Fruit shop tool: Fruit classification and Recognition using Deep Learning

R. Shantha Selvakumari*, V.Gomathy

(Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi-626005)

Abstract: Fruit image classification and recognition is a challenging application of computer vision. The computer vision system is used to recognize a fruit based on artificial neural networks. Deep neural network is widely used for various classification problems. In this paper convolutional neural network (CNN) is used to recognize the fruits. The dataset contains 1877 images of ten categories which are used for the experimental purpose. CNN is constructed with sixteen layers which are used to extract the features from images and support vector machine (SVM) classifier is used for classification. The proposed system has the classification accuracy of 99.2% and the recognition accuracy of 99.02%.

Keywords: fruit recognition, deep learning, convolution neural network, support vector machine.

Citation: R. Shantha Selvakumari1 and V.Gomathy . 2023. Fruit shop tool: Fruit classification and recognition using deep learning. Agricultural Engineering International: CIGR Journal, 25(2): 312-321.

1 Introduction

In fruit industries, it is required to construct an automatic fruit recognition system to recognize the different kinds of fruits. Fruit classification and recognition system can be used as a tool for children to improve learning capabilities. It can also be used in supermarket applications and fruit industries.

In traditional technique (Zawbaa et al. 2014, Garcia et al. 2016) the extracted features may over specified and take a long time to design and validate. All the extracted features without selecting are used for the classification of fruits which causes classification error. In order to avoid the misclassification error, it is necessary to select the appropriate features from the extracted features (Zeiler and Fergus, 2014). So that the unwanted features are not used for fruit recognition and the classifier uses only the best selected features. Feature selection requires expert knowledge and it consumes more time. To overcome the above problem in traditional technique, deep learning architecture is implemented.

Convolutional neural network (CNN) is a type of deep neural network. In deep learning approach, hand-crafted feature extraction is eliminated. In CNN, the fruit image is directly given as an input to the network (Zeiler and Fergus, 2014). It avoids the data reconstruction process in traditional technique. The basic structure of CNN contains: convolutional layers, pooling layers and fully connected layers. Convolution layer performs the convolution operation over the input and works on every part of the image. Pooling layers are usually present after the convolutional layer and it provides a down sampling operation. Fully connected layer is the last stage of CNN.

This paper proposes a classification and recognition system which is based on convolutional neural network.

Received date: 2021-08-04 **Accepted date:** 2022-10-27

***Corresponding author: R. Shantha Selvakumari,** Senior Professor and Head, Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi-626005. Tel :+91 9486636774. Email: rshantha@mepcoeng.ac.in.

This fruit recognition system has five convolutional layers, five rectified linear layers, three pooling layers, one dropout layer and one fully connected layer. For classification and recognition, support vector machine (SVM) classifier is used.

The structure of the paper is as follows: Section 2 presents related work, section 3 provides methodology. Section 4 describes the results and discussion. This paper ends with conclusion in section 5.

2 Related works

There are few works that address the classification and recognition of fruits using machine learning algorithm. The related works for the fruit recognition, found in the state of art are presented.

An artificial vision technique is used to classify strawberry fruit for industrial application (Constant et al. 2016). Deep learning with multiple stage neural networks is implemented to classify the strawberry fruit. Each stage consists of three layers. Back propagation method is used to train the network.

Deep Neural Network (DNN) is implemented to recognize the different varieties of vegetables (Sakai et al. 2017). Vegeshop tool for vegetable recognition which is based on CNN is developed. To train the network, Back propagation method is used. It has multiple hidden layers between input and output layer.

CNN based fruit recognition system is implemented (Hou et al. 2016). By using selective search algorithm, image regions are extracted. To train the network, region information is used. Softmax classifier is used to classify the different kinds of fruits. Back propagation network is used to learn and store the input and output. The weight is to be adjusted to minimize the misclassification error.

Deep CNN is used to detect the fruit images (Sa, I., Z. Ge et al. 2016). Faster Region-based convolutional neural network (R-CNN) is used for fruit detection. The faster R-CNN uses RGB image to perform object detection. Region of interest is extracted from the image and give it to the convolutional layers. VGG-16 model is used. Softmax classifier is used to classify the different kinds of fruits.

