Machine learning approach for classification of mangifera indica leaves using digital image analysis

Tanveer Aslam a, Salman Qadri a,*, Syed Furqan Qadri b, Syed Ali Nawaz c, Abdul Razzaq a, Syeda Shumaila Zarren d, Mubashir Ahmad e, Muzammil Ur Rehman f, Amir Hussain g, Israr Hussain h, Javeria Jabeen i, and Adnan Altaf a

Abstract

There is a wide range of horticulture farming in Asia. Mangifera Indica belongs to the species of flowering plant, also publicly recognized as mango. It has a significant local demand as well as a broad export marketplace throughout the world, and is considered as ‘King of Fruits.’ There are many mango varieties and each has its own business market. Efficient identification of the mango varieties is still difficult because of untrained growers and obsolete farming culture, especially in remote areas of the Asia. The primary purpose of this research study was to discriminate mango varieties with the potential of machine learning techniques by analyzing their leaves. For the purpose, we selected leaves of eight mango varieties, namely: Anwar-Ratul (AR), Chaunsa (CHAUN), Langra (LANG), Sindhri (SIND), Saroli (SARO), Fajri (FAJ), Desi (DESI), Alo-Marghan (ALM). A digital cell phone camera captured these datasets in open atmosphere without any well-equipped lab and infrastructure. Binary, histogram, RST, spectral, and texture features were employed for machine learning (ML)-based mango leaf image discrimination. A k-fold (k = 10) cross-validation method was used for ML classification. The k nearest neighbors (KNN) classifier achieved maximum overall classification accuracy (OCA) from 88.33% to 97%.

Introduction

Agriculture and farming are as historic as the human civilization breathes. The world has no means without agriculture and farming. Humanity survives with the art and state of cultivation and livestock. Almost all of the food supplements in our world are based on cultivation and livestock. Agriculture is one of the fundamental pillars of economic growth, and it is considered the backbone of the economy of any country. Agriculture is the largest financial sector in Pakistan’s economy. 1,2 Pakistan is a substantially fertile and growing land of many different crops, seeds, plants, and fruits. The researchers of Pakistan are also playing a significant role in the agriculture sector and trying very hard to uplift the economy of Pakistan in this sector. Pakistan’s agricultural sector is based on four seasons known as winter, autumn, spring, and summer. 3 During different agrarian seasons, Pakistan cultivates and exports many crops, which include cotton, rice, sugarcane, wheat, vegetables, and fruits.

Fruits are a big source of minerals, vitamins, and antioxidants. Fruits are also produced according to their farming seasons. Mango is of them also known as the ‘King of Fruits.’ The notion behind the
'King of Fruits' is that because it is a very delicious, tasty fruit that has a lot of potential benefits too.[4] Mango is not only the most loving fruit of Pakistan in the summer season, but it is also adored throughout the world because of its taste, attractive scent, repertoires of vitamins, and minerals.[5] Mango has a lot of health benefits, including the immense source of magnesium and potassium, and it is also used in different food items such as juices, bakery, jams, and medicine .[6]

Mango is widely farmed in Asia, and the yearly production of mango is given in India (18,779 tons), China (4664 tons), Thailand (34243 tons), Mexico (2197 tons), Indonesia (2184 tons), Pakistan (1606 tons), Brazil (1417 tons), Egypt (1277 tons), Bangladesh (1162 tons), and Nigeria (918 tons). In Pakistan, mango is also the first-ranked summer fruit for exporters and growers in the international market. Pakistan exported 125,000 tons of mangoes worldwide and gained revenue of almost $72 million in the fiscal year 2020. According to the report of the planning commission of Pakistan in the year 2016, Pakistan is among the top ten mango-producing countries, and Pakistan is the sixth-largest exporter of mango fruit.[3] Different mango varieties are produced throughout the country. Provincial mango production and percentage of yield in Pakistan are described as: 399.2 tons in Sindh (36%), 1313.6 tons in Punjab (62%), and 3.0 tons in KPK (1%), and 1.1 tons in Balochistan (1%). Pakistan’s economic and production growth of mangoes (in tons) has improved from 2010 to 2020. District Multan, Bahawalpur, and Rahim Yar Khan are significant areas that produce high-quality mangoes.[7]

Mango is a very delicate fruit with a wide variety of mangoes. For good and quality production, every variety needs to be looked after. The gardeners require much well-timed and useful information for unharmed farming and quality mango production. The essential information is needed to identify the exact mango variety plant. Formally, the experts measure and discriminate the plant through their visual observation and domain knowledge. Recently, identifying the variant plants and categories has been one of the most significant research issues in the agricultural domain. Accurate, well-times, and efficient classification of leave species such as mango, citrus, guava, apple, dates, etc., is highly interested in many reasons. Some of the dominant uses include quality standard farming, prompt disease identification, efficient storage, crop yield, and the efficient food processing.[10]

Leaves of any plant differ based on the complex properties such as their venation pattern, shape, mid-rub, apex, lamina, or margin. And a minor change in these properties is easily misinterpreted and deceived by the human eye, leading to wrong discrimination against the plant. Human pattern recognition based on shape, domain knowledge, size, pattern, and color is inefficient, tiring, and error-prone.[11] For this reason, automated leaf-based plant classification has become a major research issue in the agriculture domain, which leads to the quality standard. Accordingly, leaf-based mango variety classification based on human eye pattern recognition also becomes incredible, irreconcilable, and unreliable, especially becomes more complicated in the field when thousands of the mango plants are without fruiting in the off-season . Therefore, it is compulsory to provide an automated machine-based solution to discriminate the mango plants of different varieties based on their leave characteristics.

