Deep Learning Architecture For Fruit Classification

Sangeetha B[1], Senthil Prabha R[2] and Ravitha Rajalakshmi N[3]
{bsg.it@psgtech.ac.in[1], rsp.it@psgtech.ac.in[2], nrr.it@psgtech.ac.inaddress[3]}

Assistant Professor (Senior Grade), Department of Information Technology, PSG College of Technology, Coimbatore, India, TamilNadu - 641004[1,2,3]

Abstract. The agricultural industries are one of the cost demanding fields placing a requirement on skilled laborers for harvesting. To meet the demands, robots are employed to harvest which mandates the need for accurate fruit detection system. The robot has to scan the image and recognize the fruit, which is the crucial process as the recognition system faces unprecedented challenges like occlusion, deformation, illumination conditions. The objective of this work is to build an accurate and reliable fruit recognition system by addressing these challenges in image recognition. Convolutional neural network, a deep learning algorithm is used to identify the features of an image and classify the image in the fruit recognition system. The system is evaluated with Fruit-360 dataset consisting of 43329 images of 60 different categories. With the aid of the proposed system, quantifiable improvement of about 97% accuracy is achieved and the total loss of the system is about 0.13.

Keywords: Image Recognition; Deep Learning; Convolutional Neural Network; Fruit Recognition system.

1 Introduction

Harvesting is considered to be a labour intensive task in the agro industries as it posts the requirement of skilled seasonal labours. Many countries are marching towards deploying fruit picking robots, in order to reduce the labour cost on harvesting. The robots are trained with fruit detection system [1], which detects the images and identifies the fruit and plucks it. The trained robots [9] show significant performance when compared to the average performance of the skilled worker. However, the robots take up a back stage to meet the demands of the practical applications like speed, accuracy and low-cost. The critical step in the development of automated harvesting system [12] is to develop an accurate fruit detection system [3], which can identify the fruits among the other parts of the plant and pick the appropriate fruits.

Image classification [14] is the task of assigning one label to the image from a fixed set of categories. The process in image recognition is two faceted: feature extraction and classification. Extracting high perceptual data from the pixels of an image is feature extraction. This phase is usually done with varied algorithms and the extracted features are fed to the machine learning algorithms [2] for classification. The performance of the classification algorithm relies on the extracted features. Recognizing an image is trivial in the human perspective; however there exist a number of challenges in image recognition and is stated as follows: (i) Image occlusion (ii) Illumination conditions (iii) Background clutter (iv) Deformation (v) Scale variation and (vi) Intra-class variation. An accurate image identification system must be invariant to the aforementioned challenges.

ICCAP 2021, December 07-08, Chennai, India

DOI 10.4108/eai.7-12-2021.2314483

The concept of machine learning is been widely adopted in diverse areas such as computer vision, natural language processing, computational intelligence due to its inherent ability in learning from the data and make data-driven decisions. The concept of deep learning rooted from artificial neural networks [4] has started to gain prominence in recent years,. The Deep Neural Network (DNN) [7] gains its popularity as it follows a layer-by-layer greedy approach in search for the features from the input data. It provides high level of abstraction of the data. This characteristic has gained its importance in various fields for extracting the features from the input data without human intervention. In addition, the technology growth has also complemented the applicability of DNN, as it requires high computational power, large storage and parallel processing. The architectures proposed for deep learning neural network are Autoencoders, Deep belief network, Deep Boltzman machine, Recurrent Neural Network and Convolutional Neural Network [8]. The architectures of these DNN and their pros and cons are discussed in Section 3.

The research work employs a deep convolutional neural network algorithm [10] for training the system to recognize image subject to several variations. The performance of the system is evaluated using the fruit dataset 360. The evaluation metrics used are the accuracy and categorical cross entropy loss.

The main contributions of this work include:

- Developing a robust image detection system with higher accuracy trained using deep convolutional neural network.
- Image recognition system addressing occlusion, illumination conditions and deformation variations.

The reminder of this paper is organized as follows: Section 2 depicts various DNN architecture, their pros and cons. Section 3 shows the proposed model for image recognition. The experimental setup and the research findings are discussed in Section 4. The conclusion and future directions are highlighted in Section 5.

