

# Date Fruit Recognition using Feature Extraction Techniques and Deep Convolutional Neural Network

Aurangzeb Magsi\*, Javed Ahmed Mahar and Shahid Hussain Danwar

Department of Computer Science, Shah Abdul Latif University, Khairpur, Pakistan;  
aurangzeb.magsi.sef@gmail.com, mahar.javed@gmail.com, Shahid.danwar@salu.edu.pk

## Abstract

Numerous biotechnology software applications are developed to provide computational solutions to complex agricultural problems like identification of diseases and monitoring plant growth. Dates are healthy fruit and its contribution in total G.D.P of Pakistan is approximately 4% in which District Khairpur provides approximately 81% production. Approximately 22 types of Dates are produced in different areas of Pakistan. It is observed that national as well as international empor are unable to correctly identify the type of dates. **Objectives:** This study aims to presents a framework for recognition of Dates using Deep Learning technique based on color, shape and size feature extraction methods. **Methods:** We have established fruit images dataset of 500 images for evaluation purpose likewise 360-dataset. Three types of Dates were selected for experiments like Aseel, Karbalain and Kupro. The range of 500date fruit samples were collected out of which 350 used for training dataset and 150 used for testing purpose. **Findings:** Experiment performed on the selected samples following the proposed framework. For better accuracy, we have used combination of several hidden layers and 100 epochs which gives the best performance result of 97.2% at 4th epoch. A confusion matrix is used to analyze and measure the results accuracy through which we get 89.2% as a True Positive. **Application and Improvements:** The outcome will be beneficial for the empor, researchers and also for automated factory classification.

**Keywords:** Biotechnology Applications, Color Feature Extraction, Confusion Table, Dates

## 1. Introduction

Date fruits are sweet and edible, consumed as a staple diet. Date is practically at no cost have sodium and cholesterol elements providentially. The dates are very useful, helpful and they are well thought-out a divine gift specially for heart patients. Dates are one of the major crops of Pakistan and possess an enormous significance in global market. There are more than 400 varieties of date fruit found around the world among which 40 types of dates

provide extensive taste, shape and colors having special prices as per their importance<sup>1</sup>.

Various advanced computational techniques are used to recognize and identify the types and quality of different fruits using number of different methods. The machine vision is used on multispectral imaging, monochrome imaging and color imaging for grading and sorting<sup>2</sup>. Some applications for uncovering the disease, defective region and contamination on vegetables and fruits are discussed in<sup>3</sup>. Deep Neural Network (DNN) are very efficient in

\*Author for correspondence

order to identify and classify fruit images. They provide better accuracy results than other machine learning algorithms. The most common example of the Artificial Neural Network is Convolutional Neural Networks and is considered as the most efficient Deep Learning Algorithms<sup>4</sup>. Furthermore, many other algorithms are also proposed to extract the feature of fruit characters by using its image for identification and classification. Fruit can be detected and graded according to its quality by using the color, shape and size features and such process can be done by using various classifiers. To identify the class of the fruit, morphological feature and color feature were used with neural network<sup>4,5</sup>.

The lack of skills to identify different types of dates in to consumers make it a matter of importance to develop such an application which should be prove helpful and use in marketplace for such purpose. The Dates recognition framework using DNN based on features extraction methods is proposed and implement in this study.

The rest of the study classifies in different sections; related work with the research contributions is described in Section 2. In Section 3, the collected data is represented and the flow of the developed system is given in Section 4. The calculated results and concluded comments are presented in Section 5 and 6 respectively.

## 2. Literature Survey

Lot of work is done regarding the classification and recognition of fruits including date fruit using different statistical and artificial intelligence techniques. In<sup>6</sup> described the method of fuzzy inference. In<sup>7</sup> proposed a complete mechanism based on Gaussian mixture for Date classification. The taken samples of Date pertaining to every module were further checked by Mardia's multivariate tests. In<sup>8</sup> introduced a fruit recognition system based on convolutional neural network with experiments performed on the selected samples following the proposed framework array of high quality images dataset of 38409 samples of 60 different fruits. They trained a neural network based model with 40000 iterations to batches of 50 images for fruit detection and

achieved accuracy rate of 96.3% during testing phase. A R-CNN based framework proposed for orchard based on state-of-the-art detection technique<sup>9</sup>. Three different fruits were selected for individual study i.e. Apple, Mango and Almond. With the large number of training images flip and scale data augmentation techniques were found best for desired results. The experimental results produced accuracy in detection with an F1-score of >0.9 for apples and mangoes.

Likewise, A method introduced by and others for generating large-scale data set of synthetic images that are semantic<sup>10</sup>. The main purpose behind this generation of segmentation data set in agriculture was to bootstrap or pre-train models of image identification based on computer vision. It presented an approach on adaptive method for detection of area of interest from fruit images<sup>11</sup>. The co-working solution was applied on novel neural architecture and heuristic search was applied on three types of different fruits i.e. Apple, Banana and Orange. They performed the Segmentation process on the input images for easier process then heuristic search was performed based on Adaptive Artificial Neural Network in order to detect the area of interest. Similarly, presented a method of automatic fruit recognition and counting from multiple images<sup>12</sup>. They use bag-of-words model to find fruit image and novel statistical approach for counting. Correlation of 74.2% was achieved with experiments on 28000 color images.