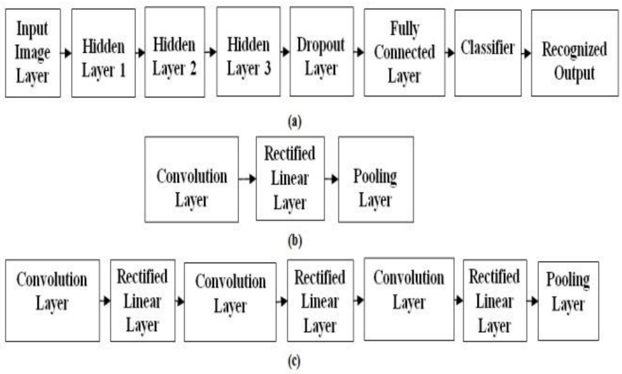
A vegetable category recognition system using deep neural network is implemented (Sakai et al. 2016). Different kinds of vegetable images were recognized by using CNN. Eight different categories of vegetables are used for recognition. The recognition system is learned by using back propagation in a multi-layer neural network. By using the convolution and pooling operations, CNN recognize the optimized features. To speed up the training process, CNN uses rectified linear units.

3 Methodology

The block diagram of the proposed method is shown in Figure 1. In CNN, fruit image is directly given as an input to the network (Yanai et al., 2015). There is no need for feature extraction and feature selection. So, hand-crafted feature extraction is eliminated in CNN. CNN can avoid the data reconstruction process. The fruit recognition system using CNN is discussed which contains many layers: convolutional layers, pooling layers, rectified linear layers, dropout layer and fully connected layer.

3.1 Image input layer

In CNN, image input layer is the first layer. Image is directly given to the CNN. It provides the advantage that, there is no need for feature extraction. The input image size is 227×227×3. These numbers correspond to the height, width and channel size. Channel size of the color image is 3, which corresponds to RGB value.



(a) Convolutional neural network (b) Hidden layer 1 & 2 (c) Hidden layer 3

Figure 1 (Proposed Block Diagram)

3.2 Convolutional layer

It performs the convolution operation over the input and works on every part of the image. It has two parameters, filter size and number of filter. The size of the filter is h×w, which represents the height and width of the filter. The number of filter provides the number of neurons that connect to the input. Totally five convolutional layers are constructed. First layer has 96 filters with the size of 3×3. Second layer consists of 256 filters with the size of 5×5. 384 filters with the size of 3×3 are used in third and fourth layer. Layer five contains 256 filters of size 3×3.

Given an image of size $W_1 \times H_1$, the output image size can be calculated as,

$$W_2 = \frac{W_1 - F + 2P}{S} + 1 \quad (1)$$

$$H_2 = \frac{H_1 - F + 2P}{S} + 1 \quad (2)$$

Where W_2 is the output image width, W_1 is the input image width, H_2 is the output image height, H_1 is the input image height, F is the filter size, P is the padding value and S is the stride value.

3.3 Rectified linear layer

Convolutional layer is followed by non-linear layer or activation layer. The main purpose of rectified linear layer is to introduce the nonlinearity to the system. It is also used to overcome the vanishing gradient problem. This layer applies the function $f(x)$ to all the input values (Fard et al., 2016).

$$f(x) = \max(0, x) \quad (3)$$

This layer changes all the negative activations to 0. This layer can increase the nonlinear properties of the model. Totally, five rectified linear layers are used in this deep learning architecture.

3.4 Pooling layer

Rectified linear layer is followed by a down sampling operation that reduces the size of the feature map and eliminates redundant information (Nagi et al., 2011; Wu et al., 2015). One method of down sampling is to use a pooling layer. Pooling layer can combine the cluster of neurons at one layer to a single neuron in the next layer. Max pooling is one of the pooling methods which is used in CNN. Max pooling can return the maximum value from each neuron cluster. Pooling size will represents the neuron cluster size. The pool size which is used in the CNN is 3×3. It provides the maximum number in every sub region of size 3×3. Totally three pooling layers are used in this architecture. The stride of different size in different layers and zero padding are used.

Given an image of size $W_1 \times H_1$, the output image size can be calculated as,

$$W_2 = \frac{W_1 - F}{S} + 1 \quad (4)$$

$$H_2 = \frac{H_1 - F}{S} + 1 \quad (5)$$

Where W_2 is the output image width, W_1 is the input image width, H_2 is the output image height, H_1 is the input image height, F is the filter size and S is the stride value.

3.5 Dropout layer

Dropout is a technique used to reduce the over fitting problem. The term dropout refers to dropping out of some hidden units in the network. It will temporarily remove the units and its connection by random choice. The key idea of the dropout layer is to randomly drop some units and its connections from the network during training phase to avoid the over fitting problem. Here, 50% dropout is performed. It is used to improve the performance of the CNN.