In the recent past, researchers greatly inspired machine learning approaches to address the issues that required quick machine vision techniques to replace the inefficient human vision and understanding for pattern recognition in the agricultural domain. In a study, the researchers discriminated the cotton and sugarcane varieties based on spectral features.[13] Mangoes maturity levels were identified on the basis of their color variations, as different maturity levels have their own usages in the various industrial sectors as a raw material . In a research study, automated robots of harvesting system for fruit identification in real environment and quality measurement system are evaluated . Authors investigated machine learning and image analysis techniques in various domains such as medical, natural, and agriculture sectors in . In another study, healthy and defective seeds are examined with an overall accuracy of 90%. The research authors identified gin cu mango variety by employing C language, computer vision, and ANN techniques with 94% accuracy result in. Another research has been conducted to identify automated mango fruit structure by applying texture indices and pattern algorithms, and gained overall 92% accuracy. The authors applied K-means clustering for mango grading using
mango size and ripeness, and obtained 88.88% accuracy.\cite{20} Barley and wheat seeds are identified by employing morphological, statistical, and color features with accuracy of 99%.\cite{21} In a research study, the economic worth of mango is analyzed in Pakistan by the author investigate the in Pakistan by surveying the stakeholders of Sindh and Punjab province.\cite{22} Machine learning model is proposed for the identification of mango varieties based on shape and texture features with overall 70% accuracy.\cite{23} Bio-diversity systems for environmental protection and climate cleanness are discussed at four levels of species, space land, eco-system, and genetic are discussed in the research investigation.\cite{24} In another research study, a machine learning mango grading technique is proposed with overall 95% accuracy.\cite{25} A research experiment was conducted to identify strawberry varieties based on Speeded-up-robust (S-U-R) and ANN by gaining overall 94.34% accuracy.\cite{26} The authors applied a train-test deep learning model for the segmentation and classification of apple image datasets, and achieved overall 97% accuracy.\cite{27} In a research experiment, mango maturities are discriminated by applying SVM classifier with an overall 95% accuracy.\cite{28} The authors investigated agro-ecological change problem, such as floods, strong winds, poor quality, and shortage of grains in West Africa and particularly in Burkina Faso. Different social, structural and technical approaches are discussed to address these problems.\cite{29} Crop management methods to enhance rice index by maintaining good water management and bio_geo_chemical process are investigated in research study.\cite{30} In another research investigation, the agriculture land of rice in China was successfully assessed using an optical sensing Land_Set device, and obtained 75% overall accuracy.\cite{31} The authors proposed a machine learning model for mango quality grading by applying four machine learning classifiers, and achieved classification accuracies between 87.9% and 98.1%.\cite{32} In another research experiment, mango varieties are discriminated by employing 12 simple sequence repeat (SSR) methodologies, and the experiment achieved accuracies ranging from 75% to 100%.\cite{33} In a research study, the authors employed weed kernel ANN model to classify mango images based on the geometrical, color and texture features, and gained accuracies between 90% and 100%.\cite{34}

CNN and DNN models also have been employed in the agricultural computer vision and in a research, mango quality is measured by applying of fully convolutional network (FrCNet) model based on segmentation and feature extraction, and gained 98.9% result accuracy.\cite{35} In another study, the authors examined the colored images of mangoes by employing the region of CNN (R-CNN) feature and achieved overall 80% accuracy.\cite{36} In research study, the authors, applied vgg-16 CNN model to discriminate the four date classes: khala, tamar and rutab and faulty by collecting dataset using a Smartphone, and gained 96.98% result accuracy.\cite{37} The research authors designed a mango grading system based on CNN and ultra-light-weight squeeze-Net algorithm, and achieved overall 97.37% accuracy.\cite{38} There are also some limitations of CNN. A CNN tells you about the class of the objects but not where they are located. It is possible to regress bounding boxes directly from a CNN, but that can only happen one at a time. If multiple objects are in the visual field, then the CNN bounding box regression cannot work well due to interference.\cite{39}

This research study was organized to develop an effective and progressive automated system to discriminate the mango varieties based on their leaf features. Our system overcomes all the indicated problems in mango varieties identifications. The designed system was based on the state-of-the-art machine learning techniques to discriminate the eight dominant mango varieties using their leaf properties. Generally, Pakistan produces many varieties of the mangoes, but the selected eight varieties are more popular and common in use.\cite{40–43} The eight selected mango varieties include Anwar-Ratul (AR), Chaunsa (CHAUN), Langra (LANG), Sindhri (SIND), Saroli (SARO), Fajri (FAJ), Desi (DESI), Alo-Marghan (ALM). Thus our primary objectives are to design a mango leaves based classification system, which comprise setting up mango leave dataset acquisition process, ROIs creation, mango leave feature extraction and optimization, and classification.
Material and methods

Proposed methodology

The proposed methodology comprised the image acquisition, pre-processing, segmentation, feature extraction, optimization classification, and evaluation. Image acquisition contains the capturing of leaf images by using a fixed heighted cell phone camera. Pre-processing involves enhancements of the captured images. Segmentation phase isolated the leaf area and remove extra surfaces and damage parts. During feature extraction we extracted the leave properties of different categories for their texture analysis. Feature optimization phase gave us the most relevant properties for texture analysis and all irrelevant features were removed, we obtained optimized features dataset. Classification step deployed LMT and KNN to discriminate the leave varieties. Finally, we evaluated the performance of the classifiers. The next sections discuss these steps in detail.