Another automatic fruit segmentation and classification approach was presented by<sup>13</sup>. Extracted most relevant features from image by combining feature learning algorithm with conditional random field and achieved 88% accuracy in order of classification. Furthermore, also proposed a method of fruit recognition based on compressed sensing technique to reduce the complex degree of algorithm<sup>14</sup>. Different class fruits were experimented and they achieved recognition rate of 83%. Also, a fruit harvesting autonomous robot was designed by<sup>15</sup>. In their work advanced method of neural network and stereo matching technique was applied for fruit image identification. Likewise,<sup>16,17</sup> developed an innovative approach of classification on the basis of image processing

and machine learning algorithms using various classifiers like SVM, KNN and Nave Bayes. Similarly, presented an approach of fruit detection based on Deep convolutional Network<sup>18</sup>. Fruit detection was performed by adapting the faster region-based CNN model with the combination of color RGB image and Near-Infrared (NIR).

After reviewing the literature, it is observed that majority of the researchers prefer color, shape and size based techniques for the classification and recognition of date and other fruits. Fadel<sup>19</sup> developed a classifier for the separation of date fruit on the RGB color criteria using probabilistic neural network. Similarly, <sup>20</sup> presented a system using RGB images which classify dates in to three categories by extracting their qualities. Likewise,<sup>21</sup> introduced classification method of computer vision and pattern recognition, which was tested on an image containing seven different types of dates and extracting fifteen features from image. The classification system was proposed by Muhammad<sup>1</sup> extract features like texture, size and colors from image of dates then decomposed it into components and applied weber local descriptor

to each component for classification. A novel method presented by <sup>22</sup> for Date fruit grading, used KNN classifier with the combination of shape and texture feature by using contour of the date fruit, curvelet transform and local binary pattern respectively.

The research we have done is also based on three features i.e. Color, Shape and Size however some of the techniques that are used in this research are taken from the above mentioned literature. The selected specimen of date fruit in this research are completely different from all the other types that were used in previous research due to the changing of weather and environmental conditions.

### 3. Data Collection

Authors of this research are basically belongs to the district Khairpur which is the major date growing city of Pakistan. The season of the dates expired in the month of august 2018. Moreover, different types of Dates grow in district Khairpur. Initially, three types of dates i.e. Aseel, Karbalain and Kupro are selected for experiments. The



**Figure 1.** Sample image of aseel date.

**Table 1.** Statistical information of collected date images

Date Types	Captured images	Selected Images
Aseel	200	185
Karbalain	200	168
Kupro	200	147
<b>Total</b>	<b>600</b>	<b>500</b>

sample image of Aseel is shown in Figure 1. The samples of the mentioned date types were collected at the end of July 2018.

For the collection of data, a digital single-lens reflex camera device was used at the 120 mega pixel standard in the clear blue sky day with 120,000 lux luminance. The captured date images then saved in to the computer memory. Images were cropped and converted in to 300 dpi by using Adobe Photoshop CS6 software.

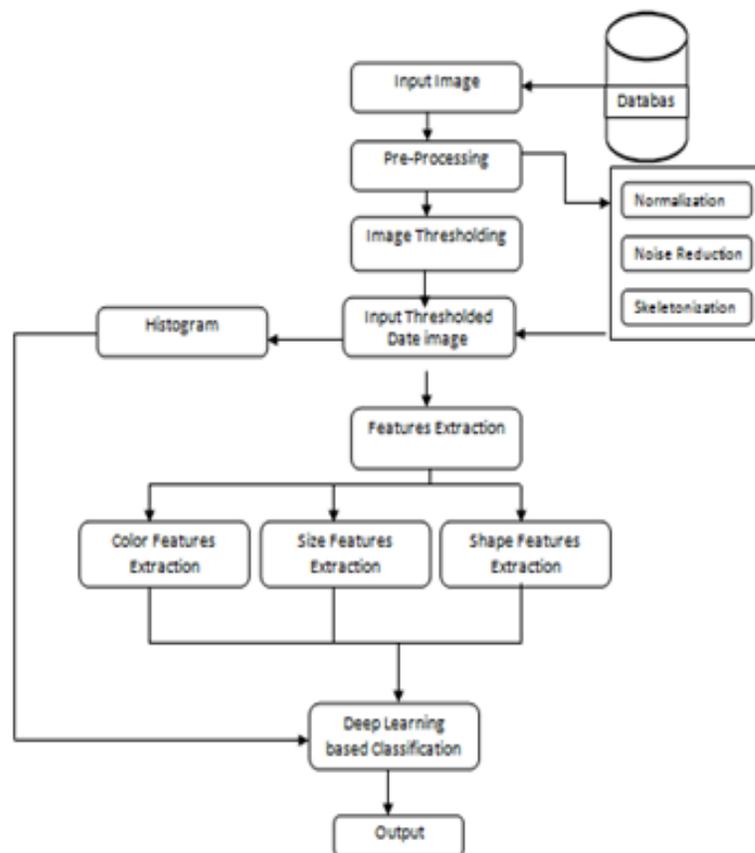
The captured images of dates are 600 among them 200 are of Aseel date, 200 are of Karbalain and remaining 200 images are the types of date Kupro. High maturity level of Date fruit is required for experimentation because the color of date fruit varies day by day and the shape structure of captured date should be at vertical level. With the help of leading farmers, 200 samples for each type selected for

process and experiment. The statistical information of the collected date images is given in Table 1.