3.6 Fully connected layer

Fully connected layer is the last stage of CNN. The fully connected layer connects all the neurons in the previous layer. The features which are used for classification and recognition are taken from the fully connected layer which can connect the entire nodes in the previous layer to the next layer. Output size in the fully connected layer corresponds to the number of categories. Here, the output size is ten, which represent ten different categories of fruits.

3.7 Classifier

SVM classifier is used to classify the different kinds of fruit images (Mahajan et al., 2016; Wu et al., 2014). SVM is an efficient tool for high dimensional space. SVM is used for linear as well as non-linear classifications. If the features are non-linear, then it is mapped into high dimensional space. After mapped into high dimensional, those features are easily separable. Separation between the classes is highly non-linear. So there is a flexibility in the decision boundaries, which leads the classifier to perform better.

4 Results and discussion

4.1 Data set

The Supermarket Produce data set contains 10 different categories of fruits: Cashew, Diamond Peach, Fuji Apple, Granny Smith Apple, Honeydew Melon, Kiwi, Nectarine, Orange, Plum, and Spanish Pear. Totally 1877 images are available in the dataset which are used in CNN. Figure 2 shows the different kinds of fruits in the data set.

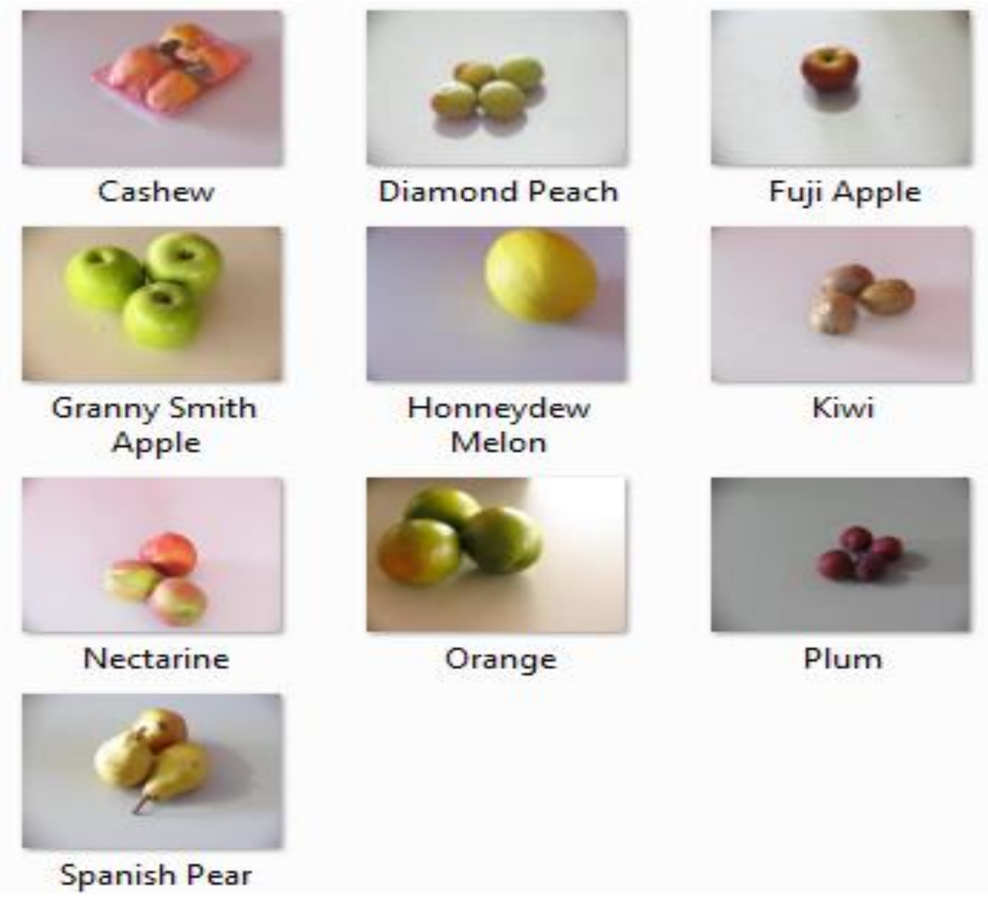


Figure 2 Dateset used in CNN

The Supermarket Produce data set images are collected at different time, different day for similar category of fruit class. The data set contains images with different lighting. This data set also contains images with difference in position and difference in number of elements in an image.

The implemented CNN structure is shown in the Figure 3. Image input layer is the first layer in CNN. The size of the input layer is specified as 227×227×3. These numbers correspond to the height, width, and the channel size respectively. For color images, the channel size is 3, corresponding to the RGB values.

