JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

SICAIET
2024

Research Article
Collection: Artificial Intelligence and Emerging Technologies

Classification for Phoenix Dactylifera L. Varieties Using Statistical Features

Hafiz Muhammad Ijaz[1], Tanveer Aslam[2*], Syed Ali Nawaz[2], Muhammad Shehzad[3], Muhammad Sabir[1], Syed Zohair Quain Haider[1], Shehzad khan[1], and Muhammad Yasir Khan[3]

[1]Department of Computer Science & IT, Institute of Southern Punjab Multan (ISP-Multan), Multan, Pakistan.

[2]Department of Information Technology, Islamia University of Bahawalpur, Bahawalpur Punjab, Pakistan.

[3]Institute of Computing, MNS University of Agriculture, Multan, Pakistan.

[*]Corresponding Author: Tanveer Aslam. Email: tanveerchuhan786@gmail.com

Academic Editor: Salman Qadri **Published:** February 01, 2024

Abstract: Food plays a crucial role in human life since it completes the human body's vital nutrients, vitamins, minerals, and antioxidants. Phoenix Dactylifera L is one of several food items that are grown in the farming sector. Phoenix dactylifera L., is also known as "date fruit" was a blooming species of plant in the palm family. The main objective of this experimental process is to identify a technique for classifying the categories of date fruits using Machine Vision (MV) methodology. There are different categories of date fruit, and every category has its own importance. In this study, we were collected six different varieties of date fruit namely: Fard, Gulistan, Hussaini, Mozawati, Shakar, and Chohara. These data sets were captured with a digital camera on a bright day without any necessary tools or a lab. A total of 43 statistical features were examined for each date fruit's image in the interest of region (IOR) using three different statistical features, namely: Binary, Histogram, and Texture. The 10-K fold cross-validation method was applied to several classifiers. The instance-based K-Nearest Neighbor (IB-KNN) classifier achieved the best overall accuracy result (OAR) of 97%.

Keywords: Date Fruit Data; Feature Optimizing; Classification Process; Machine Vision.

1. Introduction

Fruits are the greatest source of antioxidants, fact-checked nutrients, vitamins and minerals [1]. Farming and agriculture are the largest areas where different categories of food items are cultivated around the world [2]. Due to the growing population, agriculture plays an extremely beneficial role in the provision of food. In Pakistan, the farming and agriculture sectors were the most important pillars of economic development [3]. Pakistan is a developing country where most of the population is directly or indirectly dependent on agriculture and farming [4]. In Pakistan, 67.2% of women and 30.2% of men work in the agricultural sector, which plays an important role in the country's development [5]. In the agricultural sector of Pakistan, different vegetable and fruit items are one of the most important sectors for fulfilling food requirements [6]. Many fruits are cultivated in Asian countries, and Phoenix dactylifera L. is one of them [7]. Phoenix dactylifera L. is also known as "date fruit". Date fruits are the most important food item in the world. In Pakistan, a number of varieties of date-fruit orchard farms are cultivated, especially during the date-fruit season. Furthermore, the area devoted to date fruit orchards has increased by 3.1 percent, while Pakistan's GDP has increased by 0.6 percent in the Financial-Year-Survey 2021 (FYS-2021) [5]. In Pakistan, numerous varieties of date fruits are grown in Sindh, Punjab, Balochistan, and Khyber Pakhtunkhwa (KPK), among other areas [8].

Pakistan is a cultivated and fertile land with different categories of plants, seeds, fruits, and crops. The anticipated yearly production of date fruit in Pakistan is 540,000 to 640,000 metric tons, with the following distribution: 57.3% in Sindh, 33.4% in Balochistan, 7.0% in Punjab, and 2.3% in KPK [9]. The most energizing and nutrient-dense food is the date fruit, which also contains lipids, proteins, and other nutrients, including vitamins and minerals. To meet the increasing demand for fruits from the public, a sufficient quantity of fruits should be made available in the fruit market. For research studies on food processing, it is crucial to identify the many types of date fruits [10]. Different varieties of date-fruit, including Aseel, Basra, Chohara, Fard, Gulistan, Hussaini, Kanri Wala, Mitha, Mozawati, Pimaza, Shakar, Sherin, Zard, and others, are farmed in Pakistan [11]. Pakistan is the biggest exporter around the world of food items and generates millions of dollars as revenue, according to the report (FYS-2021) [5]. Different varieties of date fruits were identified using domain expertise and field research methods. These procedures are expensive, fruitless, and time-consuming. The date fruit has unique properties in terms of area and color, compared to other edible fruits. The average diameter and length of date fruits are 2.1–6.2 cm (in height) and 1.3–4.6 cm (in diameter), respectively [12]. Due to the variety of date fruits, it can be challenging to determine the correct type by looking at it with open eyes [13].

Researchers have been working on different groups of date fruits based on their production and demand. This research study was designed to create a progressive automated system that can distinguish between different date-fruit types based on their plotted areas of characteristics. Date fruits come in a large number of types, but the six varieties that have been chosen are the most consumed and popular [9]. In this study, it was suggested that the following six date fruit varieties be identified: Fard, Gulistan, Hussaini, Mozawati, Shakar, and Chohara. Its varied characteristics, including geometry, shape, texture, and size, allow for the most accurate identification of a date fruit. Therefore, we outline a novel method for classifying different types of date fruit. (i) First, the date fruit images dataset are pre-processed to evaluate it further processing; (ii) Secondly, segmentation of threshold clustering (STC) technique is employed for eliminating the fore and background of image; (iii) Third, evaluate the statistical feature dataset; (iv) Finally, the optimized feature datasets were observed and deploying different MV classifiers. In the end, we evaluate a better accuracy result. The main objectives of our research study are as follows:

- The experimental processes develop a date fruit identification framework by using the MV technique based on statistical features.
- Due to the different characteristics of date fruit, a rigorous evaluation process is required. This framework was examined using date fruit image datasets.