## 4. Overview of Dates Recognition System

Various mechanisms are proposed for the identification and classification of fruits and vegetables. In this research project, date fruit is selected because dates are core fruit of district Khairpur. Approximately 22 types of dates produced from which some types are grown in fewer amount while some on the basis of gustatory perception are grown in anenormous amount.

From the most grown types of date fruit, three well known types i.e. Aseel, Karbalain and Kupro were selected for experimentation which are most liked and favorable. Experiments for the identification of different types of



**Figure 2.** Flowchart of proposed dates recognition system.

date, performed using Matlab R2015a software. The basic operations basically performed on the methodologies of Threshold, Histogram and pixel information. The flowchart of the proposed methodology is depicted on “Figure 2”.

A dataset for training and testing purpose created in order to save and process captured images. The next step is the preprocessing where normalization technique is applied to all the images in order to make them normalize in size and shape protocols. Then we apply the Noise reduction operation to all the collected date images to remove the very small scale features. The third part of the preprocessing is the Skeletonization which extracted the region-based shape of date fruit from image.

To remove the background and separate only the shape of date, the Image Threshold technique is used which removed the entire background of the image (except date fruit itself) successfully. The threshold image then used as an input to histogram. Using Histogram technique, graph of RED, GREEN and BLUE component of threshold image generated. Furthermore, to extract the color pixel values of the preprocessed image, color feature extraction method was used. Area, perimeter, length, width and centroid of the date image calculated in order

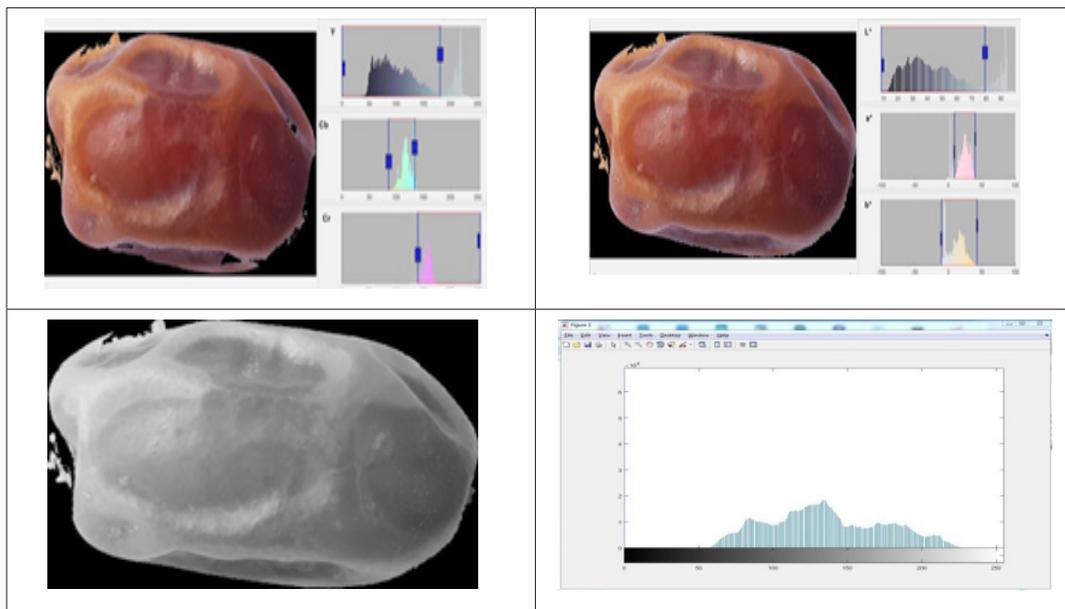
to extract its size features. While the shape features of the respective image extracted through the Scale-Invariant Feature Transform (SIFT) method.

By using the Histogram and extracted feature’s values, classification of Date fruit took place through Deep Learning method. The trained Neural Network is able to recognize and classify the specific Date type which has similarities in term of graph value, pixel information, and similarities in shape key points, length, width, color percentage to the targeted values.

## 5. Experiments and Results

The collected images of date fruit are experimented through the proposed mechanism. The 200image samples were collected from the each selected date type for experimentation. The confusion matrix is also interestingly used by the researchers for the better accuracy in classification of fruits<sup>23</sup>. Hence, the outcome of the developed application is represented using the confusion table for identification of proper date fruit.