Input layer is followed by convolutional layer. In the convolutional layer, the first parameter is filter size and the second parameter is the number of filters, which is already discussed in section 3. Convolutional layers are followed by rectified linear layers. It is used to overcome the vanishing gradient problem.

Next, it is followed by down-sampling operation that reduces the spatial size of the feature map and removes redundant spatial information. This layer is called pooling layer. Here, max pooling is used and the size of the rectangular region is [3, 3].

Dropout layer is present after the max pooling layer. The last layer in the CNN is the fully connected layer. The output size parameter in the last fully connected layer is equal to the number of classes in the target data. Here, the output size is 10, corresponding to the 10 classes. SVM classifier is used to classify the different category of fruit classes in the dataset. Table 1 shows the layer description and its input and output dimensions.

Table 2 shows the types and number of fruits in the dataset. The plum has maximum number of fruits and the orange has least number of fruits in the dataset.

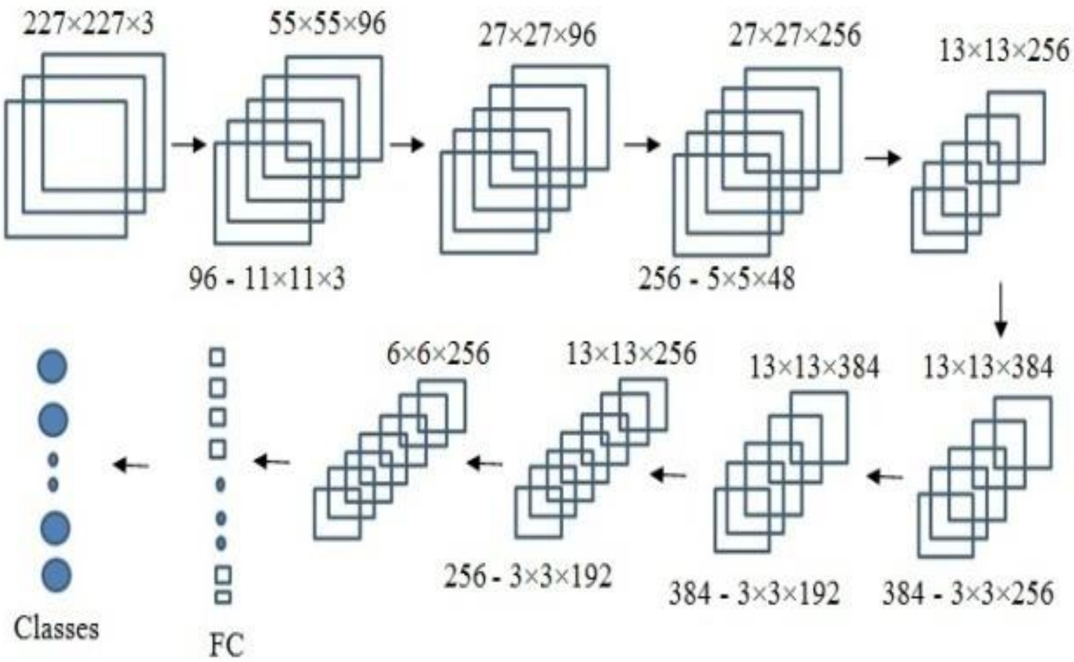


Figure 3 Convolutional neural network structure

Table 1 Layer description and dimension

Layer Description	Input Dimension	Output Dimension
Convolution Layer 1	227×227×3	55×55×96
Pooling Layer 1	55×55×96	27×27×96
Convolution Layer 2	27×27×96	27×27×256
Pooling Layer 2	27×27×256	13×13×256
Convolution Layer 3	13×13×256	13×13×384
Convolution Layer 4	13×13×384	13×13×384
Convolution Layer 5	13×13×384	13×13×256
Pooling Layer 3	13×13×256	6×6×256

Table 2 Types and number of fruits in Dataset

S. No.	Name of the Fruits	Total number of fruits
1	Cashew	210
2	Diamond Peach	211
3	Fuji Apple	212
4	Granny Smith Apple	155
5	Honeydew Melon	145
6	Kiwi	171
7	Nectarine	247
8	Orange	103
9	Plum	264
10	Spanish Pear	159
	Total	1877