2. Literature Review

Date fruits are grown a numerous variety all over the world, and different researchers have worked on classifying vegetables and other foods using computer vision [14]. Producing the high-quality of food and managing the agriculture system were both significantly impacted by the early development of food identification technology [15]. The primary components of plants were employed in the categorization process on the leaf. It is examined the more work in [16], the authors discussed a cell-based application for classifying plant kinds by analyzing a plant's leaf characteristics. A different experiment was suggested in [17] which the researchers properly distinguish the sugarcane and cotton varieties using multispectral radiometer data with a result accuracy of 98%. In [18], the researchers used 12 different leaf traits to investigate 32 cultivars, with a 90% accuracy rate as a result.

Hussain et al. [19] described an approach that makes use of deep convolution neural networks (DCNN). The accuracy of the results is 99% overall. In [20], the experimental approach included five different wheat varieties and two kinds of temporal variants. For classification, the dataset relies on photography and radiometry. The accuracy results are 93.14% and 96%, respectively. Jana et al. [21], devised a method to recognize different fruit groups by utilizing the texture and grey level co-occurrence matrix (GLCM) characteristics. A Support Vector Machine (SVM) classifier is used during the experimental process. A novel based feature is used in this work [22], which assesses the color and shape characteristics. An artificial neural network (ANN) was used to classify the eight different kinds of pepper seeds, and an accuracy of 84.94% was achieved. Astuti et al. [23], describes a technique that uses an ANN in conjunction with an SVM classifier to perform classification. Without using a segmentation technique, the ANN achieves 66.7% classification accuracy.

Roche and his fellows [24], a statistical feature-based automated method for classifying fruits and vegetables has been proposed. The experimental method used the K-means algorithm, but classification accuracy has not been mentioned. In [25], The accuracy of the fungi diseases assessed by the researchers using basic threshold and triangular dataset diseases of leaves was 98.60% when they looked at the different sugar cane kinds. In [26], Citrus leaves are used to test the colored-based feature extraction. The accuracy result of the feature reduction strategies was 95%. In [27], Using a deep learning (DL) approach, the researchers assess the classification of plant features based on leaves. A recently created hybrid model that describes the various details of the characteristics of leaves. A study found that the categorization process can be improved by using DL techniques. The author presented a DL-based leaf classification method using a discriminative feature from a plant image.

Hossain and his researchers [28], was discover new fruit types for the industrial sector, use the DL approach. The suggested method is based on the DL model's Oxford-Net and CNN. There are two different kinds of datasets of fruit photos utilized in experiments. Detecting the type of mango used in form and texture-based picture enhancement algorithms. The accuracy result was describe to following and it is reported in [29], SVM technique was deployed for classification process. Attique and his authors [30], proposed a novel genetic algorithm or methodology for the detection of four distinct viruses in various apple varieties. The classification method used an SVM classifier, and the accuracy of the results was 98.10%.
Mohammed Faisal and his fellows [31], introduced a real-time DL and Computer Vision (CV) approach for the categorization of date fruit kinds and also looked at their weight and maturity level. Three subsystems, namely the Dates-Weight-Estimation-System (D-W-E-S), Dates-Maturity-Estimation-System (D-M-E-S), and Date-Type-Estimation System, form the basis of this system (D-T-E-S). For the classification process, a DL and SVM were employed. A greater accuracy result of 99.175% is obtained using the DTES approach. The following analysis of the literature concludes that more research is required before date fruit types may be classified using various feature-based methodologies. Therefore, we require an automated, inexpensive, dependable, and effective system for the kinds of date fruit variations employing the MV approach.

3. Materials and Methods:

3.1. Proposed Methodology

The proposed methodology consists of the following steps: acquisition of the image dataset, preprocessing of the image dataset, segmentation, and feature extraction, optimization of the feature dataset, classification, and result discussion. We found the date fruit images dataset that are the most appropriate and clear by taking them at a height of 2 feet above the date fruit during the experimentation procedure. We deployed a still based stand for fixing a camera. The date fruit images that were obtained must be improved during preprocessing. The segmentation process segregated the date fruit area and removed superfluous surfaces and damaged components. For their texture analysis, we extracted the date fruit qualities of various categories throughout the feature extraction process. During the feature optimization process, we were given the information that was most important for texture analysis, and all extra features were removed to create the optimized features dataset. Instances-based K Nearest Neighbor (IB-KNN) was used in the classification step to distinguish between different date fruit types. Finally, we assessed the performance of the classifiers. Detailing these steps is covered in the following sections.

3.2. Acquisition of Images

As previously mentioned, this study reflects the identification of six types of date fruits, including Fard, Gulistan, Hussaini, Mozawati, Shakar, and Chohara. These date fruit data sets were collected from two distinct farms in the Pakistani districts of Multan and Bahawalpur, and a full digital image dataset was collected in an open environment at the Muhammad Nawaz Sharif University of Agriculture Multan (MNS-UAM) in Pakistan, which had a 42° temperature between June to August and was situated at 30.1475° N and 71.4436° E. The 64-megapixel cell phone camera (Samsung A12) was used to capture all of the images [32], and date fruit are placed on white paper and a clean surface. Each date fruit variety was represented by 100 images in the dataset that was gathered. As a part of our testing, we discovered that taking images of date fruit 2.0 feet above the ground produced the clearest and most suitable results [33], which was shown in Figure 1. Additionally, all captured datasets were gathered in Pakistan at mid-day (12:00 P.M. to 4:00 P.M). Finally, from the 600 image dataset, 100 of each variety of date fruit were chosen, and 10 pieces of date fruits were assembled in each category. As a result, for our investigation, we used a high resolution of 600 (100 x 6) color image datasets with an image resolution of 4190 x 3010 and a 64-bit joint photographic (JPG) format. As a result, the experimental procedure required images of six distinct date fruits, as seen in Figure 1.