2 Convolutional Neural Network Architecture

A number of deep learning architecture [11] put forth in the literature are in widespread use across diverse applications. This section explores five architectures which are extensively used. Notably, Autoencoders, Deep belief network, Deep Boltzman machine, Recurrent Neural Network and Convolutional Neural Network. A comparision of the varied DNN architecture is shown in Table 1.

3 Proposed Methodology

The proposed model uses convolutional neural network for fruit recognition system. The system is loaded with an input image and is pre-processed to feed it to the convolutional layer. The kernels in the convolutional layer are varied to extract multiple features. The kernel coefficients are initialized randomly before training.

Table 1: Comparison of Architectures

Architecture Type	**Principle**	**Benefits**	**Limitations**	**Applications**

Autoencoders	• A 3 layered unsupervised neural network with input, hidden and output layers. • The number of input and output nodes are equal as they are mainly used in the reconstruction of the images	• No prior knowledge on data is needed • Intuitive	• Requires pre-training • Expensive • Redesigned and retrained for each application	• Used to reconstruct images from noisy images. • Used to compress the data by utilizing the correlation
Deep Belief Network	• Multi layered connected networks. • Consists of stochastic binary variables with weighted connections	• Layered procedure for learning the weights top-down • Values of latent variables are inferred with a single bottom-up pass	• Catastrophic forgetting • Unsupervised learning may fail due to lack of knowledge and reasoning	Drug discovery
Deep Boltzmann Machine	• Bipartite connections between the nodes are established. Random weights are chosen • Undirected connection between the nodes in each layer	• Robust inference in a top down fashion	• Time it takes to settle down in equilibrium is higher • May result in variance trap • Optimization is difficult for large datasets	• Speech Recognition • Scene de-noising • Object recognition
Recurrent Neural Network	• Has memory unit to retain the computation • Weights of the neurons are the same	• Generate model for dependent tasks	• Vanishing gradient problem • expensive to apply back-propagation	NLP tasks and also in dependent tasks
Convolutional Neural Network	• Based on the biological model Convolutional layers to learn the features. Sub-sampling layers –reduce spatial dimension	• Sparse Connectivity • Parameter Sharing	Requires more training samples	image recognition, natural language processing

The basic architecture of the proposed model is presented in Figure 1 and the performance of the system is evaluated by increasing the number of convolution and sub-sampling layers. The working of convolutional neural network is presented as follows:

3.1 Convolutional Neural Network

Convolutional neural network (CNN or Convnet) is a feed forward artificial neural network which is mainly used for object recognition, natural language processing, video recognition and recommender systems. The architecture of CNN consists of input, multiple hidden layers and an output layer. The hidden layers may perform a combination of any of the operations like convolution, subsampling, normalization and fully connected layers. The CNN enhances the decision making of the network by capturing the local and global features. The working of each layer is discussed as follows:

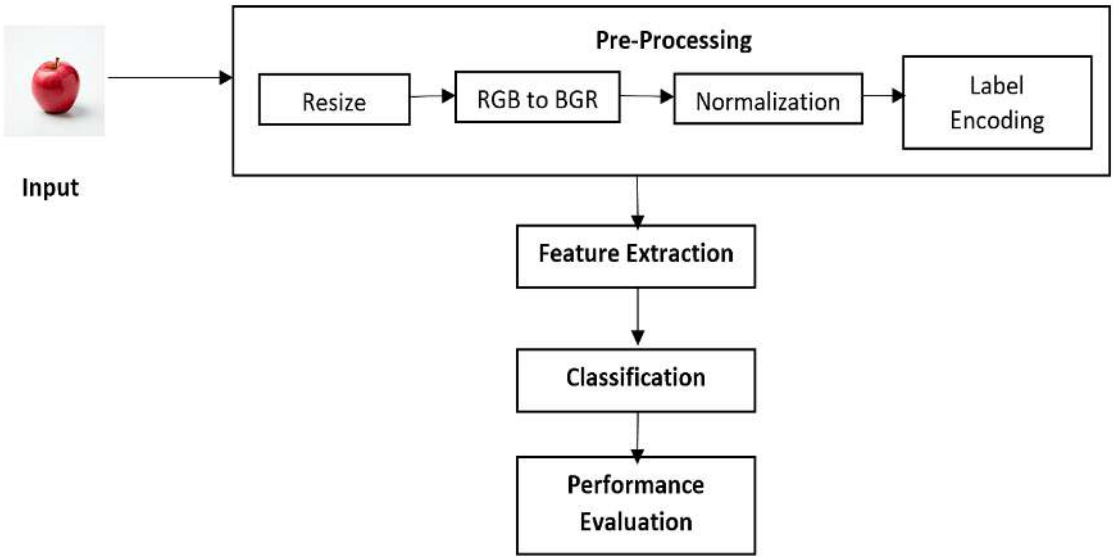


Figure 1: Design of the proposed system

3.2 Convolution layer

One of the foremost layers used in CNN is the convolutional layer. The word convolve indicates “roll together”. To mathematically state, this indicates an integral measure of how one function overlaps on the other. The convolutional layer takes in an image of size m x n x r as an input, where m, n and r are the width, height and the number of channels of an image. The numbers of channels vary based on the modality of the image. For instance, for coloured image the value of r is 3 and for gray scale image it is 1. This layer generates feature maps by using various filters or otherwise termed as kernel on the input image. The filter is an n x n x q matrix where n is smaller than the dimension of the image and q shall take the same value as the number of channels of the given image or may take a lesser value. The number of kernels and the size of the kernel may vary resulting in multiple feature maps. Each kernel is used to learn a feature from the given image.

In the process of generating the feature map, stride 1 refers to the kernel convolving on every pixel of the image. The stride value shall be varied to reduce the size of the feature map. Padding refers to the addition of zeros to the input image across the borders. This padding turns out to be beneficial if the spatial dimension of the image has to be retained otherwise the size of the feature map reduces as it moves to multiple convolutional layers than expected

thereby resulting in information loss. As each neuron produces a linear output, each feature map is then subjected to an activation function. A few popular activation functions are sigmoid, tanh, ReLU, Leaky ReLU. Table 2 shows the representation of the activation functions.

3.3 Pooling

For images of larger dimensions, the dimension of the convolved feature is also larger. Since it is difficult to handle such large dimensions of data, sub-sampling otherwise known as pooling is carried out, which also tends to retain the important features. The operations carried out in the sub-sampling layer are max or average pooling. The convolved feature is segmented into overlapping or non-overlapping regions on which the pooling operations are performed. The images are not physically segmented, however the disjoint or overlapping regions of the images are done with the help of stride value.

3.4 Fully connected layer

The fully connected layer is similar to the traditional neural network where every neuron of one layer is connected to every neuron of the subsequent layer. Unlike the convolutional and pooling layers which is used for extracting the features of the input image, this layer is used to classify the image belonging to a class based on the training data. Figure 2 shows the baseline architecture used in our experiments.

4 Experimental Results

The objective of the experiment is to aid a robot in recognizing the given image using CNN. The experiment was carried out using the Tensor flow CPU version [5]. The empirical evaluation of the system was performed using Fruit-360 [6] dataset consisting of 43329 images of size 100 x 100 pixel on 60 different categories. Python is used for purpose of implementation and opencv, matplotlib packages are used for pre-processing and plotting functionalities.

The images in this dataset were pre-processed to a size of 45 x 45 maintaining the aspect ratio. The experimentation was done with the images by narrow convolution. The number of training images is 28736 images and a validation set of 14593 images were used for evaluating the performance of the system. Figure 3 shows the sample images each of which differs by the shape, colour and orientation [13].

The following pre-processing is done on the input images and Figure 4 shows a sample of it.

- Resize: The images in the Fruit-360 dataset is of size 100 x 100 pixels with 3 channels. As the spatial volume of the image increases, to handle the data the images are resized to 45 x 45 with the same aspect ratio.
- RGB to BGR: In our implementation, OpenCV is used for image processing functions and since it uses BGR representation of the pixel information, the conversion is performed accordingly.