Image acquisition

For the research experiment, we collected Mango leaves for all the mentioned varieties, namely: AR, CHAUN, LANG, SIND, SARO, FAJ, DESI, from the garden located in Tehsil Ahmad Pur East of District Bahawalpur Punjab, Pakistan. All the images were captured by a 64 megapixels camera of Samsung cell phone A-12 by placing them on a clean white-colored surface. The collected dataset comprised 100 images of each variety of mangoes. During the experimentation process, we found the leave images most suitable and clear by capturing them at the height of 1.5 feet above the leaves. These images were not too much close and not they were out of focus. We used a stand to fix the cell phone at the required height 1.5 feet above the leaves to standardize the whole image acquisition process. The image acquisition setup is shown in Figure 1. These images were captured on a sunny day at noontime (11:30 am to 3:00 pm). This research is distinguishing because this empiric process is employed without any well-furnished ecosystem and fully equipped research lab. Sample images are shown in Figure 2.

Image pre-processing

During the pre-processing phase each leaf image was resized using image convert and resize software. All non-ROI equal-sized 800 (100 x 8) images were converted into color to gray-scale with the 8-bit format, shown in Figure 3. To remove the noise from these non ROI and standardized images we applied Laplacian filter. 

![Image Acquisition Setup](image.png)
Since, we placed the taken leave images on a surface, so we applied the segmentation procedure to remove the surface and defective leaf parts. For the purpose, we applied range oriented by pixel resolution (RO-PR) algorithm on the leave images. RO-PR algorithm is employed to describe the clustering region of the leave pixels. The total value of pixel (TVP) examines the determined threshold value by considering a unique leaf image cluster. This thresholding level was used as a pixel value of the leaf (PVL) image or initial of the cluster. Resolution of the all neighboring pixels was compared, and the total value of image resolution is examined to create the cluster. In the end, it is evaluated that the purview dataset value of TVP is assessed the same as the PVL, then consider it the same clustering of pixel region (PR) and increase the total data of cluster by evaluating its ROIs. In the similar manner every next adjacent cluster is determined. The complete algorithmic steps of RO-PR scheme are shown in Figure 5.

Feature extraction

After segmenting the mango leaves, next step was to take ROIs and extracting the image properties of these leaves for texture analysis. Different automated, semi-automated, and manual methodologies exist for creating ROIs. Automatic schemes depend highly on image enhancement,[46] whereas the semi-automated and manual methods are based on expert’s judgment.[47] To establish mango leaf dataset, we created three non-overlapping ROIs of sizes 512 × 512 on the images of each category, resulting 300 (100 x 3) ROIs images of each one. The sample ROI’s are shown in Figure 4. In this way, we got a total of 2400 (300 x 8) ROIs image dataset for eight mango varieties. We employed CVIP) and WEKA 3.8.1, the computer vision and image processing software tools.[48–50] Feature acquisition process is most vital process in machine learning-based dataset classification, as it gives the necessary
information for texture analysis. For the study, 28 binary, five histograms, seven RST, 10 texture, and seven spectral features were designated for texture analysis. In this way, total 57 features were extracted from each of the created ROI. The total features vector space (FVS) comprised 136,800 \((8 \times 100 \times 3 = 2400 \times 57)\) features. Next we present the overview of some these features.

**Histogram feature**

A histogram is representation summarizes a series of discrete or continuous form of numerical data in a format that is easy to interpret. It groups the data into logical ranges of different heights. These are also known as bins. It is represented as a frequency graph of intensity values of underlying pixels. We need the interpretation of intensity distributions of the individual pixels of mango leaves.

**Texture feature**

The texture is a significant to identify objects or regions of interest in an image. In images texture represents a specific pattern, which is usually repeated sequentially in the image. Any change in the underlying texture is determined by the texture features. As the leave changes, we need to capture the similarities and changes in the underlying patterns of the leaves.

**Binary feature**

Binary features are widely used in texture analysis, object detection, facial detection and many other areas. They are fast and powerful features working at local level to analyze the patterns. To examine the
leaves at local levels we also applied binary features, including area, centroid, orientation, euler number, thinness, aspect ratio, and projection.\textsuperscript{[12,40]}

**Rotation, scaling, and translation (RST) feature**

To analyze and identify the objects based on their scale and angular information we apply the RST properties. The geometrical interpretations are applied to determine the mango leaves.\textsuperscript{[12,40]}

**Spectral feature**

Spectral features are based on time, amplitude and frequency. Fourier transformation is used convert into the frequency domain. Density variations in the incoming signals are measures with sensitivity. Sensitive variations in the intensities are determined using the deep properties of wavelength. In this way, the closest objects are also able to distinguish. As in the case of mango leaves, we require to discriminate the minor variations in the leaves of different mango varieties.\textsuperscript{[12]}