Table 3 Results of fruit classification system

S. No.	Name of the Fruits	Classification Accuracy in %	
		50% Training 50% Testing	25% Training 75% Testing
1	Cashew	100	100
2	Diamond Peach	100	96
3	Fuji Apple	98	96
4	Granny Smith Apple	100	100
5	Honeydew Melon	100	100
6	Kiwi	100	100
7	Nectarine	98	100
8	Orange	98	100
9	Plum	100	100
10	Spanish Pear	98	96
	Overall Accuracy	99.2	98.8

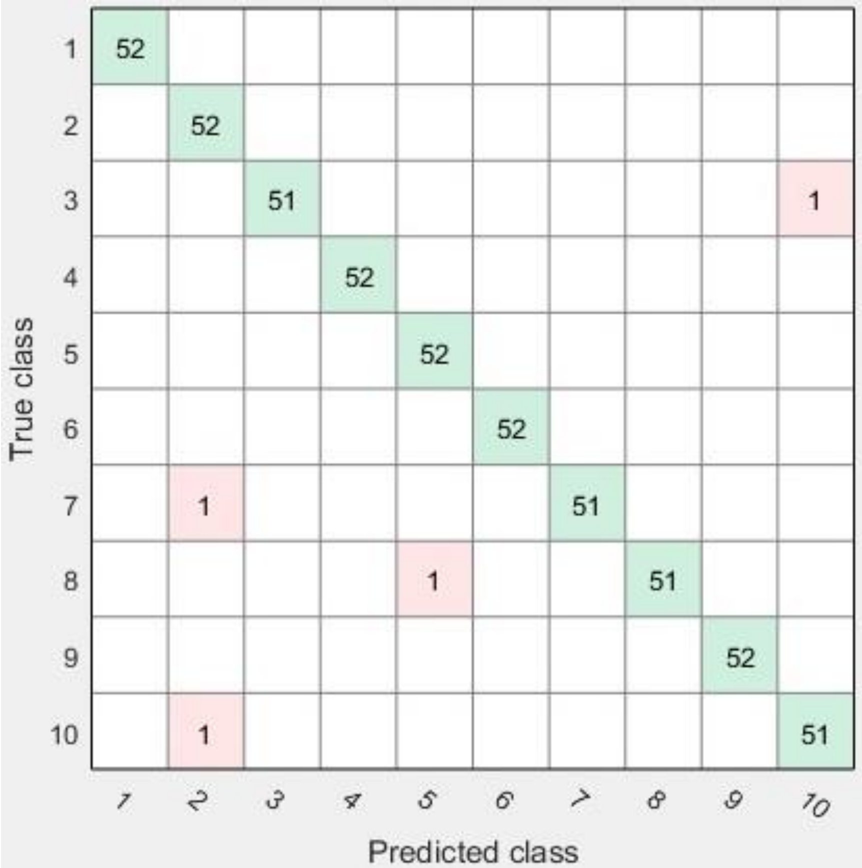


Figure 4 Confusion matrix for classification - 50% training & 50% testing

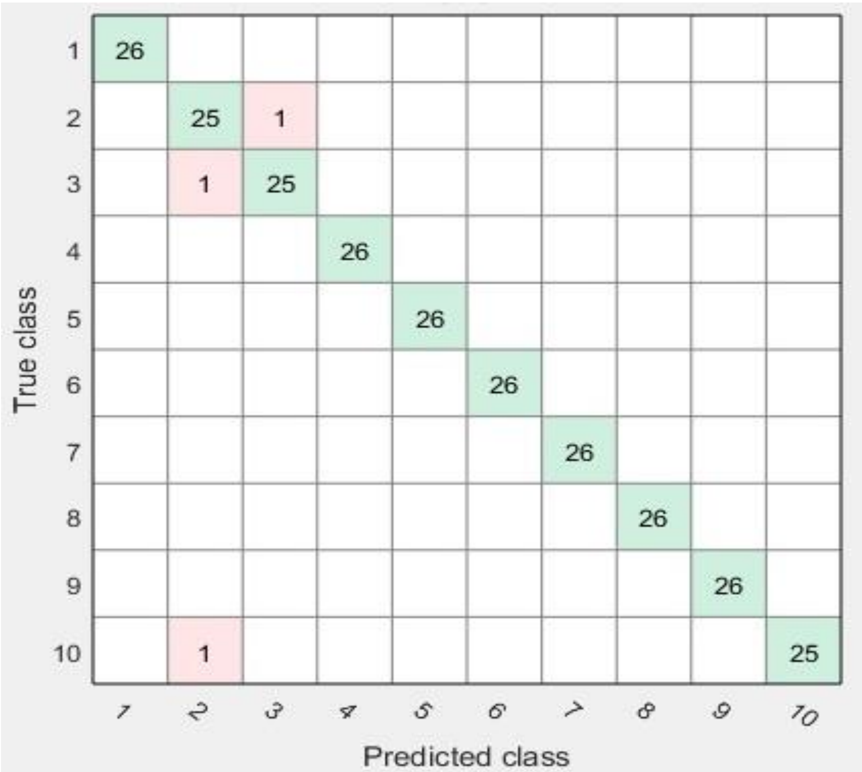


Figure 5 Confusion matrix for classification - 25% training & 75% testing

Totally, 1030 images are used for the experimental purpose. For 50% training and 50% testing, 52 images are used to train and 51 images are used to test in each class. 520 images are used for training and 510 images are used for testing. Next, for 25% training and 75% testing, 26 images are used to train and 77 images are used to test in each class. 260 images are used for training and 770 images are used for testing.

The classification accuracy of each fruit class in the data set and overall classification accuracy of the proposed system are shown in the Table 3. By using the proposed method, 100% classification accuracy is obtained for Cashew, Diamond Peach, Granny Smith Apple, Honeydew Melon, Kiwi and Plum. The remaining fruit classes are also classified with better accuracy. The overall classification accuracy of the proposed method is 99.2%.