Figure 1: Setup of Image Capturing

3.3. Preprocessing of Images

The date fruit should be put on a piece of white paper for optimal results. To increase the size of these available images, Microsoft (MS) has also deployed image scaling tools [34]. Date fruit images were scaled down to 512 x 512 pixels and made into a total of 600 gray-scale images datasets [35], which is shown in **Figure 3**. For feature extraction, three non-overlapping interests of the region (IOR) with 512 x 512 pixel dimensions and 300 (100 x 3) IOR were gathered for each variation, as shown in Figure 4. The following categories of date fruit received a total evaluation of 1800 (300 x 6) IOR. We used a median filter on the available image dataset to remove the noisy information [36]. In this experimental approach, software tools such as WEKA 3.8.1 and Computer-Vision and image-processing (CVIP) were used [37].

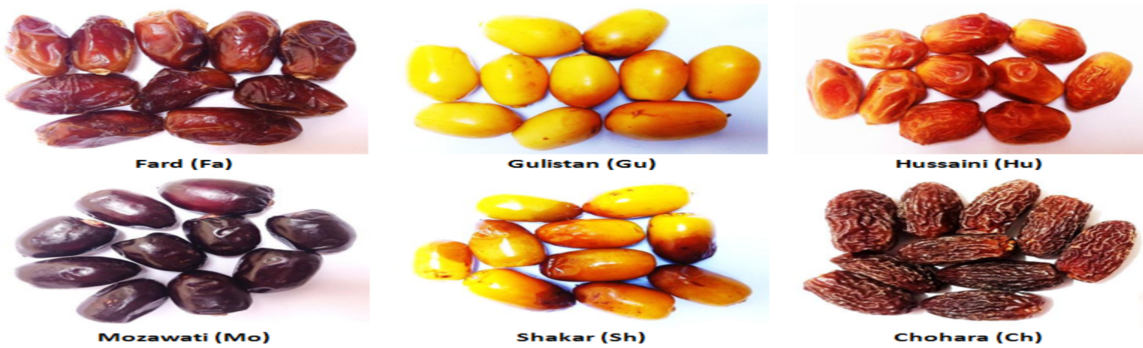

Figure 2. Six Date Fruits Varieties (Digital Images)

Figure 3. Convert Color to Gray Scale Images of 6 Date Fruit Varieties

Figure 4. Create 3 Non-overlapping Interest of Region areas of Date Fruits

3.4. Segmentation

Different date fruit varieties were spread out on a white paper, and we used the segmentation of threshold clustering (STC) technique to extract only the right date fruits section and remove the unnecessary area. With the use of preprocessing classes, image retrieval has improved. This section provides details of the suggested procedure, which is based on different steps. Image segmentation has been carried out using a newly segmentation technique known as STC, which is introduced in Algorithm 1. The three levels of this strategy are as follows.

At level 1, the STC technique is used to estimate the value of a pixel background (PB) dataset from a specific segment and to obtain all of the details regarding the image's backdrop. In the next step, a PB dataset was constructed to take into account all of the threshold method's closest pixel values and to fill in every square inch of the image. In the conclusion, we judged a region to be in the pixel foreground (PF) if the pixel value of the grey level (PVGL) was determined to be greater than (PB). We looked at the IOR to assess the entire cluster.

```
Algorithm 1: Segmentation of Threshold Clustering (STC)
Start {
        For all pixel of image pi = 1 to m
        Pixel back ground value of threshold = (PB)
        Pixel value of gray scale level = PVGL
        Clustering calculate = 0
        If (PVGL > PB)
        Clustering calculate ++
        Pixel fore calculating
        Interest of Region (IOR)
        else
        Not count region of segmentation
        Calculate PB
        End for pi
      }
```

3.5 Feature Extraction

After employed the segmentation process on date fruit, the next step is to employed IOR technique for acquiring the properties of date fruit for classification. Different semi-automatic, manual and automatic techniques were employed for creating IOR. To assess the qualities of the date fruit images dataset, many statistical features were used, including binary, histogram, and texture. Based on these texture features, the dataset's average value is calculated [38]. These statistical features are therefore based on 5 histograms, 28 binary values with 10 pixels (width and height), and 10 values from the texture collection. In this process, each IOR is subjected to an evaluation of total 43 statistical features and we were containing a total of 77,400 (43 x 1800) features dataset. We used the feature optimization method covered in the feature optimization section to increase the accuracy of the results. Here we describe the overview about these features.

3.5.1 Histogram or First Order Feature

The intensity values were examined individual information of pixel by using histogram, which was shown in equation 1.

$$P(f) = \frac{M(n)}{L} \tag{1}$$

Here, In the probability P(f) is used, L evaluates the sum-up value of the pixel, and M(n) describes the total value of gray-scale instances of n. There are following histogram methods are used for statistical evaluation.

Mean: It evaluate the average value of dark (low) and bright (high) mean, which is shown in equation 2.

$$\bar{n} = \sum_{m=0}^{H-1} nH(m)$$

$$= \sum_{e}\sum_{f} \frac{L\,(e,f)}{C} \tag{2}$$

It is representing the total value of gray scale between 0 to 225 and f is representing row and column.

Standard Deviation: Contrast of image was evaluated in standard deviation, which is shown in equation 3.

$$\boldsymbol{\sigma_n} = \sqrt{\sum_{m=0}^{H-1} (\mathbf{n} - \overline{\mathbf{n}})^2 \, \mathbf{H(m)}} \quad (3)$$

Skew: The central value (Median, Mean and Mode) was calculated by skewness, which is shown in equation 4.

Skewness

$$= \frac{1}{\sigma_m^3} \sum_{n=0}^{H-1} (m - \overline{m})^3 \, H(m) \quad (4)$$

Energy: The gray level distribution is measured by energy, that is defines in equation 5.

$$\text{Energy} = \sum_{m=0}^{H-1} [H(m)]^2 \quad (5)$$

Entropy: The number of bits code is measured in image data is evaluate by entropy, which is shown in equation 6.

$$\text{Entropy} = -\sum_{m=0}^{H-1} H\,(m) \log_2 [H(m)] \quad (6)$$

3.5.2 Texture Feature

The coordinates of image area is examined by row and column in texture [34]. Texture has been evaluated on five different methods that are entropy, correlation, energy, inertia and inverse difference.