**Feature optimization**

Feature optimization is very significant in machine learning and dataset classification. During this phase, the most valuable features for the texture analysis are extracted and the irrelevant features are eliminated.\textsuperscript{[51]} In the previous phase, we extracted 57 features from each ROI image obtained a multi-feature FVS, comprising 136,800 features. Such a large-scaled FVS was not sufficient for efficient texture analysis. Thus, it was compulsory to minimize the size of FVS. For this intent, we applied a supervised correlation-based feature selection (CFS) technique on the FVS dataset, and obtained the most significant multi-feature FVS. The equation of CFS is given below\textsuperscript{[10]}

$$S_T = \frac{N_{\sigma_{ab}}}{\sqrt{N + N(N - 1)\bar{\sigma}_{ab}}}$$

where it describes the prominent subset T with b number of features. The equation computes the mean correlation of features, called the average feature correlation. CFS scheme gave a reduced FVS volume of 19 the most dominant set of features, shown in Table 1. Hence, volume of FVS was reduced from 136,800 to 45,600 (2400 x 19). This minimized multi-feature dataset was input to the ML classifiers.

**Classification**

Classification is the final phase in the machine learning-based texture analysis. During this step, the extracted feature dataset is given input to the algorithm for the discrimination of different image classes.\textsuperscript{[52]} However, in our methodology, the next step was to classify the optimized mango leave’s multi-feature dataset into the corresponding mango varieties. For the purpose, we deployed two machine learning classifiers on the optimized mango leave’s multi-features FVS. We tried different classifiers, but we got promising results on LMT and KNN. LMT is a nice supervised train-test classification model that fuses regression and decision tree.\textsuperscript{[53]} Finally, the KNN gave more better accuracy as compared to LMT. The KNN classification is traditional, nonparametric, and simple to understand, as it works by measuring the distance between a group of data points defined by the value of k. For problems where all the data points are well defined or contain less non-linearity, then KNN is the best choice. Furthermore, it is calm to implement and has no complex parameters to regulate tasks.\textsuperscript{[54,55]} The complete framework of the proposed methodology is shown in Figure 6.
The purpose of this research was to identify the variant mango varieties using ML classifiers. In this section, we present and discuss the output values after the classification. We examine all the outputs with field-based dataset values. In this experimental process, multi-feature ROIs dataset were created, and we input optimized feature set to different machine learning classifiers and completed the experimental procedures. In the start, the optimized multi-features dataset was evaluated by employing ML classifiers, meta bagging, tree J48, Logistics, and tree random forest. We observed that the output results for the classifiers were not sufficient and the result accuracy remained less than 80%.

During the experimentation process, when the same procedure was repeated for LMT, a Tree-based classification method, we observed that the results were slightly improved and the classification accuracy
rate increased from 80.33% to 88.33%, for the different mango varieties. The overall classification accuracy (OCA) results of the following eight mango varieties, AR, CHAUN, LANG, SIND, SARO, FAJ, DESI, and ALM were 86%, 80.67%, 86.33%, 88.33%, 82.33%, 82%, 80.33%, and 82%, respectively. The LMT classification rates are shown in Table 2. The values in Table 2 present the overall classification accuracy for each mango class, the cross-class occurring of the sample or the missclassifications. The difference in classification results was observed due to the variants in the acquired features. The graph of the LMT classification performance to discriminate the eight mango varieties is shown in Figure 7.

Since the classification results were not still satisfactory on LMT classifier, so, we repeated the experiment for KNN, a straightforward, conventional nonparametric classification scheme by determining the k distance between different groups of data points. The classification accuracies of mango varieties were improved on KNN classifiers. When we applied the KNN classifiers, the overall classification accuracy (OCA) results of the following eight mango varieties, AR, CHAUN, LANG, SIND, SARO, FAJ, DESI, and ALM were 94.33%, 97%, 96%, 96.67%, 94.67%, 94.33%, 88.33%, and 89.67% respectively. The KNN classification rates are shown in Table 3. The values in Table 3 show the overall classification accuracy for each mango class, the cross-class occurring of the sample or the missclassifications. The graph of the KNN classification performance to discriminate the eight mango varieties is shown in Figure 8.

Since the classification results were not still satisfactory on LMT classifier, so, we repeated the experiment for KNN, a straightforward, conventional nonparametric classification scheme by determining the k distance between different groups of data points. The classification accuracies of mango varieties were improved on KNN classifiers. When we applied the KNN classifiers, the overall classification accuracy (OCA) results of the following eight mango varieties, AR, CHAUN, LANG, SIND, SARO, FAJ, DESI, and ALM were 94.33%, 97%, 96%, 96.67%, 94.67%, 94.33%, 88.33%, and 89.67% respectively. The KNN classification rates are shown in the Table 3. The values in Table 3 show the overall classification accuracy for each mango class, the cross—class occurring of the sample

| Mango Varieties | AR  | CHAUN | LANG | SIND | SARO | FAJ  | DESI | ALM | Total Datasets | OCA (%) |
|-----------------|-----|-------|------|------|------|------|------|-----|----------------|---------|
| AR              | 258 | 26    | 2    | 3    | 4    | 5    | 1    | 1   | 300            | 86      |
| CHAUN           | 29  | 242   | 3    | 6    | 8    | 8    | 1    | 3   | 300            | 80.67   |
| LANG            | 2   | 5     | 259  | 16   | 2    | 5    | 8    | 3   | 300            | 86.33   |
| SIND            | 1   | 5     | 14   | 265  | 4    | 8    | 2    | 1   | 300            | 88.33   |
| SARO            | 6   | 9     | 2    | 4    | 247  | 22   | 1    | 9   | 300            | 82.33   |
| FAJ             | 6   | 8     | 5    | 4    | 22   | 246  | 6    | 3   | 300            | 82      |
| DESI            | 0   | 3     | 2    | 1    | 3    | 8    | 241  | 42  | 300            | 80.33   |
| ALM             | 3   | 1     | 1    | 1    | 4    | 6    | 37   | 247 | 300            | 82      |

Figure 7. LMT Classification Performance Graph for the Eight Mango Varieties.
The graph of the KNN classification performance to discriminate the eight mango varieties is shown in Figure 8.