Accuracy of the proposed system can be calculated as,

$$Accuracy = \frac{Number\ of\ images\ correctly\ recognized X 100}{Total\ number\ of\ images} \quad (6)$$

Figure 4 shows the confusion matrix for classification using 50% of different category of fruits in dataset. A confusion matrix is a matrix that is used to describe the classification model or classifier performance on a given set of data. The confusion matrix is used to identify the number of correctly classified fruit images and the number of misclassified fruit images of each fruit class. The diagonal element of the confusion matrix shows the number of correctly classified fruit images of each class in the dataset. In each fruit class, 52 images are used for classification. For Cashew, Diamond Peach, Granny Smith Apple, Honeydew Melon, Kiwi and Plum, the proposed system is correctly classified all the 52 fruit images.

Table 4 Results of fruit recognition system

S. No.	Name of the Fruits	Recognition Accuracy in %	
		50% Training 50% Testing	25% Training 75% Testing
1	Cashew	100.0	100.0
2	Diamond Peach	98.00	96.10
3	Fuji Apple	100.0	98.70
4	Granny Smith Apple	98.00	98.70
5	Honeydew Melon	100.0	100.0
6	Kiwi	98.00	94.81
7	Nectarine	98.00	100.0
8	Orange	100.0	97.40
9	Plum	98.00	100.0
10	Spanish Pear	100.0	98.70
	Overall Accuracy	99.02	98.44