The actual process begins with the background removing by image Thresholding technique. For Image enhancement we need a change in color space of the



**Figure 3.** Experimented image of YCbCr and La\*b\* color space with gray scaled image and its histogram.

selected image to obtain single channel for luminance information and two other channels for chrominance information. Therefore various color spaces are applied on the original color image in order to obtain specific color space.

From the various applied color space operations, the output of YCbCr and La\*b\* results are shown in Figure 3. By using RGB technique, image loses its shadow. This means it leads to black crush as it turns dark area in to black. So in this case information of that area also loses.

YCbCr threshold technique is used for experiment which split image in to luminance Y and Chrominance CB and Chrominance Cr. The complex background and different illumination conditions are removed by segmenting the date fruit in this technique. Effects on date region segmentation are avoided as Chromaticity of YCbCr color space is being used.

La\*b\* technique which describes all perceivable colors in three dimensions *L* for lightness while *a* and *b* green-red and blue-yellow color opponent mathematically. It consist of three layers luminosity ‘L\*’, chromaticity

‘a\*’ where colors falls along the axis and chromaticity ‘b\*’ which indicate the color falls along the blue-yellow axis. Among three defined techniques, La\*b\* is the most suitable which was selected for further processing.

After removing the background, it can clearly be visualized that some noise like holes appeared in image region and also some of major boundary contents removed. This is because of their similarities with the background pixels values. To overcome this complicatedness, image is converted into Grayscale level as shown in Figure 3 in order to recover the lost data. The Histogram technique is applied to the gray scaled image to find out the graphs for further process.

Luminance of the image shows in horizontal (X) axis in the histogram from left to right edge of the graph as from pure black to pure white. While comparative quantity of light for the luminance is show on the vertical (Y) axis in the graph.

The color histogram technique is also applied on with and without background images. This technique basically used to apply for global contrast of images but in this

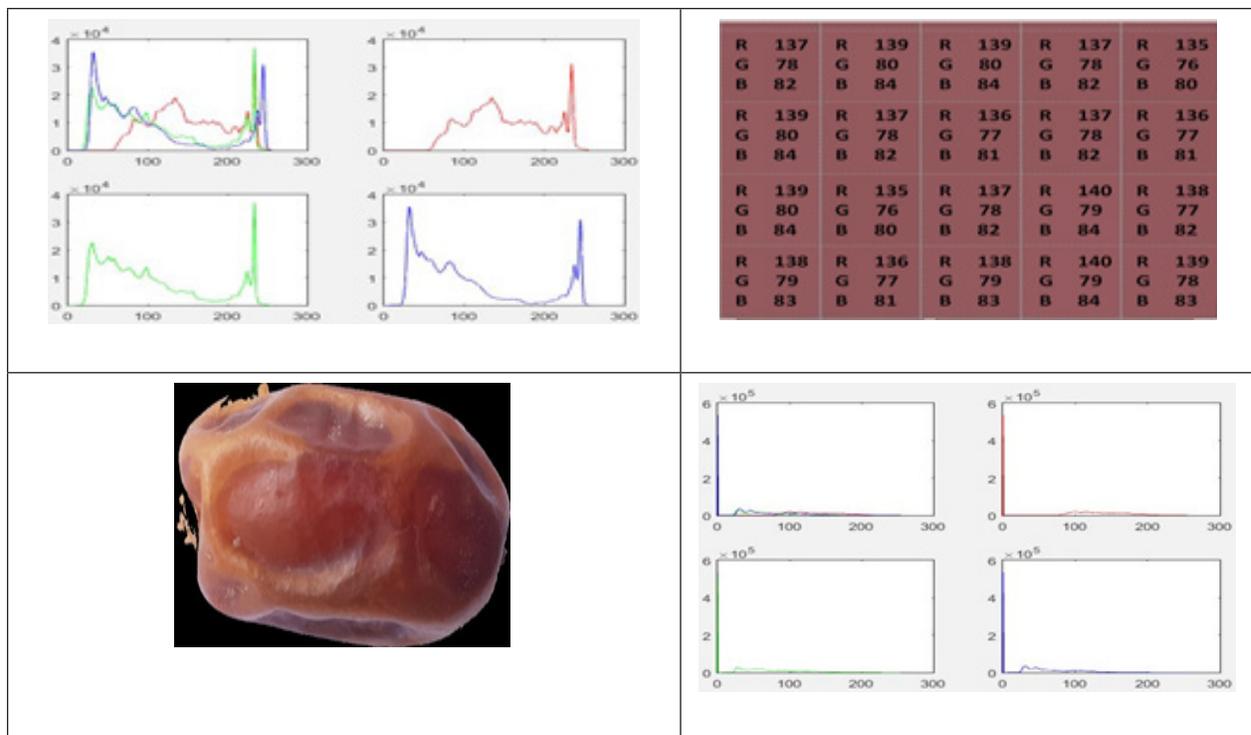


Figure 4. Color histogram and pixel information.