A comparison of our proposed technique with the other state-of-the-art methods is presented in Table 4. The table lists the determined features, main methodology, classification scheme, and results obtained. We can observe that our proposed classification framework outperformed most of the other mango classification schemes.

**Table 3. Confusion Matrix Table of KNN Classifier for Eight Mango Leaves Varieties.**

| Mango Varieties | AR   | CHAUN | LANG | SIND | SARO | FAJ  | DESI | ALM  | Total Datasets | OCA (%) |
|-----------------|------|-------|------|------|------|------|------|------|----------------|---------|
| AR              | 283  | 12    | 0    | 0    | 0    | 3    | 0    | 2    | 300           | 94.33   |
| CHAUN           | 6    | 291   | 0    | 0    | 1    | 1    | 0    | 1    | 300           | 97      |
| LANG            | 0    | 1     | 288  | 8    | 2    | 1    | 0    | 0    | 300           | 96      |
| SIND            | 0    | 0     | 9    | 290  | 1    | 0    | 0    | 0    | 300           | 96.67   |
| SARO            | 1    | 0     | 0    | 0    | 284  | 12   | 2    | 0    | 300           | 94.67   |
| FAJ             | 0    | 0     | 0    | 1    | 14   | 283  | 0    | 2    | 300           | 94.33   |
| DESI            | 0    | 1     | 1    | 0    | 1    | 3    | 265  | 29   | 300           | 88.33   |
| ALM             | 1    | 1     | 0    | 0    | 2    | 0    | 27   | 269  | 300           | 89.67   |

**Figure 8. KNN Classification Performance Graph for the Eight Mango Varieties.**

**Table 4. Comparison between proposed and existing approach.**

| Link/References | Methodology/ Technique | Classification | OCA (%) |
|-----------------|------------------------|----------------|---------|
| [12]            | Texture + Histogram + Binary + Spectral + RST | Artificial Neural Network (ANN) | 95–98   |
| [40]            | Binary + RST Invariant | Artificial Neural Network (ANN) | 96      |
| [18]            | C Language + Deep Learning | Artificial Neural Network (ANN) | 94      |
| [36]            | F1 Score, Threshold    | Region Convolutional Neural Network (R-CNN) | 80–90   |
| [23]            | Entropy + Variance + Correlation etc. | LDA + PCA | 70, 83  |
| [56]            | RGB + Recursive Feature Elimination (RFE) | Support Vector Machine (SVM) | 90      |
| Proposed Technique | Binary + Histogram + Texture + RST Invariant + Spectral | LMT Tree & K Nearest Neighbors (K-NN) | 88.33, 97 |

or the missclassifications. The graph of the KNN classification performance to discriminate the eight mango varieties is shown in Figure 8.

A comparison of our proposed technique with the other state-of-the-art methods is presented in Table 4. The table lists the determined features, main methodology, classification scheme, and results obtained. We can observe that our proposed classification framework outperformed most of the other mango classification schemes.

**Conclusion**

This research was accomplished to discriminate eight mango varieties using multi-feature analysis of their leaves, based on machine learning classifiers. A 64 megapixels Samsung cell phone camera
captured the leave images of the selected mango varieties in an open climate, and mango leaves image dataset was constructed. We created ROIs on the leave images and extracted binary, histogram, RST, spectral, and texture features from the ROIs and formed a multi-feature vector space. Large volume of the features vector was reduced by applying CFS optimization. Different machine learning classifiers were experimented on the optimized feature vector, using k-fold (k = 10) cross-validation. Two of the classifiers, the LMT and KNN gave good classification accuracies to discriminate the eight mango varieties. LMT set out classification accuracy rates between 80.33% and 88.33%, for different mango varieties, whereas the KNN classifier achieved the maximum overall classification accuracy between 88.33% and 97%. LMT has a regressive nature and it establishes logistic iterations, cross validation, and tree pruning. The architecture of LMT fuses two algorithms, the regression and tree induction. This property makes LMT able to exhibited good performance on the multi-\(^^\)\ A KNN classifier depending on measuring the distances, and performs well without normalizing when incoming features belong to small scaled variants.\cite{54,55} Since, we applied it on the mangos, which have small scaled variations, so it outperformed on this dataset. Finally, observing the other studies, it is shown that the classification results high on the designated set of features. In the future, our proposed classification framework may be experimented on other datasets comprising small-scaled features variations. Another study may be conducted to test the proposed classification model on a dataset having opposite characteristic, means on dataset comprising large scaled feature variations.

**Acknowledgments**

This research was supported by the Researchers Supporting Project number (RSP2022R476), King Saud University, Riyadh, Saudi Arabia.