Energy: It described homogeneity or smoothness by calculating gray levels distribution is defined in equation 7.

$$\text{Energy} = \sum_{e} \sum_{f} (K_{ef})^2 \quad (7)$$

Here K_{ef} is evaluated the values of pixel co-occurrence matrix.

Correlation: It is defining the similarity of pixel at specified pixel distance, which was shown in equation 8.

$$\text{Correlation} = \frac{1}{\sigma_e \sigma_f} \sum_{c} \sum_{d} (c - \mu_e)(d - \mu_f) K_{ef} \quad (8)$$

Here $\boldsymbol{\mu_e}$ **and** $\boldsymbol{\mu_f}$ are the means of e and f respectively.

$$\mu_e = \sum_{np} t \sum_{p} k_{np} \quad (8.1)$$

$$\mu_f = \textstyle\sum_p p \sum_n k_{np} \quad (8.2)$$

$$\sigma_e^2 = \sum_{n} (n - \mu_e)^2 \sum_{n} k_{np} \quad (8.3)$$

$$\sigma_f^2 = \sum_{p} (p - \mu_f)^2 \sum_{p} k_{np} \quad (8.4)$$

Entropy: The content value of image is measured by entropy, which is shown in equation 9.

$$\text{Entropy} = -\sum_{x} \sum_{y} k_{xy} \log_2 k_{xy} \quad (9)$$

Inverse difference (I.D): Homogeneity of image value is measured by I.D, that defined as follow in equation10.

$$\text{Inverse Difference} = \sum_{x}\sum_{y}\frac{k_{xy}}{|x-y|} \tag{10}$$

Inertia: The contrast of image is measure by inertia, which is shown in equation 11.

$$\text{Inertia} = \sum_{x}\sum_{y}(x-y)^2\,k_{xy} \tag{11}$$

3.5.3 Binary Feature

In image processing, binary features use the area, axis of least second moments, Euler number, center of area, and projection to identify the objects and their shapes. We used the feature optimization method covered in the feature optimization section to increase the accuracy of the results [34].

3.5.3.1 Feature Optimization

Feature optimization is the most crucial strategy and methodology in the machine vision approach. By eliminating all superfluous or irrelevant features, just the most noteworthy and accurate features are acquired using classification process [4]. Before putting the categorization approach into practice, it was determined that 43 IOR features are equally important for the date fruit dataset of 77,400 features, nevertheless. It's difficult to manage this enormous amount of feature data. Thus, it has necessary to reduce the feature dataset for accuracy of result. We employed a correlation feature selection and Genetic Search (CFS-GS) methods were used on available optimized feature datasets. The most prominent features were evaluated, shown in **equation (12)**.

$$T_L = \frac{S_{\bar{\sigma}_{xy}}}{\sqrt{S + S\,(S-1)\,\bar{\sigma}_{xy}}} \tag{12}$$

Here, x and L evaluate the most important feature areas, define the feature class of the correlation mean (x & L), and discuss the average feature correlation. CFS-GS technique reduces the original dataset's feature space to 20 statistical features, as shown in Table 1. Finally, the volume of statistical features dataset are reduced from 77,400 (43 x 1800) to 36,000 (1800 x 20) features. A different MV classifiers technique was deployed on these optimized features datasets.

Table 1. A Correlation base Feature (CFS-GS) Selection and Genetic Search Data

Sr. No.	Features Dataset	Sr.No.	Features Dataset
1.	Objects_row_coordinate	11.	Texture_energy_average
2.	Objects_column_coordinate	12.	Texture_energy_range
3.	Area	13.	Inertia_average
4.	Centroid(row_column)	14.	Inertia_Range
5.	Orientation(Axis_of_least_second_moment)	15.	Correlation_range
6.	Histo_Mean	16.	Correlation_average
7.	Histo_Standard_Deviation	17.	Inverse_diff_average
8.	Histo_Skew	18.	Inverse_diff_range
9.	Histo_Energy	19.	Texture_entropy_average

10.	Histo_Entropy	20.	Texture_entropy_range

3.5.3.2 Classification:

Classification is the final step for feature analyzing in the MV approaches. In this phase, the optimized texture feature dataset is given to MV algorithm for the classification of classification of different images dataset [39]. Thus, in our technique, the next level has to evaluate the available dataset of optimal date fruit statistical feature and we employed different MV classifiers on these optimized datasets. We deployed different MV classifiers techniques, but we examine that the instances-based K-Nearest Neighbor (IB-KNN) classifier was achieved a better accuracy. The most fundamental instance-based method is the K-Nearest Neighbor algorithm. In this algorithm, it is assumed that every instance corresponds to a point in n-dimensional space [40] [41]. In the outcome section, a description of the feature dataset that is readily available for the experimental procedure is given first. Figure 5 illustrates the classification architecture for the instances-based K-Nearest Neighbor (IB-KNN) framework that was used on several variations of date fruit.