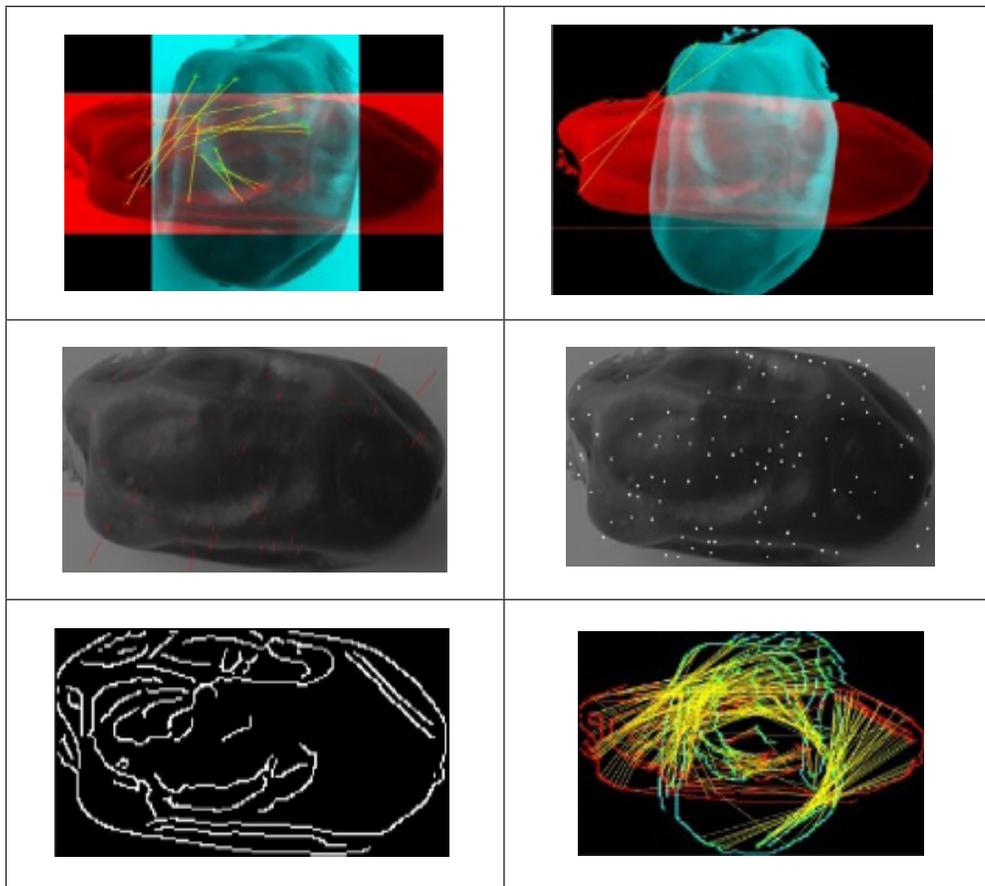
research histogram technique is applied to extract the pixel information in term of graph.

When the color histogram technique was applied, the output we get is in a graphical form which shows the ratio of the RED, GREEN and BLUE component of the image individually and combined as well. Along with the graph, percentage is also acquired for RED, GREEN and BLUE pixels of the image without background. To analyze the texture of the input image, we need pixel information and Matlab facilitates with the built-in tool to provide the equivalent pixel matrix of the image as shown in Figure 3.

In the Figure 4, first quarter show RED, GREEN and BLUE components collectively. But for recognition of more than one date, graphical representation of all the three components individually i.e. RED, GREEN and BLUE along with their values is required for comparison and classification. So the color histogram of RED, GREEN

and BLUE-sensitive pixels shows their values respectively. The mentioned histogram technique is applied to the threshold image of date fruit as well which are without background. And likewise the RED, GREEN and BLUE components of threshold image as a color histogram and pixel information is also acquired.

The intensity level is measured by extracting pixel information of both images with and without background. The concluded results show that the image with background has higher intensity level as compared to the image without background. The pixel information within the boundary of the date fruit image does not show any difference in the both images but the pixels that are out of the fruit region contain different values which hardly effects the next processes. To attain the accurate results we need image with low intensity therefore the image without background is selected for further practice.



**Figure 5.** SIFT and Harris feature extracted techniques and key point localization.

To extract color feature, an array of M by N by 3 is used to store color image data which define Red, Green and Blue color components for each individual pixel. Combination of R,G,B intensities defines color of each pixel and stored in each color plane at pixel's location. Values of each color component vary between 0 and 1 in an RGB array. A pixel has value (0,0,0) displayed as black and pixel with value (1,1,1) displayed as white.

Three color components R,G,B for any pixel stored along the third dimension of the data array. For example, RED, GREE and BLUE color components of the pixel (10,2)are stored in RGB(10,2,1), RGB(10,2,2) and RGB(10,2,3) respectively.

Along with the color features, also shape feature extracted in order to acquire meaningful results. Using various operations of SIFT (scale invariant feature transformation algorithm), different values for shape feature were generated. Using Harris feature detector, similar features were extracted and matched in original and gray scaled images as shown in Figure 5. SIFT feature generator eliminates the edge responses which leads to point out the major and minor edges in an image and allow to localize the number of accurate key points for shape identification. The edge responses and key point

localization also depicted in Figure 5. For better results canny edge detector is used along with the harris feature detector in order to extract the maximum number of similar shape features. Using canny edge detector, number of shape features extracted as it detect number of internal edges in an image along with the boundary edges which allow Harris feature detector to match maximum number of features as shown in Figure 5.