**Data availability**

The data used to support the findings of this study are available from the corresponding author upon request.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**ORCID**

Salman Qadri \(\text{http://orcid.org/0000-0002-3503-6535}\)

**References**

[1] Syed, A.; Raza, T.; Tufail Bhatti, T.; Eash, N. S. Climate Impacts on the Agricultural Sector of Pakistan: Risks and Solution. *Environ. Challenges*. January 2022, 6(10.1016/j.envc.2021.100433): 1–5.

[2] PBS, Pakistan Bureau of Statistics . (2022, August) Agriculture Statistics. [Online]. https://www.pbs.gov.pk/content/agriculture-statistics PBS (Pakistan, Islamabad: Goverment of Pakistan).

[3] Mubarik, A. Ministry of Planning Development and Special Initiatives , Islamabad, Pakistan, Cluster Development Based Agriculture Transformation Plan Vision-2025 . February 2022.

[4] Kumari Behera, S.; Sangita, S.; Kumar Rath, A., and Kumar Sethy, P. Automatic Classification of Mango Using Statistical Feature and SVM. *Springer Nat. Singapore Pte Ltd* 4 . 2019, 469–475.

[5] Shih-Lun, W.; Tung, H.-Y., and Hsu, Y.-L. Deep Learning for Automatic Quality Grading of Mangos: Methods and Insights. in *19th IEEE International Conference on Machine Learning and Applications (ICMLA)* MIAMI, FLORIDA, DOI:10.1109/ICMLA51294.2020.00076, 23 November 2020, pp. 1–8.

[6] Harna Yossey, E.; Pranata, J.; Wijaya, T.; Hermawan, H.; Budiharto, W. Mango Fruit Sortation System Using Neural Network and Computer Vision, in *2nd International Conference on Computer Science and Computational Intelligence 2017, ICCSCI 2017*, Bali, Indonesia, 2017, pp. 596–603.
[7] Imtiaz Ahmad, D., “Pakistan Economic Survey 2019-20,” Economic Adviser’s Wing, Finance Division Government of Pakistan, Islamabad, Pakistan, June, 2020.

[8] Nagle, M. Determination of Surface Color of ‘All Yellow’ Mango Cultivars Using Computer Vision. *Int. J. Agric. Biol. Eng.* 2016, 9(1), 42–50. January.

[9] Truong Minh Long, N.; Truong Thinh, N. Using Machine Learning to Grade the Mango’s Quality Based on External Features Captured by Vision System. *Appl. Sci. Art* August 2020, 10(5775): 1–25.

[10] Qadri, S.; Furqan Qadri, S.; Husnain, M.; Saad Missen, M. M.; Khan, D. M.; Muzammil-Ul, R.; Razzaq, A., and Ullah, S. Machine Vision Approach for Classification of Citrus Leaves Using Fused Features. *Int. J. Food Prop.* 2019 December, 22(1), 2071–2088. doi: 10.1080/10294291.2019.1703738.

[11] Saleem, R. Mango Leaf Disease Recognition and Classification Using Novel Segmentation and Vein Pattern Detection. *Appl. Sci. Art*. 2021, 11(1901), 1–12. December.

[12] Qadri, S.; Furqan Qadri, S.; Razzaq, A.; Ul Rehman, M.; Ahmad, N.; Nawaz, S. A.; Saher, N.; Akhtar, N., and Khan, D. M. Classification of Canola Seed Varieties Based on multi-feature Analysis Using Computer Vision Approach. *Int. J. Food Prop.* 2021 March, 24(1), 493–504. doi: 10.1080/10294291.2021.1900235.

[13] Shahzad Shifa, M.; Shahid Naweed, M.; Omar, M.; Zeeshan Jhandir, M.; Ahmed, T. Classification of Cotton and Sugarcane Plants on the Basis of Their Spectral Behavior. *Pak. J. Bot.* August 2011, 43(4): 2119–2125.

[14] Castillo-Castaneda, E. Neural Network Training in SCILAB for Classifying Mango (Mangifera Indica) according to Maturity Level Using the RGB Color Model. *Aim Sci.* 2021 July, 1(I), 186–198. doi:10.3934/steme.2021014.

[15] Lin, G.; Tang, Y.; Zou, X.; Cheng, J.; Xiong, J. Fruit Detection in Natural Environment Using Partial Shape Matching and Probabilistic Hough Transform. *Precis. Agric.* April 2019, 21(I), 160–177. DOI: 10.1007/s11119-019-09662-w.

[16] Strzelecki, M.; Szczypinski, P.; Materka, A.; Klepaczko, A. A Software Tool for Automatic Classification and Segmentation of 2D/3D Medical Images. *Elsevier Nucl. Instrum Methods Phys. Res. A*. February, 2013, 702, 137–140. DOI: 10.1016/j.nima.2012.09.006.

[17] José, M. V. G.; Andreu-Garcia, G.; Potter, P.; Rodas-Jordá, Á. Automatic Corn (Zea Mays) Kernel Inspection System Using Novelty Detection Based on Principal Component Analysis. *Biosyst. Eng.* January 2014, 117 (Special), 94–103. DOI: 10.1016/j.biosystemseng.2013.09.003.

[18] Harna Yossy, E.; Pranata, J.; Wijaya, T.; Hermawan, H.; Buddiharto, W. Mango Fruit System Using Neural Network and Computer Vision. *Procedia Comput. Sci*. Elsvier January 2017, 116(7), 596–603. DOI: 10.1016/j.procs.2017.10.013.