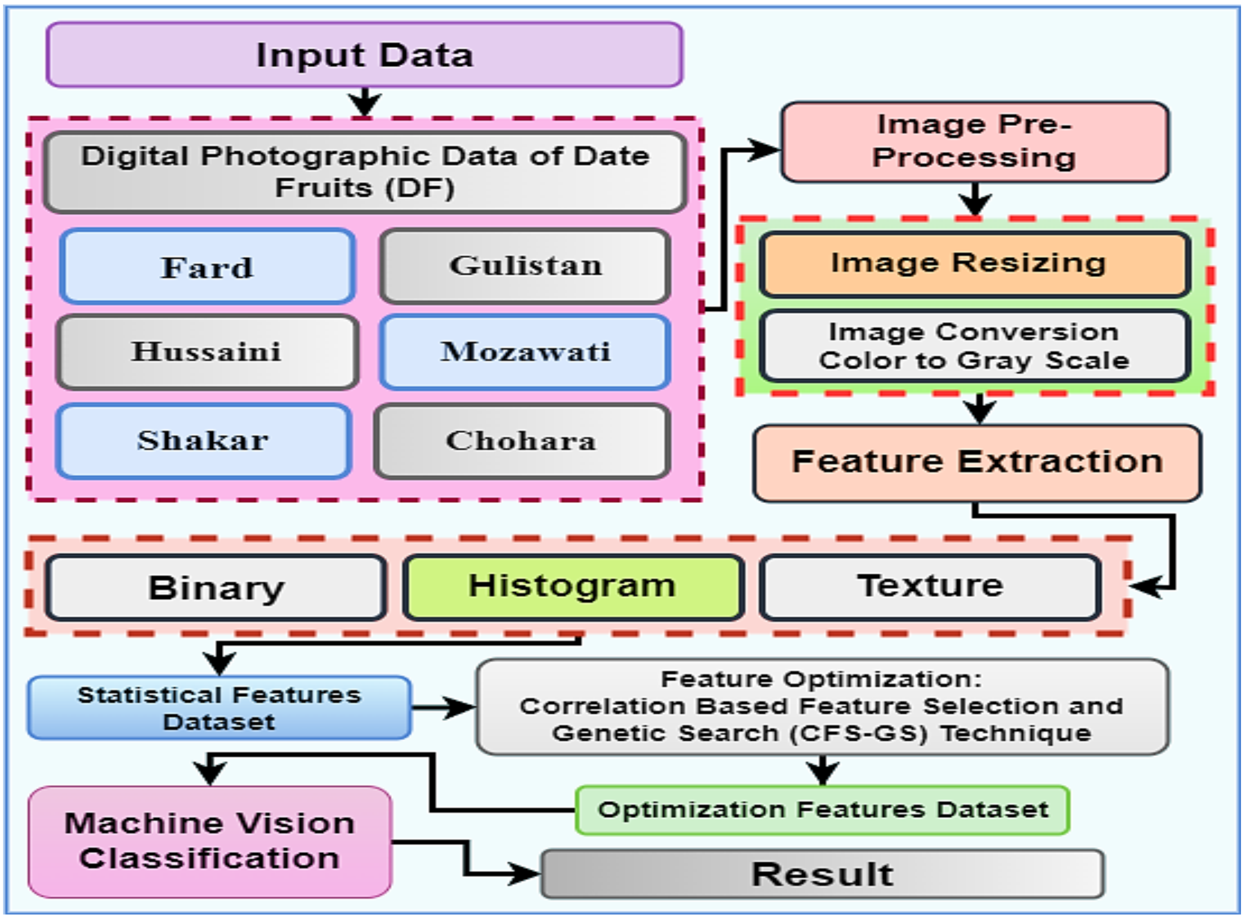

Figure 5. Framework of Six Date Fruit Classifications

4. Result and Discussion:

The main object of this experimental procedure is to identify the different varieties of date fruit based on optimized feature datasets using MV techniques. In the following section, we discuss and present the dataset of classification values. Furthermore, we evaluated a statistical feature dataset using IOR, employed different MV classifiers on the available optimized dataset, and completed all necessary experiments. At the start, the available optimized statistical feature dataset was described based on the following MV classifiers: meta bagging, J48, random forest trees, and decision trees. We found that the following result accuracy was not satisfactory, and the OAR is less than 75%.

Table 2. Matrix of Confusion for Date Fruit types with Naive Bayes (NB) classifier

Date Fruit Varieties	Fard	Gu-listan	Hussaini	Mozawati	Shakar	Chohara	Total Date Fruit Data	OAR
Fard	255	22	8	4	6	5	300	85%
Gulistan	30	252	3	6	6	3	300	84%
Hussaini	8	0	262	15	8	7	300	87.33 %
Mozawati	6	11	12	259	5	7	300	86.33 %
Shakar	8	15	6	4	240	27	300	80%
Chohara	8	4	6	1	19	262	300	87.33 %