Extracting the region properties of an image, information of area, perimeter, major axis, minor axis, centroid and bounding box is used to find out the size of an image. The perimeter of an image is measured in the number of pixels. First the connected components in a gray scaled or binary image labeled to be identified then using region props() function, multiple regions were extracted and structured in an array. To get the outlines of an array of (x,y) coordinates of individual i.e. horizontal, vertical and diagonal pixels, bw boundries() function along with plot() is used to plot all the possible pixel values over the original image and also total number of pixels count from the binary image as shown in Figure 6. The properties of region of an image such as Area, Perimeter, Bounding box, Length, Width, Centroid and total number of pixels inside the boundary were calculated as size feature and

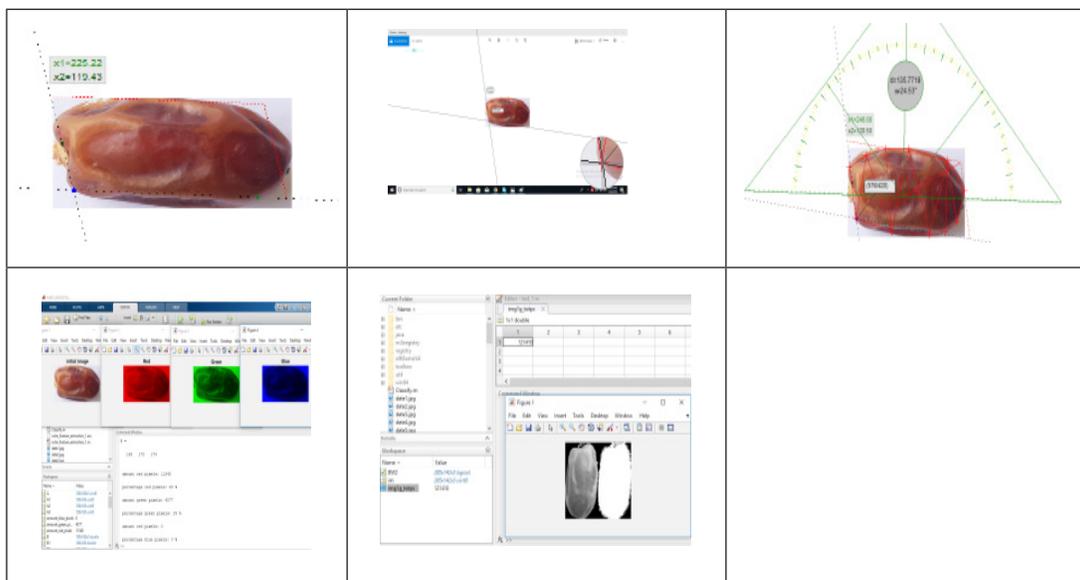


Figure 6. Extraction of length, width in pixel form and percentage of color components.

merged with other two features such as Color and Shape feature for better classification and acquire high accuracy in results.

The matrix of the true color image, the total number of key points and the calculated values of area, perimeter, centroid and values of other extracted features were stored in a variable form which is used as an input to train the neural network for classification. The developed neural network is trained by defining the target class or target values with respect to the input class of dataset through proper feature matrix.

The training data set contain 350 image values while the testing data set contain values of 150 images. Images used for feature extraction were in a fixed resolution of

150x150x3(Length\*Width\*Chanel). Images information i.e. values of various extracted features were fed in to the CNN as Input. The developed network was trained with the Gradient algorithm.

$$w = w - \alpha \nabla_w J(w)$$

Which measures positive and negative gradient at each epoch during the iteration process w.r.t parameters at the learning rate of 0.001? The network was trained for 100 epoch through the training set and generated the best training performance of 3.3254e-26 at 4th epoch as shown in Figure 7.



Figure 7. Best performance result at 4<sup>th</sup> epoch.

Table 2. Calculated results of DRS

Dates Types	True Positive	False Positive	False Negative	True Negative
Aseel	126	0	22	10
Karbalain	93	6	5	0
Kupro	75	8	5	0