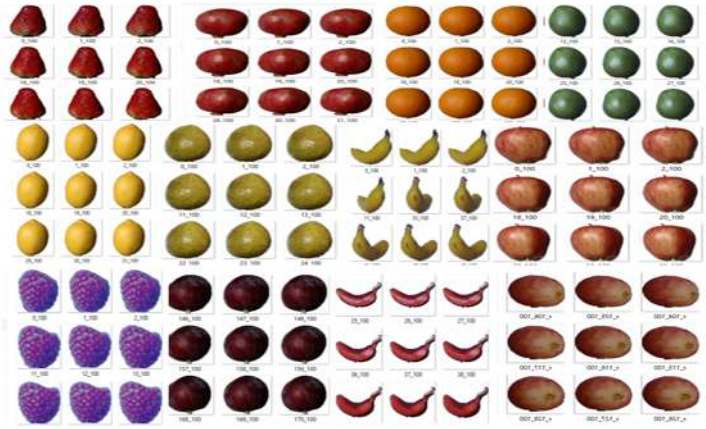


Figure 3: SampleFruit-360 image dataset

Table 2: Activation Functions in DNN

Activation Function	**Representation**	**Range of value**	**Weakness**
Sigmoid	$\sigma(z) = \frac{1}{1 + e^{-x}}$	0 to 1	Vanishing Gradient
Tanh	$\sigma(z) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$	-1 to 1	Vanishing Gradient
ReLU	$R(z) = \max(0, z)$	0 to ∞	Dying ReLU (used in hidden layers)
Leaky ReLU	$R(x) = \begin{cases} \alpha x & x < 0 \\ x & x > 0 \end{cases}$	0.01*x to ∞	Non - consistent
Softmax	$f(x_j^i) = \frac{\exp(z_j^i)}{\sum_x \exp(z_x^i)}$	0 to 1	Indicates the probability of certain classes (used in output layer)