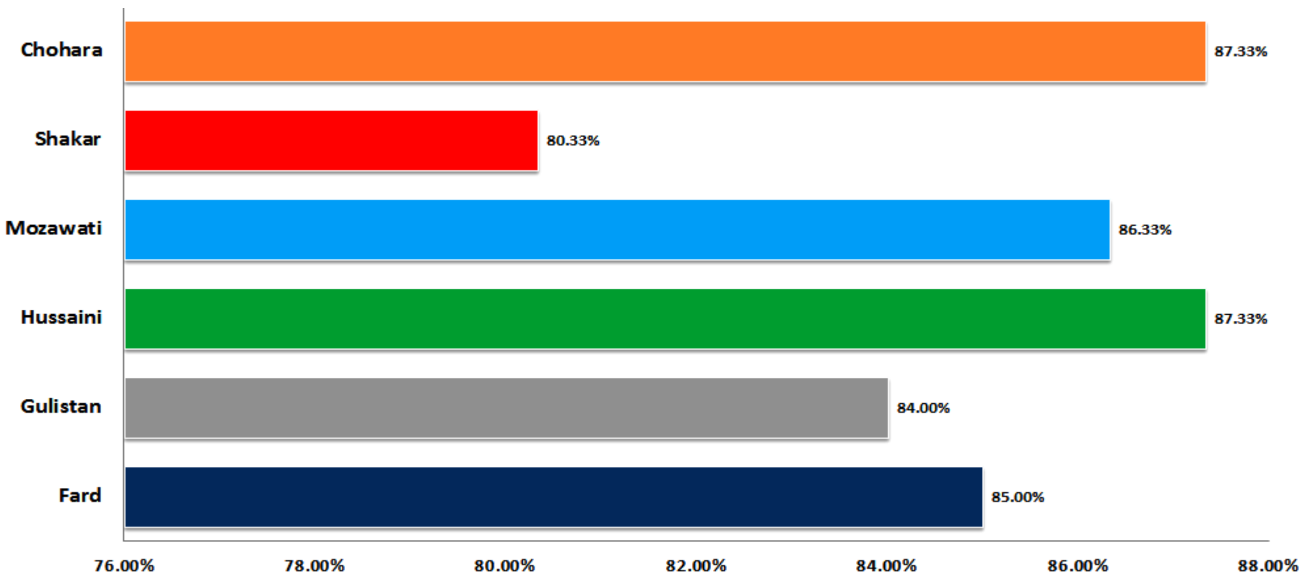

Figure 6. Date Fruit varieties of classification OAR for NB classifier

Furthermore, we repeated the same experimental process using the Naive Bayes (NB) classifier technique, and we describe that the NB classifier result is much better and the OAR value improved from 80% to 87.33% using different varieties of date fruit. The OAR of the following six date fruit categories, including Fard, Gulistan, Hussaini, Mozawati, Shakar, and Chohara, were 85%, 84%, 87.33%, 86.33%, 80%, and 87.33%, respectively. The NB classification output values were represented in Table 2 along with the confusion values matrix. The graphical representation of NB output dataset which was shown in Figure 6.

Table 3. Matrix of Confusion for Date Fruit types with IB-KNN classifier

Date Fruit Varieties	Fard	Gulistan	Hussaini	Mozawati	Shakar	Chohara	Total Date Fruit Data	OAR
Fard	287	9	0	0	1	3	300	95.67%
Gulistan	6	291	1	1	1	0	300	97%

Hussaini	0	0	289	10	1	0	300	96.33%
Mozawati	0	0	8	291	0	1	300	97%
Shakar	1	1	0	1	286	11	300	95.33%
Chohara	0	0	1	1	10	288	300	96%

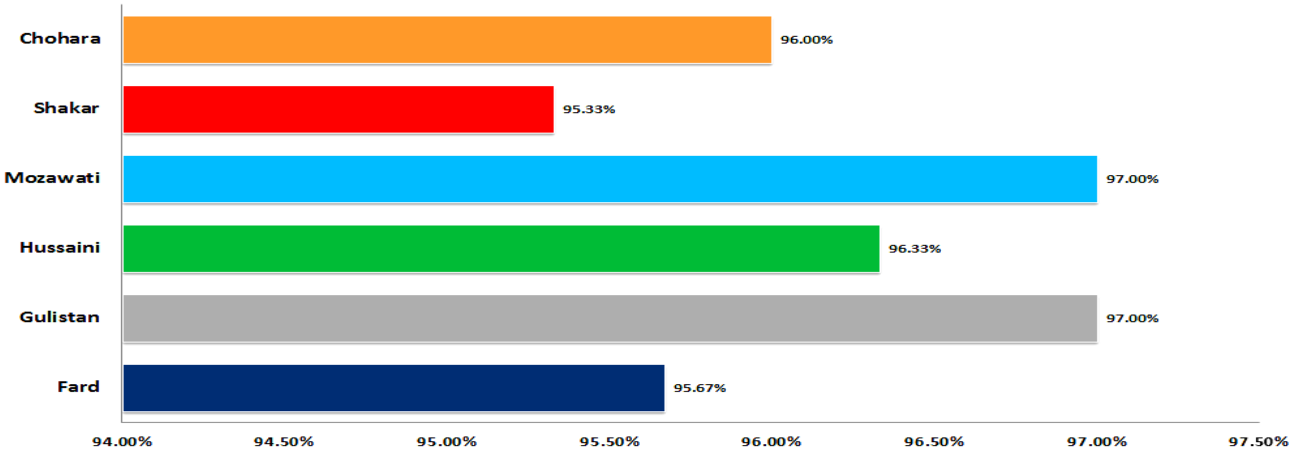

Figure 7. Data Fruit varieties of classification OAR for IB-NN classifier

Thus, we observe that the NB classifier result was not sufficient. Therefore, we deployed the same experimental process using an instance-based nearest neighbor (IB-NN) classifier. We observed that this result is superior, and OAR values improved between 95.33% and 97% on the available statistical feature dataset. The OAR of the following six date fruit categories, including Fard, Gulistan, Hussaini, Mozawati, Shakar, and Chohara, was 95.67%, 97%, 96.33%, 97%, 95.33%, and 96%, respectively. The instances-based K nearest neighbor (IB-KNN) classification output values were represented in Table 3 along with the confusion values matrix. The graphical representation of IB-KNN output dataset which was shown in Figure 6.

A comparison between the existence of the proposed methodology and the other methodologies is shown in Table 4. We conclude that our proposed framework is better than most of the other date fruit results.

Table 4. Comparison between existence and proposed approaches

Ref.	Methodology / Algorithms	Classifiers	OAR
[18]	PCA + Geometric Features	Probabilistic Neural Network (PNN)	90%
[20]	Radiometric and Photographic	Artificial Neural Network	96%
[22]	Harlick + Gabor Feature	Multi-Layer Perception	84.94%
[23]	Fast Fourier Transform (FFT)	Artificial Neural Network	66.7%
[26]	Color Co-occurrence Method (CCM)	1B Model + CCM Textural Feature	95%
Proposed Technique	Binary + Histogram + Texture	Instances Based K Nearest Neighbor (IB-KNN)	97%

5. Conclusion

This experimental process was described for six-date fruit varieties using a statistical feature dataset and MV classifiers. A cell phone camera (Samsung A12) was used to capture the date fruit images of the available dataset in an open environment. We created IOR on the date fruit images and evaluated Binary, Histogram, and Texture based statistical features. A huge volume of feature datasets was decreased by using the CFG-GS optimization methodology and different MV classifiers. NB and IB-KNN classifiers were examined to determine the better classification result using six date fruits, and the NB classifier was evaluated for result accuracy between 80% to 87.33%, but the IB-KNN classifier achieved a better accuracy of 97%. In IB-KNN classification, all computation was postponed until after the function was evaluated, and the function was only locally approximated. Since this approach relies on distance for classification, normalizing the training data can significantly increase accuracy if the features reflect several physical units or have wildly different sizes [42] [43]. In the feature, we used the same techniques or procedures on a farm growing date fruit to detect date fruit infections with expensive but readily controllable tools and resources.