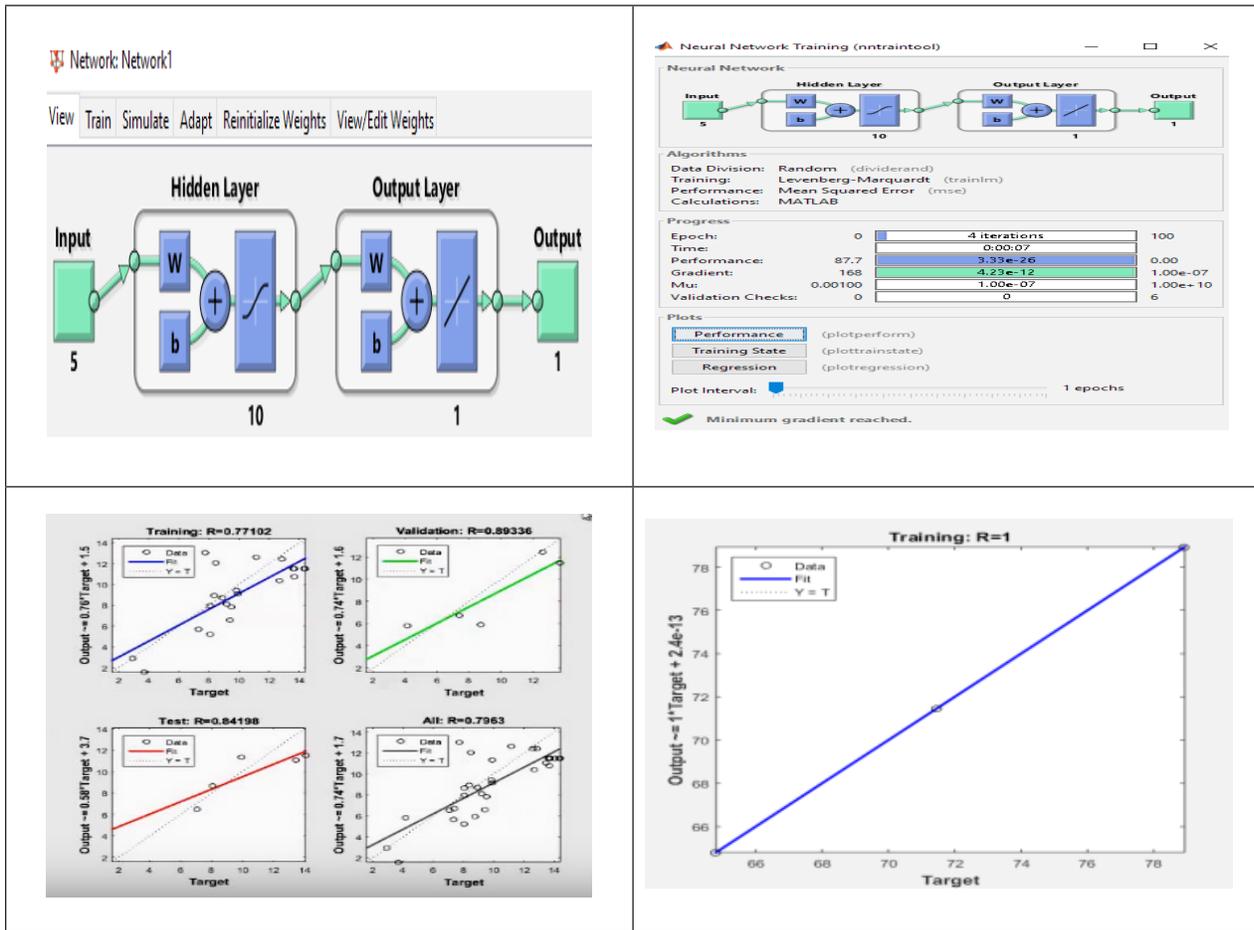


Figure 8. Execution process of neural network.

In this experiment we have trained a network based on 5 neurons as an input, a fully connected hidden layer consist of 10 neurons and an output layer with the minimum batch size of 50. Simulation is performed in MATLAB software and acquired 97.2% accuracy in identification with certain inputs and targeted values as depicted in Figure 8. Acquired results are analyzed through confusion table.

As the image based results are always displayed in a matrix form or table form. Hence, in this research, confusion table is preferred to represent the data because the developed system sometimes efficiently identifies images and gives positive result and negative output produces by system seldom as very few times the system shows true negative and false positive results which

are unexpected results. The detailed calculated results represented in Table 2.

The selected three types of dates collectively observed, experimented and achieved successful identification results in terms of true positive with Karbalain date while unexpected result such as true negative also received with the Aseel date. Surprisingly, the results True Positive of Aseel and Kupro dates are quite acceptable. The successful false negative and unpredicted false positive also received with images of Aseel and Kupro.

## 6. Conclusion

It is demonstrated that the proposed framework is very effective in the classification of dates based on color, size

and shape features extraction from the images of dates. Approximately 22 types of dates are grown in province Sindh; only three types of dates i.e., Aseel, Karbalain and Kupro were selected for recognition and classification. 500 samples were collected among them 350 required samples were experimented. The experiments were performed in Matlab using the proposed framework. Before the recognition, three preprocessing operations were performed on each input image. Different color based techniques; shape and size algorithms were used during the entire process of feature extraction and recognition and concluded results at acceptable level by evaluating graph and extracting pixel information. The calculated results represented with confusion table. The cumulative best accuracy of 97.2% is achieved during the experiments. For future work; number of features like texture and more and also by types of date fruit will be increased to develop more efficient and powerful application of recognition and classification.