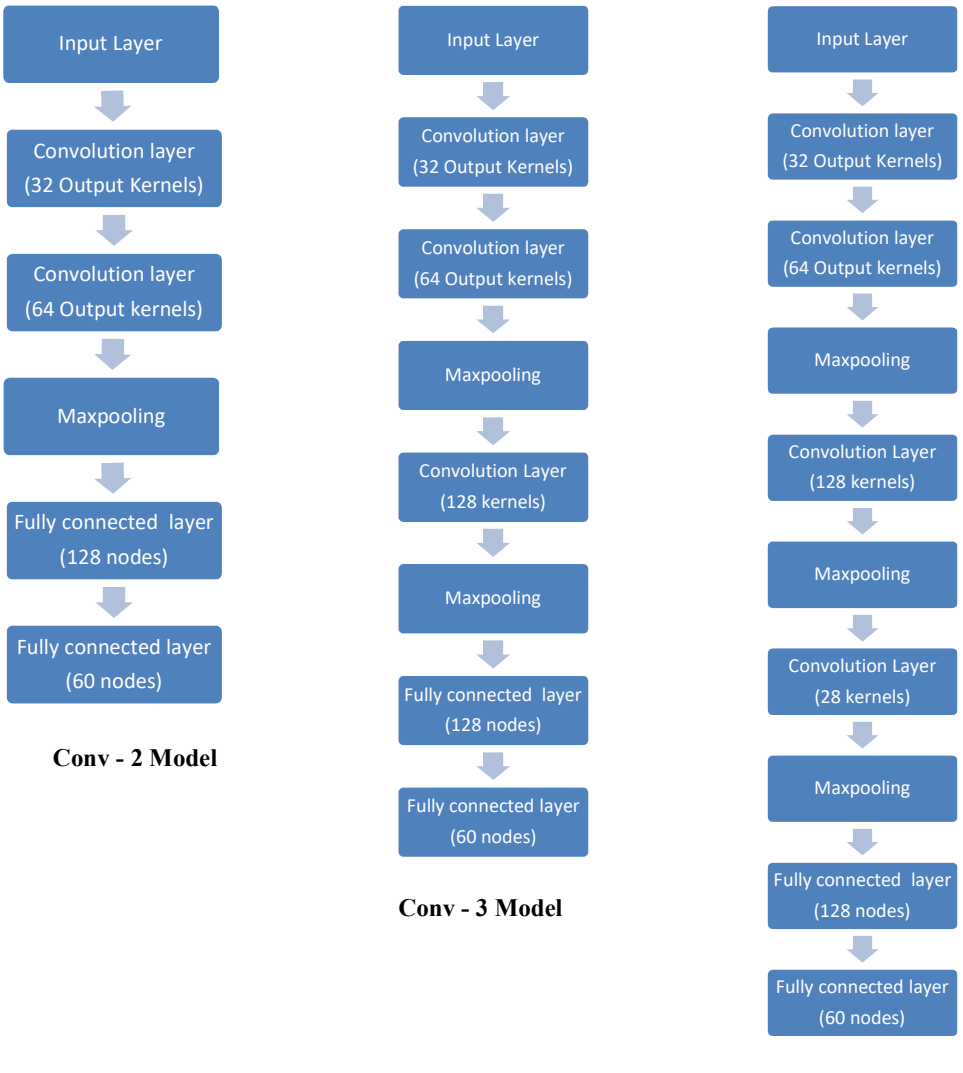


Figure 2: Baseline Architecture for fruit recognition system

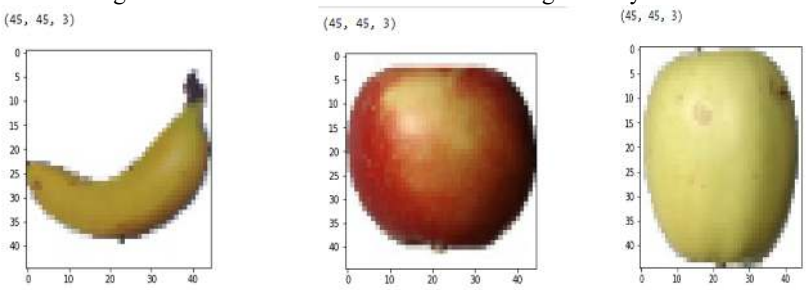


Figure 4: Pre-processed image

4.1 Normalization

To get rid of distortions caused by light and shadow in an image. The pixel of a coloured image takes a value in the range 0 to 255. In order to converge faster the colour range is normalized in the range 0 to 1. This is done by dividing each pixel value by the sum of pixel values over all the channels. For instance, let R, G and B be the pixel in the respective channels and S be the sum of pixel values across all the channels.

S= R+G+B

The normalized value for each pixel would be (R/S, G/S, B/S).

4.2 Label Encoding

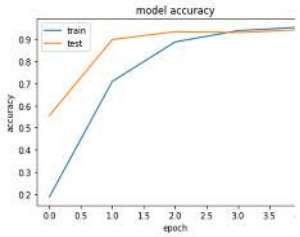
The categorical labels of the images are encoded to nominal values by using one-hot encoding. The pre-processed images are fed to the convolutional layer for capturing various features of the given image. The pragmatic evaluation of the proposed system is performed by varying the number of convolutional layers and changing the activation functions. The observation reveals the fact that increasing the number of convolutional layers improves the accuracy of the system by enabling it learns varied features. Table 4 shows the results of the fruit detection system by varying the convolutional layers and also testing with ReLU and Tanh activation functions. Increasing the kernel size has reflected in the improvement in accuracy of the recognition system. The accuracy and the loss of the recognition system using the ReLU and Tanh activation functions which are trained for 100 epochs is presented in Figures 5 and 6 respectively. Categorical cross - entropy loss function is used to adjust the parameters values in the network. The loss function is computed for an observation/example as

$$L = -\sum_{i=1}^{60} y_c \log(p_c)$$

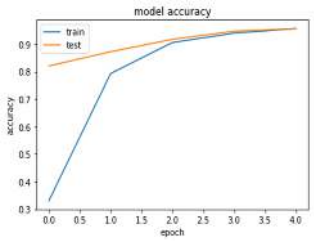
where P_c denotes the probability value computed for class c and denotes the whether the predicted label agrees with the actual label. It takes a value of 1 for the correct class and 0 for incorrect classes. The network is trained using stochastic gradient descent algorithms. The learning rate is adapted for each parameter separately using Adadelta. Both the activation functions perform equally well on the dataset. With the increase in the convolution layers, accuracy increases.