References

1. Mohamed A. Farag, Asmaa Otify, and Mostafa H. Baky, "Phoenix Dactylifera L. Date Fruit By-products Outgoing and Potential Novel Trends of Phytochemical, Nutritive and Medicinal Merits," Food Reviews International, vol. 39 (1), no. https://www.tandfonline.com/doi/abs/10.1080/87559129.2021.1918148, pp. 488-510, May 2021.
2. Salman Qadri et al., "Classification of canola seed varieties based on multi-feature analysis using computer vision approach," International Journal of Food Properties, vol. 24, no. 1, pp. 493-504, March 2021.
3. Abdul Rehman et al., "Prediction of Major Agricultural Fruits Production in Pakistan by using an Econometric Analysis and Machine Learning Technique," International Journal of Fruit Science , vol. 18, no. 4, pp. 445-461, 2018.
4. Tanveer Aslam et al., "Machine learning approach for classification of mangifera indica leaves using digital image analysis," INTERNATIONAL JOURNAL OF FOOD PROPERTIES, vol. 25, no. 1, pp. 1987-1999, September 2022.
5. Shaukat Fayaz Ahmed Train et al., "Pakistan Economic Survey 2020-21," Finance Division Government of Pakistan, Islamabad, Economic Survey FY-2021, June, 2021.
6. Ejaz Wasti et al., "Pakistan Economic Survey 2017-18," Finance Division, Government of Pakistan, Islamabad, Economic Survey 2018 April, 2018.
7. Adel Ahmed Abdul Soad, Shaimaa Mahdi, and Ghulam Sarwar Markhand, "Date Palm Status and Perspective in Pakistan," Springer Science+Business Media Dordrecht 2015, vol. 2, pp. 153-205, Jamuary 2015.
8. Pakistan Defence. (2019, May) Dates Production in Pakistan. [Online]. https://defence.pk/pdf/threads/dates-production-in-pakistan.617337/
9. Muhammad Jehanzed Khan, Zafar Hassan, Babar Ehsan Bajwa, Mubarik Ali, and Muhammad Javed Tareen, "CLUSTER DEVELOPMENT BASED AGRICULTURE TRANSFORMATION PLAN VISION - 2025 ," Planning Commission of Pakistan, Islamabad, Dates Cluster Feasibility and Transformation Study February, 2020.
10. GHULAM SARWAR MARKHAND, ADEL A. ABUL-SOAD, AMEER AHMED MIRBAHAR, and NISAR AHMED KANHAR, "FRUIT CHARACTERIZATION OF PAKISTANI DATES," Pakistan Journal of Botany, vol. 42, no. 6, pp. 3715-3722, December 2010.
11. Summar A. Naqvi and Iqrar Ahmad Khan. (2021, February) Research Gate. [Online]. https://www.researchgate.net/publication/349006477_Date_Palm_in_Pakistan
12. Ghulam Sarwar Markhand, Adel A. Abdul-Soad, Ameer Ahmed Mirbahar, and Nisar Ahmed Kanhar, "FRUIT CHARACTERIZATION OF PAKISTANI DATES," Pakistan Journal of Botany, vol. 42, no. 6, pp. 3715-3722, December 2010.
13. Hafsa Raissouli, Abrar Ali Aljabri, Sarah Mohammed Aljudaibi, Fazilah Haron, and Ghada Alharbi, "Date Grading using Machine Learning Techniques on a Novel Dataset," International Journal of Advanced Computer Science and Applications, vol. 11, no. 8, pp. 758-765, 2020.
14. Md. Robel Mia, Md. Jueal Mia, Anup Majumder, Soummo Supriya, and Md. Tarek Habib, "Computer Vision Based Local Fruit Recognition," International Journal of Engineering and Advanced Technology (IJEAT), vol. 9, no. 1, pp. 2810-2820, October 2019.
15. Salman Qadri et al., "Machine vision approach for classification of citrus leaves using fused features," INTERNATIONAL JOURNAL OF FOOD PROPERTIES, vol. 22, no. 1, pp. 2071-2088, December 2019.
16. Wang Shouzhi, He Dongjian, Li Wen , and Wang Yanchun, "Plant leaf disease recognition based on kernel K-means clustering algorithm," Nongye JixieXuebao = Trans. Chin. Soc. Agri, vol. 40, pp. 152-155, March 2009.