## 7. References

- Muhammad G. Date fruits classification using texture descriptors and shape-size features. *Engineering Applications of Artificial Intelligence*. 2015; 37:361–7. <https://doi.org/10.1016/j.engappai.2014.10.001>.
- Naik S, Patel B. Machine vision based fruit classification and grading—A review. *International Journal of Computer Applications*. 2017; 170(9):22–34. <https://doi.org/10.5120/ijca2017914937>.
- Mahendran R, Jayashree GC, Alagusundaram K. Application of computer vision technique on sorting and grading of fruits and vegetables. *Journal of Food Processing and Technology*. 2012:1–7.
- Moallem P, Serajoddin A, Pourghassem H. Computer vision-based apple grading for golden delicious apples based on surface features. *Information Processing in Agriculture*. 2017; 4(1):33–40. <https://doi.org/10.1016/j.inpa.2016.10.003>.
- Ninawe P, Pandey MS. A completion on fruit recognition system using K-Nearest Neighbors Algorithm. *International Journal of Advanced Research in Computer Engineering and Technology*. 2014; 3(7):2352–6.
- Alavi N. Date grading using rule-based fuzzy inference system. *Journal of Agricultural Technology*. 2012; 8(4):1243–54.
- Aiadi O, Kherfi ML. A new method for automatic date fruit classification. *International Journal of Computational Vision and Robotics*. 2014; 7(6):692–711. <https://doi.org/10.1504/IJCVR.2017.10006506>.
- Mureşan H, Oltean M. Fruit recognition from images using deep learning. *Acta Universitatis Sapientiae, Informatica*. 2018; 10(1):26–42. <https://doi.org/10.2478/ausi-2018-0002>.
- Bargoti S, Underwood J. Deep fruit detection in orchards. *IEEE International Conference on Robotics and Automation (ICRA)*; 2017. p. 3626–333. <https://doi.org/10.1109/ICRA.2017.7989417>.
- Barth R, Jsselmuiden JI, Hemming J, VanHenten EJ. Data synthesis methods for semantic segmentation in agriculture: A Capsicum annum dataset. *Computers and Electronics in Agriculture*. 2018; 144:284–96. <https://doi.org/10.1016/j.compag.2017.12.001>.
- Woźniak M, Połap D. Adaptive neuro-heuristic hybrid model for fruit peel defects detection. *Neural Networks*. 2018; 98:16–33. <https://doi.org/10.1016/j.neunet.2017.10.009>. PMID:29153957.
- Sang Y, Glasbey CA. Automatic fruit recognition and counting from multiple images. *Biosystems Engineering*. 2014; 118:203–15. <https://doi.org/10.1016/j.biosystemseng.2013.12.008>.
- Hung C, Nieto J, Taylor Z, Underwood J, Sukkarieh S. Orchard fruit segmentation using multi-spectral feature learning. *IEEE/RSJ International Conference on Intelligent Robots and Systems*; 2013. p. 5314–20. <https://doi.org/10.1109/IROS.2013.6697125>.
- Gao G, Zhao S, Zfang C. Study on fruit recognition methods based on compressed sensing. *Journal of Computational and Theoretical Nanoscience*. 2015; 12:2937–42. <https://doi.org/10.1166/jctn.2015.4203>.
- Henry AM, Williams MN. Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. *Biosystems Engineering*. 2019; 181:140–56. <https://doi.org/10.1016/j.biosystemseng.2019.03.007>.
- Mahar JA, Khalid M, Mahar MH. Off-line signature verification of bank cheque having different background colors. *Proceedings of IEEE AICCSA*; 2007. p. 738–45. <https://doi.org/10.1109/AICCSA.2007.370715>.
- Pour AS, Chegini G, Massah J. Curvature-based pattern recognition for cultivar classification of Anthurium flowers. *Postharvest Biology and Technology*. 2018; 139:67–74. <https://doi.org/10.1016/j.postharvbio.2018.01.013>.
- Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C. Deep fruits: A fruit detection system using deep neural networks. *Sensors*. 2016; 16(8):1–10. <https://doi.org/10.3390/s16081222>. PMID:27527168. PMCID:PMC5017387.

19. Fadel M. Date fruits classification using probabilistic neural networks. *Agricultural Engineering International: CIGR Journal*. 2007:1–9.
20. Ohali YA. Computer vision based date fruit grading system: Design and implementation. *Journal of King Saud University-Computer and Information Sciences*. 2011; 23(1):29–36. <https://doi.org/10.1016/j.jksuci.2010.03.003>.
21. Haidar A, Dong H, Mavridis N. Image-based date fruit classification. *IV International Congress on Ultra-Modern Telecommunications and Control Systems*; 2012. p. 357–63. <https://doi.org/10.1109/ICUMT.2012.6459693>.
22. Mohana SH, Prabhakar CJ. A novel technique for grading of dates using shape and texture features. *Machine Learning and Applications*. 2014; 1(2):15–29. <https://doi.org/10.5121/mlaj.2014.1202>.
23. Zang Y, Wu L. Classification of fruits using computer vision and a multiclass support vector machine. *Sensors*. 2012; 12:12489–505. <https://doi.org/10.3390/s120912489>. PMID:23112727. PMCID:PMC3478854.