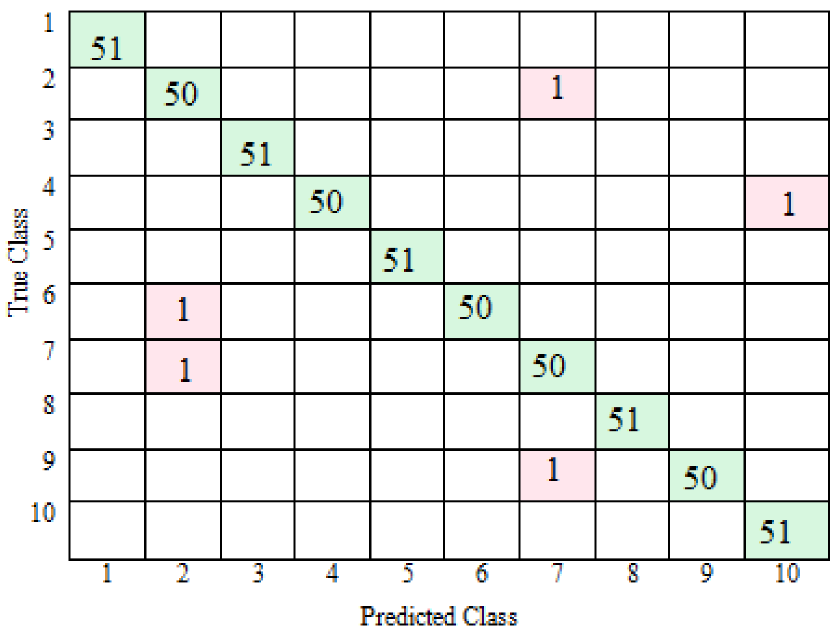


Figure 6 Confusion matrix for recognition - 50% training & 50% testing

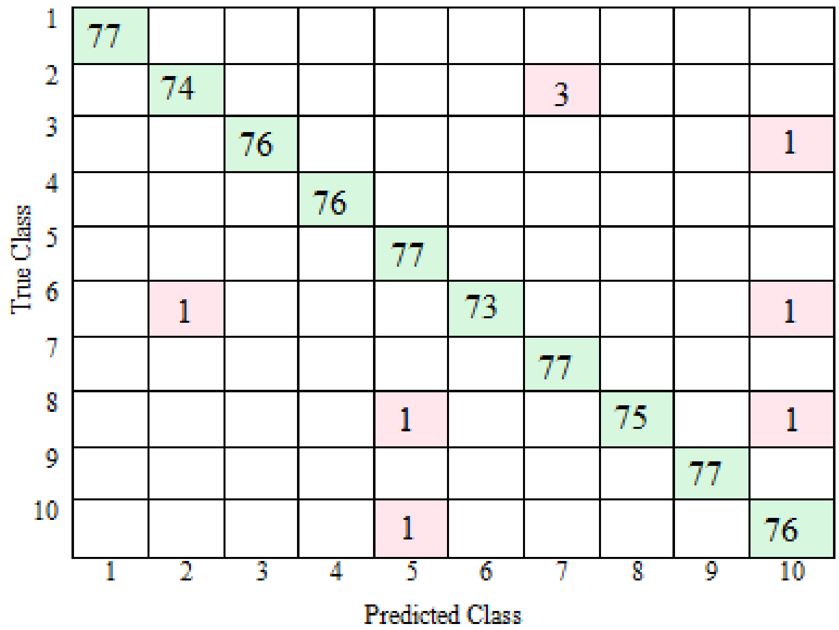


Figure 7 Confusion matrix for recognition - 25% training & 75% testing

Table 5 Comparison of overall accuracy for fruit recognition

Reference Paper	Methodology	Recognition Accuracy in %
Proposed	CNN + SVM	99.02
Cornejo, J.Y.R. and Pedrini (2017)	CENTRIST and HS-Histogram representation + SVM	97.23
Dubey et al. (2015)	Global Color Histogram + Local Binary Patterns + SVM	93.84

Figure 5 shows the confusion matrix for classification using 25% of different category of fruits in dataset. In each fruit class, 26 images are used for classification. All the 26 images of Cashew, Granny Smith Apple, Honeydew Melon, Kiwi, Nectarine, Orange and Plum are correctly classified by the proposed system.

Table 4 shows the recognition accuracy of each fruit class in the dataset and overall recognition accuracy of the proposed system. By using the proposed method, 100% recognition accuracy is obtained for Cashew, Fuji Apple, Honeydew Melon, Orange and Spanish Pear. The other fruit classes in the dataset are also recognized well by using the proposed system. The overall recognition accuracy of the proposed method is 99.02%.

Figure 6 shows the confusion matrix for recognition using 51 images in each category of fruits in the dataset. For Cashew, Fuji Apple, Honeydew Melon, Orange and Spanish Pear, the proposed system is correctly recognized all the 51 fruit images which leads to 100% accuracy for these fruit images.

Figure 7 shows the confusion matrix for recognition using 77 images in each category of fruits in the dataset. For Cashew, Honeydew Melon, Nectarine and Plum, all the 77 images are correctly recognized by the proposed system.

Table 5 shows the performance comparison results on the Supermarket Produce data set. In (Dubey et al., 2015)–background subtraction, feature extraction and classification are the three steps involved in fruit and vegetable recognition. For background subtraction, K-means clustering-based image segmentation is used. For feature extraction, global color histogram and local binary patterns are used. SVM classifier is used for classification and recognition. The overall recognition accuracy is 93.84%.

In (Cornejo et al. 2017), preprocessing, feature extraction, feature reduction and classification are the four stages used in the vegetable classification system. CENTRIST, color CENTRIST and hue - saturation histogram representation are used for feature extraction. In feature reduction, principal component analysis (PCA) is used. SVM classifier is used for classification and recognition. The overall recognition accuracy is 97.23%.

By using the CNN, the proposed method obtained

99.02% recognition accuracy. The obtained accuracy is higher when compare to that of two papers accuracy.