17. MUHAMMAD SHAHZAD SHIFA, M. SHAHID NAWEED, MUHAMMAD OMAR , M. ZEESHAN JHANDIR, and TANVEER AHMED, "CLASSIFICATION OF COTTON AND SUGARCANE PLANTS ON THE BASIS OF THEIR SPECTRAL BEHAVIOR," Pakistan Journal of Botany, vol. 43, no. 4, pp. 2119-2125, August 2011.
18. Stephen Gang Wu et al., "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," in 2007 IEEE international symposium on signal processing and information technology, IEEE, Cairo, Egypt, 2007, pp. 11-16.
19. Israr Hussain, Qianhua He , and Zhuliang Chen, "AUTOMATIC FRUIT RECOGNITION BASED ON DCNN FOR COMMERCIAL SOURCE TRACE SYSTEM," International Journal on Computational Science & Applications (IJCSA), vol. 8, no. 2/3, pp. 1-14, June 2018.
20. E. A. REHMANI et al., "A Comparative Study of Crop Classification By Using Radiometric and Photographic Data," Sindh University Research Journal - SURJ (Science Series), vol. 47, no. 2, 2015.
21. Susovan Jana, Saikat Basak, and Ranjan Parekh , "Automatic Fruit Recognition from Natural Images using Color and Texture Features," in DevIC 2017, March, 2017, pp. DOI:10.1109/DEVIC.2017.8074025.
22. Ferhat Kurtulmuş, İlknur Alibaş, and Ismail Kavdir, "Classification of pepper seeds using machine vision based on neural network," Int J Agric & Biol Eng, vol. 9, no. 1, pp. 51-62, January 2016.
23. Winda Astuti, Satrio Dewanto, Khristian Edi Nugroho Soebandrija, and Sofyan Tan, "Automatic fruit classification using support vector machines: a comparison with artificial neural network," in The 2nd International Conference on Eco Engineering Development 2018 (ICEED 2018), 2018, pp. 1-11.
24. Anderson Rocha, Daniel C. Hauagge, Jacques Wainer, and Siome Goldenstein, "Automatic fruit and vegetable classification from images," Computers and Electronics in Agriculture, vol. 70, no. 1, pp. 96-104, January 2010.
25. Sanjay B. Patil and Shrikant K. Bodhe, "LEAF DISEASE SEVERITY MEASUREMENT USING IMAGE PROCESSING ," International Journal of Engineering and Technology, vol. 3, no. 5, pp. 297-301, November 2011.
26. R.Pydipati , T.F.Burks , and W.S.Lee , "Identification of citrus disease using color texture features and discriminant analysis," Computers and Electronics in Agriculture, vol. 52, no. 1-2, pp. 49-59, June 2006.
27. Sue Han Lee, Chee Seng Chan, Simon Joseph Mayo, and Paolo Remagnino, "How deep learning extracts and learns leaf features for plant classification," Pattern Recognition, vol. 71, no. https://doi.org/10.1016/j.patcog.2017.05.015, pp. 1-13, November 2017.
28. M. Shamim Hossain, Muneer Al-Hammadi, and Ghulam Muhammad, "Automatic Fruit Classification Using Deep Learning for Industrial Applications," IEEE Transactions on Industrial Informatics, vol. 15, no. 2, pp. 1027-1034, Febrauary 2019.
29. Waqar Shahid Qureshi et al., "Machine vision for counting fruit on mango tree canopies," Precision Agriculture, vol. 17, no. DOI:10.1007/s11119-016-9458-5, pp. 224-244, April 2017.
30. Muhammad Attique Khan et al., "An Optimized Method for Segmentation and Classification of Apple Diseases Based on Strong Correlation and Genetic Algorithm Based Feature Selection," IEEE Access, vol. 7, no. DOI: 10.1109/ACCESS.2019.2908040, pp. 46261-46277, March 2019.
31. Mohammad Faisal, Fahad Albogamy, Hebah Elgibreen, Mohammad Algabri, and Abdu Fattoh Alqershi, "Deep Learning and Computer Vision for Estimating Date Fruits Type, Maturity Level and Weight ," IEEE Access Multidisciplinary, vol. 8, no. DOI: 10.1109/ACCESS.2020.3037948, pp. 206770-206782, November 2020.
32. MNS-University of Agriculture, Multan. (2022, September) MNS-University of Agriculture, Multan. [Online]. https://mnsuam.edu.pk/
33. Tanveer Aslam et al., "A Spatial Model of K-Nearest Neighbors for Classification of Cotton (Gossypium) Varieties based on Image Segmentation," LGU Research Journal for Computer Science and IT, vol. 5, no. 1, pp. 24-39, March 2021.

34. Tanveer Aslam et al., "Emotion Based Facial Expression Detection Using Machine Learning," Life Science Journal, vol. 17, no. 8, pp. 35-43, August 2020.
35. Imran Maqbool, Salman Qadri, Dost Muhammad Khan, and Muhammad Fahad, "Identification of Mango Leaves by Using Artificial Intelligence," International Journal of Natural and Engineering Sciences, vol. 9, no. 3, pp. 45-53, October 2015.
36. G. Devarajan , V.K. Aatre , and C.S. Sridhar , "ANALYSIS OF MEDIAN FILTER," in IEEE ACE '90. XVI Annual Convention and Exhibition of the IEEE In India, India, 1990, pp. 274-276.
37. CVIP Tools. (2022, September) CVIP Tools at SIUE. [Online]. https://cviptools.ece.siue.edu/index.php
38. Salman Qadri et al., "A machine vision approach for classification the rice varieties using statistical features," INTERNATIONAL JOURNAL OF FOOD PROPERTIES, vol. 24, no. 1, pp. 1615-1630, September 2021.
39. Tanveer Aslam, Hafiz Muhammad Ijaz, Muzammil Ur Rehman , Syed Ali Nawaz, and Salman Qadri, "Machine Vision Approach for Identification of Four Variant Pakistani Rice Using Multi-Features Dataset," LGU Research Journal of Computer Science and IT, vol. 7, no. https://lgurjcsit.lgu.edu.pk/index.php/lgurjcsit/article/view/348, pp. 35-45, April 2023.
40. Data Cadamia. (2023, April) Data Cadamia. [Online]. https://datacadamia.com/data_mining/knn
41. David W. Aha , Dennis Kibler, and Marc K. Albert , "Instance-based learning algorithms," Machine Learning, vol. 6, no. https://link.springer.com/article/10.1007/BF00153759, pp. 37-66, January 1991.
42. Wikipedia. (2023, April) Wikipedia The Free Encyclopedia. [Online]. https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm#cite_note-3
43. Murat Koklu, Ramazan Kursun, Yavuz Selim Taspinar, and Ilkay Cinar, "Classification of Date Fruits into Genetic Varieties Using Image Analysis," Mathematical Problems in Engineering, no. https://www.hindawi.com/journals/mpe/2021/4793293/, pp. 1-13, November 2021.