[19] He, J. Research on Mango Shape Evaluation Method Based on Machine Vision. *Mod. Agric. Equip.* 2021, 42(1), 56–60.

[20] Supekar, A.; Wakode, M. Multi-Parameter Based Mango Grading Using Image Processing and Machine Learning Techniques. *INFOCOMP*. December 2020, 19(2): 175–187.

[21] Guevara-Hernandez, F.; Gomez-Gil, J. A Machine Vision System for Classification of Wheat and Barley Grain Kernels. *Spanish J. Agric. Res.* 2011, 9(3), 672–680. DOI: 10.5424/siar/20110903-140-10.

[22] Badar, H.; Ariyawardana, A.; Collins, R. Dynamics of Mango Value Chains in Pakistan. *Pak. J. Agric. Res.* April 2019, 56(2): 523–530.

[23] Abbas, Q.; Muddasir Iqbal, M.; Niazi, S.; Noureen, M. Mango Classification Using Texture & Shape Features. *Int. J. Comput. Sci. Network Secur.* August 2018, 18(8): 132–138.

[24] Shiming, L. U. O. Agroecological Rice Production in China: Restoring Biological Interactions. . 2018. (Rome: FOOD AND AGRICULTURE ORGANI ZAT ION OF THE UNI TED NAT IONS)

[25] Ahmad Saad, F. S. Y. A. H. I. R.; Firdaus Ibrahim, M.; Md Shakaff, A. Y.; Zakaria, A.; Zaki Abdullah, M. Shape and Weight Grading of Mangoes Using Visible Imaging. *Comput. Electron. Agric.* May 2015, 5(6), 51–56. DOI: 10. 1016/j.compag.2015.05.006.

[26] Shin, J.; Chang, Y. K.; Heung, B.; Nguyen-Quang, T.; Price, G. W., and Al-Mallahi, A. Effect of Directional Augmentation Using Supervised Machine Learning Technologies: A Case Study of Strawberry Powdery Mildew Detection. *Biosyst. Eng.* 2020 March, 194, 49–60. doi: 10.1016/j.biosystemseng.2020.03.016.

[27] Li, L. S.; Eng, P. P. Apple Target Detection Based on Improved faster-RCNN Framework of Deep Learning. *Mach. Des. Res.* 2019, 35(5), 24–27.

[28] Sekhar Nandi, C.; Tudu, B.; Koley, C. A Machine Vision Technique for Grading of Harvested Mangoes Based on Maturity and Quality. *IEEE Sen. J.* June 2016, 16(16), 6387–6396. DOI: 10.1109/JSEN.2016.2580221.

[29] Korbeogo, G. Reinventing New Systems of Crop Production in Time of agro-ecological Change in Burkina Faso. *Afr. J. Environ. Sci. Technol.* May 2015, 9(5), 464–472. DOI: 10.5897/AJEST2012.046.

[30] Yang, J.; Zhang, J. Crop Management Techniques to Enhance Harvest Index in Rice. *J. Exp. Bot.* April 2010, 61(12), 3177–3189. DOI: 10.1093/jxb/erq112.

[31] Froliking, S.; Qiu, J.; Boles, S.; Xiao, X.; Liu, J.; Zhuang, Y.; Li, C., and Qin, X. Combining Remote Sensing and Ground Census Data to Develop New Maps of the Distribution of Rice Agriculture in China. *Global Biogeochern. Cycles* 2002 November, 16(4), 1–11. doi: 10.1029/2001GB001425.

[32] Truong Thinh, N.; Truong Minh Long, N. Using Machine Learning to Grade the Mango’s Quality Based on External Features Captured by Vision System. *Appl. Sci. Art.* August 2020, 10( doi:10.3390/app10175775): 1–25.
[33] Nazish, T.; Shabbir, G.; Ali, A.; Sami-ul-Allah, S.; Naeem, M.; Javed, M.; Batool, S.; Arshad, H.; Hussain, S. B.; Aslam, K.; et al. Molecular Diversity of Pakistani Mango (Mangifera indica L.) Varieties Based on Microsatellite Markers. *Genet. Mol. Res.*** 2017, *16*(2), 1–8. doi: 10.4238/gmr16029560.

[34] Granitto, P. M.; Navone, H. D.; Verdes, P. F.; Ceccatto, H. A. Weed Seeds Identification by Machine Vision. *Comput. Electron. Agric.* November 2001, 33(2002), 91–103. DOI: 10.1016/S1861-1699(02)00004-2.

[35] Saleem, R.; Hussain Shah, J.; Sharif, M.; Jillani Ansari, G. Mango Leaf Disease Identification Using Fully Resolution Convolutional Network. *Comput. Mater. Continua* May 2021, 69(3), 3581–3601. DOI: 10.32604/cmc.2021.017700.

[36] Borianne, P.; Sarron, J.; Borne, F.; Faye, E. Deep Mangoes: From Fruit Detection to Cultivar Identification in Colour Images of Mango Trees. In *Disp* 19, Oxford, United Kingdom, 2019; pp 1–8.

[37] Nasiri, A.; Taheri-Garavand, A.; Zhang, Y.-D. Image-based Deep Learning Automated Sorting of Date Fruit. *Postharvest. Biol. Technol.* April 2019, 153, 133–141. DOI: 10.1016/j.postharvbio.2019.04.003.