4 Conclusion

Fruit classification and recognition system is proposed based on CNN. The proposed method contains many layers: convolutional layers, pooling layers, rectified linear layers, dropout layer and fully connected layer. SVM classifier is used to classify the different kinds of fruit images in the dataset. Totally, 1030 images are used for the experimental purpose. CNN randomly splits the training as well as testing data. The accuracy value is calculated by using both 50% training & 50% testing and 25% training & 75% testing in the given dataset. For 25% training and 75% testing, 260 images are used for training and 770 images are used for testing. The overall classification accuracy and recognition accuracy for the proposed system is 98.8% and 98.44% respectively. For 50% training and 50% testing, 520 images are used for training and 510 images are used for testing. The overall classification accuracy and recognition accuracy for the proposed system is 99.2% and 99.02% respectively.

References

Constant, P., A. Gordon, O. Chang, E. Pruna, F. Acuna, and I. Escobar. 2016. Artificial vision techniques for strawberry's industrial classification. *IEEE Latin America Transactions*, 14(6): 2576-2581.

Cornejo, J.Y.R., Pedrini, H. 2017. Automatic Fruit and Vegetable Recognition Based on CENTRIST and Color Representation. In: Beltrán-Castañón, C., Nyström, I., Famili, F. (eds) Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. CIARP 2016. Lecture Notes in Computer Science(), vol 10125. Springer, Cham. https://doi.org/10.1007/978-3-319-52277-7_10.

Dubey, S. R., and A. S. Jalal. 2015. Fruit and vegetable recognition by fusing colour and texture features of the image using machine learning. *International Journal of Applied Pattern Recognition*, 2(2): 160-181.

Fard, M. A., H. Haddadi, and A. T. Targhi. 2016. Fruits and vegetables calorie counter using convolutional neural networks. In *DH '16: Proceedings of the 6th International Conference on Digital Health Conference*, 121-122. Montréal, Québec, Canada, 11-13 April.

Garcia, F., J. Cervantes, A. López, and M. Alvarado. 2016, Fruit chromaticity classification by extracting color, shape and texture features: towards an application for supermarkets. *IEEE Latin America Transactions*, 14(7): 3434-3443.

Hou, L., Q. Wu, Q. Sun, H. Yang, and P. Li. 2016. Fruit recognition based on convolution neural network. In *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 18-22. Changsha, China, 13-15 August.

Mahajan, M. S., and S. T. Patil. 2016. Optimization and classification of fruit using machine learning algorithm. *International Journal for Innovative Research in Science & Technology*, 3(1): 377-381.

Nagi, J., F. Ducatelle, G. A. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella. 2011. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications*, 342-347. Kuala Lumpur, Malaysia, 16-18 November.

Sa, I., Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool. 2016. Deepfruits: A fruit detection system using deep neural networks. *Sensors*, 16(8): 1222.

Sakai, Y., Oda, T., Ikeda, M., Barolli, L. 2017. VegeShop Tool: A Tool for Vegetable Recognition Using DNN. In: Barolli, L., Xhafa, F., Yim, K. (eds) Advances on Broad-Band Wireless Computing, Communication and Applications. BWCCA 2016. Lecture Notes on Data Engineering and Communications Technologies, vol 2. Springer, Cham. https://doi.org/10.1007/978-3-319-49106-6_68.

Sakai, Y., T. Oda, M. Ikeda, and L. Barolli. 2016. A vegetable category recognition system using deep neural network. In *2016 10th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 189-192. Fukuoka, Japan, 6-8 July.

Wu, D., C. Li, J. Chen, D. You, and X. Xia. 2014. A novel multi-class support vector machines using probability voting strategy and its application on fault diagnosis of gearbox. *International Journal of Performability Engineering*, 10(2): 173-186.

Wu, H., and X. Gu. 2015. Max-pooling dropout for regularization of convolutional neural networks. In *Neural Information Processing*, https://www.researchgate.net/publication/308761865_Max-Pooling_Dropout_for_Regularization_of_Convolutional_Neural_Networks, Springer International Publishing.

Yanai, K., and Y. Kawano. 2015. Food image recognition using deep convolutional network with pre-training and fine-tuning. In *2015 IEEE International Conference on Multimedia & Expo Workshops*, 1-6. Turin, Italy, 29 June - 03 July.

Zawbaa, H. M., M. Abbass, M. Hazman, and A. E. Hassenian. 2014. Automatic fruit image recognition system based on shape and color features. In *Advanced Machine Learning Technologies and Applications. AMLTA 2014. Communications in Computer and Information Science, vol 488*, eds. A. E. Hassanien, M. F. Tolba, and A. Taher Azar, 278-290. Switzerland: Springer International Publishing.

Zeiler, M.D., Fergus, R. 2014. Visualizing and Understanding Convolutional Networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8689. Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53.