Table 4: Accuracy and Loss observed varying the activation functions and number of convolutional layers

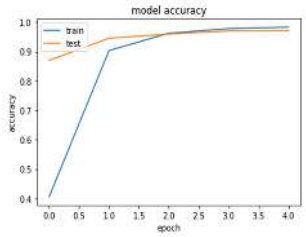
Number of layers	**Activation Function**	**Kernel size**	**Accuracy**	**Loss**
Conv- 2	ReLU	(2,2)	0.94	0.18
	Tanh		0.93	0.23
Conv -3	ReLU	(3,3)	0.95	0.13
	Tanh		0.95	0.14
Conv- 4	ReLU	(4,4)	0.97	0.13
	Tanh		0.97	0.09



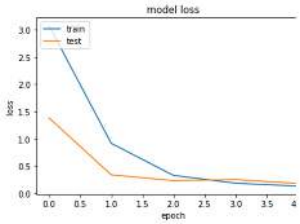
a) Model Accuracy - Convnet layers - 2



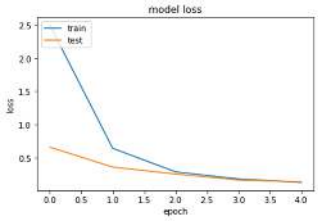
b) Model Accuracy - Convnet layers - 3



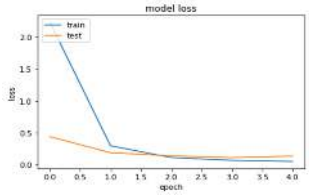
c) Model Accuracy - Convnet layers - 4



d) Model Loss -Convnet layers - 2

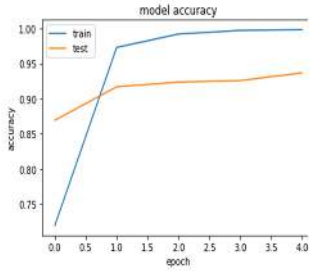


e) Model Loss -Convnet layers - 3

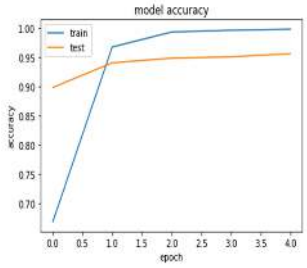


f) Model Loss -Convnet layers - 4

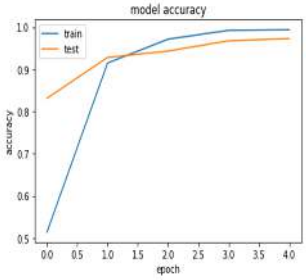
Figure 5: Accuracy and Loss of the recognition system – ReLU Activation Function



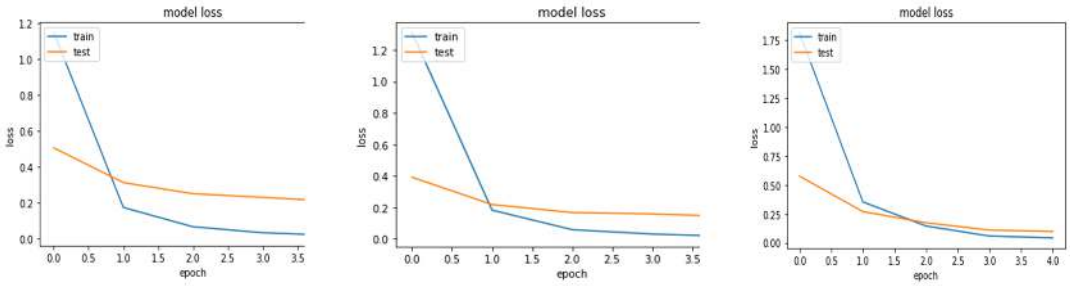
a) Model Accuracy - Convnet layers - 2



b) Model Accuracy – Convnet layers - 3



c) Model Accuracy - Convnet layers - 4



d) Model Loss - Convnet layers - 2

e) Model Loss - Convnet layers - 3

f) Model Loss -Convnet layers - 4

Figure 6 : Accuracy and Loss of the recognition system – Tanh Activation Function

