images, first clear the uploaded image by clicking on the *clear* button.

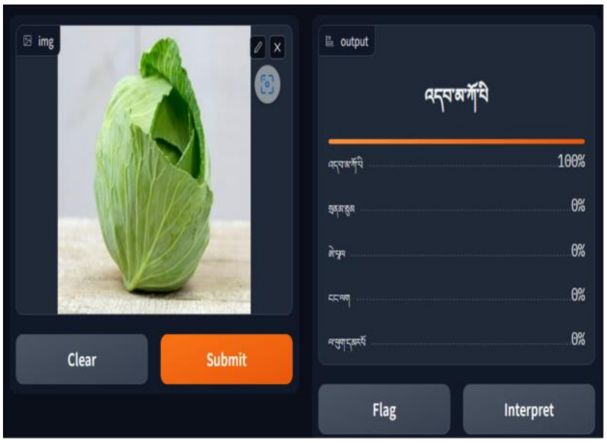


Fig. 7 Cabbage detection by the system

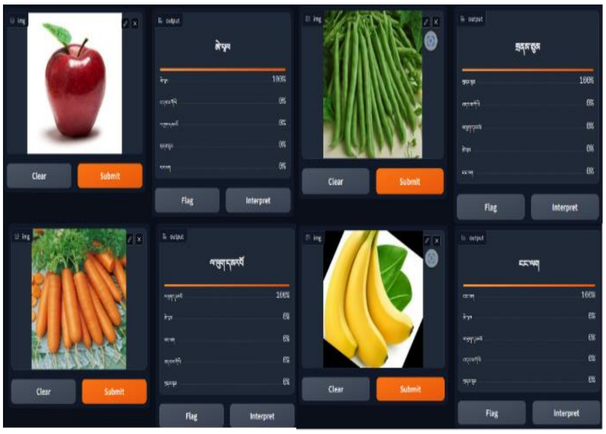


Fig. 8 Fruits and vegetables recognized in Dzongkha using the system

5. CONCLUSION

In this study, fruits and vegetable recognition system in Dzongkha was developed and tested using the VGGNet model. The proposed VGG-8 model consisted of six convolutional layers and two dense layers. The convolutional layers were divided into three blocks. In each block, two convolutional layers were stacked consecutively followed by batch normalization, max-pooling, and dropout. The early stopping mechanism was used and training stopped after 50 epochs. The training and validation accuracy was 97.76% and 97.80%% respectively with the minimum training loss of 0.746. However, the performance of the model requires further improvement.

The performance of the model can be enhanced by using high-resolution images and deeper layers. In addition, a new image dataset could be collected with varying angles and distances to increase the efficiency of the recognition system. Furthermore, various classification models could be evaluated such as the YOLO and deploy using mobile applications.

REFERENCES

Du, Z., Zeng, X., Li, X., Ding, X., Cao, J., & Jiang, W. (2020). Recent advances in imaging techniques for bruise detection in fruits and vegetables. *Trends in Food Science & Technology*, *99*, 133–141. https://doi.org/10.1016/J.TIFS.2020.02.024

Gomes, J. F. S., & Leta, F. R. (2012). Applications of computer vision techniques in the agriculture and food industry: A review. *European Food Research and Technology*, *235*(6), 989–1000. https://doi.org/10.1007/S00217-012-1844-2

Jaiswal, A., Pathak, S., Rathore, Y. K., Janghel, R. R., Jaiswal, C. A., Pathak, · S, Rathore, Y. K., Janghel, · R R, Pathak, S., & Janghel, R. R. (2021). Detection of disease from leaf of vegetables and fruits using deep learning technique. *Springer*, 199–206. https://doi.org/10.1007/978-981-15-6329-4_18

Jin, X., Sun, Y., Che, J., Bagavathiannan, M., Yu, J., & Chen, Y. (2022). A novel deep learning-based method for detection of weeds in vegetables. *Pest Management Science*. https://doi.org/10.1002/PS.6804

Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of 'MangoYOLO.' *Precision Agriculture*, *20*(6), 1107–1135. https://doi.org/10.1007/S11119-019-09642-0

Kuensel. (2017, May 22). Education policy Dzongkha curriculum questioned. *Kuensel*. https://kuenselonline.com/education-policy-dzongkha-curriculum-questioned/?msclkid=85d5d2d1b8a411ec83d3334c22fb37cf

Mai, X., Zhang, H., Jia, X., & Meng, M. (2020). Faster R-CNN with classifier fusion for automatic detection of small fruits. *Ieeexplore.Ieee.Org*. https://doi.org/10.1109/TASE.2020.2964289

Muresan, H., & Oltean, M. (2018). Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*, *10*(1), 26–42. https://doi.org/10.2478/ausi-2018-0002

Sa, I., Ge, Z., Dayoub, F., Upcroft, B., Perez, T., Mccool, C., Oliver-Codina, G., Gracias, N., & López, A. M. (2016). Deepfruits: A fruit detection system using deep neural networks. *Mdpi.Com*. https://doi.org/10.3390/s16081222

Seth, K. (2020). *Fruits and Vegetables Image Recognition Dataset*. Kaggle. https://www.kaggle.com/kritikseth/fruit-and-vegetable-image-recognition

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. http://www.robots.ox.ac.uk/

Song, Y., Glasbey, C., Horgan, G., Polder, G., Dieleman, J., & Heijden, G. van der. (2014). *Automatic fruit recognition and counting from multiple images*. Biosystems Engineering. https://doi.org/10.1016/j.biosystemseng.2013.12.008

Valentino, F., Cenggoro, T., & Pardamean, B. (2021). A Design of Deep Learning Experimentation for Fruit Freshness Detection. *Earth and Environmental Science*. https://doi.org/10.1088/1755-1315/794/1/012110

Van Driem, G. (2014). Language and identity in Bhutan Related papers. *Druk Journal*, *1*(1), 61–67.

Varghese, R., Jacob, P., S, S., Ranjan, D., Varughese, J., & Raju, H. (2021). Detection and Grading of Multiple Fruits and Vegetables Using Machine Vision. *Ieeexplore.Ieee.Org*.

Wan, S., & Goudos, S. (2020). Faster R-CNN for multi-class fruit detection using a robotic vision system. *Computer Networks*, *168*, 107036. https://doi.org/10.1016/J.COMNET.2019.107036

Zeng, G. (2017). Fruit and vegetables classification system using image saliency and convolutional neural network. *Ieeexplore.Ieee.Org*, 613–617.