4 Conclusion

This work presents state-of-the-art fruit detection system for recognizing various fruits under different lighting conditions, deformation, colour, size and shape. The feature extraction is performed by the convolutional layers in CNN, a deep learning algorithm which aids in the recognition of the images. This system is not restricted to fruit detection, it may find its application in any image recognition system. A reliable and accurate fruit detection system is built achieving a substantial accuracy of 97% and also a minimal loss of 0.13. Future work is to integrate the system with fruit harvesting robot. Further, the work shall be extended to implement the fruit detection system in a distributed environment to improve the response time of the system.

References

[1] Hetal N.Patel, Dr.R.K.Jain, Dr M.V. Joshi, "Fruit Detection using improved Multiple Features based algorithm", International Journal of Computer application(0975-8887),vol.13-No 2, Janurary 2011.

[2] Zawbaa H. M. , Hazman M., Abbass M. and Hassanien A. E. , "Automatic fruit classification using random forest algorithm", Hybrid Intelligent Systems (HIS) 4 th international conference, Kuwait, 14-16, pp: 164 – 168, 2014

[3] Ms. Snehal Mahajan , Prof. S. T. Patil "Optimization and Classification of Fruit using Machine Learning Algorithm". IJIRST –International Journal for Innovative Research in Science & Technology Vol.3- No 01, June 2016.

[4] Saswati Naskar, Tanmay Bhattacharya, Ph.D, "A Fruit Recognition Technique using Multiple Features and Artificial Neural Network", International Journal of Computer Applications (0975 – 8887) Volume 116 – No. 20, April 2015.

[5] TensorFlow. https://www.tensorflow.org. last visited on 14.01.2018

[6] Fruits Dataset on GitHub. https://github.com/Horea94/Fruit-Images-Dataset. last visited on 24.01.2018

[7] Rupesh K Srivastava, Klaus Gre_, J urgen Schmidhuber. "Training very deep networks, Advances in neural information processing systems", Twenty-Eight International Conference

on Neural Information Processing Systems, pp. 2377-2385, Montreal, Canada, December 07-12, 2015.

[8] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, Martin A.Riedmiller, “Striving for Simplicity: The All Convolutional Net”, CoRR, abs/1412.6806, 2014.

[9] Greenwald HS and Oertel CK, “Future Directions in Machine Learning”,Front. Robot. AI, 3:79, 2017. doi: 10.3389/frobt.2016.00079.

[10] Jürgen Schmidhuber, “Deep learning in neural networks: An overview”, Neural Networks, vol. 61, pp: 85-117, 2015.

[11] Weibo Liua, Zidong Wanga,∗, Xiaohui Liua et al., “A survey of deep neural network architectures and their applications”, Neurocomputing, vol. 234, pp: 11-26, 2017.

[12] Guo Feng; Cao Qixin and Nagata Masateru, “Fruit Detachment and Classification Method for Strawberry Harvesting Robot”, International Journal of Advanced Robotic Systems, Vol. 5, No. 1, pp: 41-48, 2008.

[13] S. Md. Iqbal, A. Gopal, P.E. Sankaranarayanan & Athira B. Nair, “Classification of Selected Citrus Fruits Based on Color Using Machine Vision System”, International Journal of Food Properties, 19:2, 272-288, 2016. DOI: 10.1080/10942912.2015.1020439

[14] Gennady Ososkov1,_ and Pavel Goncharov2, “Two-Stage Approach to Image Classification by Deep Neural Networks”, In EPJ Web of Conferences 2018. https://doi.org/10.1051/epjconf/201817301009

[15] M. Tholkapiyan, A.Mohan, Vijayan.D.S , “A survey of recent studieson chlorophyll variation in Indian coastal waters”, IOP Conf. Series: Materials Science and Engineering 993 (2020) 012041, doi:10.1088/1757-899X/993/1/012041.