[38] Zheng, B.; Huang, T. Mango Grading System Based on Optimized Convolutional Neural Network, *Math. Probl. Eng.*, no. 10.1155/2021/2652487, pp. 1–11, September 2021.

[39] More, S.; Shiordkar, V.; Joshi, R.; Thakur, N. Overcoming the Drawbacks of Convolutional Neural Network Using Capsule Network. *IOSR J. Comput. Eng.* March–April 2019, 21(2): 6–11.

[40] Maqbool, I.; Qadri, S.; Muhammad Khan, D.; Fahad, M. Identification of Mango Leaves by Using Artificial Intelligence. *Int. J. Nat. Eng. Sci.* October 2015, 9(3): 45–53.

[41] Dananjayan, S.; Tang, Y.; Zhuang, J.; Hou, C.; Luo, S. Assessment of state-of-the-art Deep Learning Based Citrus Disease Detection Techniques Using Annotated Optical Leaf Images. *Comput. Electron. Agric.* February 2022, 193, 106658. DOI: 10.1016/j.compag.2021.106658.

[42] Abdurakhmonov, I. Y., et al., *Plant Breeding*. 1st Rebrovic, M., Smiljanic, T., Abdurakhmonov, I. Y., Eds. Croatia: INTECH OPEN ACCESS PUBLISHER. December 2011.

[43] Kirairatitanaprak, K., et al. Development of Paddy Rice Seed Classification Process Using Machine Learning Techniques for Automatic Grading Machine. *Hindawi J. Sens.* 2020, 704(1310), 1–14. July.

[44] Aslam, T., et al. A Spatial Model of K-Nearest Neighbors for Classification of Cotton (Gossypium) Varieties Based on Image Segmentation. *LGU Res. J. Comput. Sci.* IT 2021, 5(1), 24–38. March.

[45] L3HARRIS. (2022, August) L3HARRIS Geospatial. [Online]. https://www.l3harrisgeospatial.com/docs/laplacian filters.html#:~:text=A%20Laplacian%20filter%20is%20an%20edge%20detector%20continuous%20progression.

[46] Qadri, S.; Khan, D. M.; Qadri, S. F.; Razzaq, A.; Ahmad, N.; Jamil, M.; Nawaz Shah, A.; Shah Muhammad, S.; Saleem, K.; Awais, S. A., et al. Multisource Data Fusion Framework for Land Use/Land Cover Classification Using Machine Vision. *Hindawi J Sens*. September 2017, 2017, 1–9. DOI:10.1155/2017/3515418.

[47] Shehzad, M., et al. Machine Vision Based Identification of Eye Cataract Stages Using Texture Features. *Life Sci. J.* 2020, 17(8), 44–50. August.

[48] Qadri, S.; Aslam, T.; Nawaz, S. A.; Saber, N.; Razzaq, A.; Ur Rehman, M.; Ahmad, N.; Shahzad, F.; Furqan Qadri, S.; et al. A Machine Vision Approach for Classification the Rice Varieties Using Statistical Features. *Int. J. Food Prop.* 2021, September, 24(1), 1615–1630. doi: 10.1080/10942912.2021.1986523.

[49] Aslam, T., et al. Emotion Based Facial Expression Detection Using Machine Learning. *Life Sci. J.* 2020, 17(8), 35–43. August.

[50] Zopotoczny, P. Discrimination of Wheat Grain Varieties Using Image Analysis and Multidimensional Analysis Texture of Grain Mass. *Int. J. Food Prop.* August 2014, 17(1), 139–151. DOI: 10.1080/10942912.2011.615085.

[51] Faisal, M.; Fahad Albogamy, H. E. B. A. H. E. L. G. I. B. R. E. N.; Algabri, M.; ALQERSHI, F. A. T. T. O. H. A. B. D. U.; Alqersh, F. A. Deep Learning and Computer Vision for Estimating Date Fruits Type, Maturity Level, and Weight. *IEEE Access Multidiscip.* November 2020, 8(2020), 206770–206782. DOI: 10.1109/ACCESS.2020.3037948.

[52] R Kuma Tripathi, M.; Maktedar, D. D. A Role of Computer Vision in Fruits and Vegetables among Various Horticulture Products of Agriculture Fields: A Survey. *Information Process. Agric.* July 2020, 7(2020), 183–203. DOI: 10.1016/j.inpa.2019.07.003.

[53] Fazakas, N.; Karlos, S.; Kotsiantis, S.; Sgarbas, K. Self-Trained LMT for Semisupervised Learning, *Comput. Intell. Neurosci.*, vol. Article ID 3057481, no. 10.1155/2016/3057481, pp. 1–13, January 2016.

[54] Sharma, Shivam; (2021, May). K-Nearest Neighbour: The Distance-Based Machine Learning Algorithm. [Online]. https://www.analyticsvidhya.com/blog/2021/05/knn-the-distance-based-machine-learning-algorithm/ May 2021 (Analytics Vidhya)

[55] Joshua Hutusoi, B.; Sofyan, H.; Richard Kodong, F. Classification of Mango Plants Based on Leaf Shape Using GLCM and K-nearest Neighbor Methods. *Comput. Inf. Process. Lett.* November 2021, 1(1), 1–7. DOI: 10.31315/cip.v1i1.6124.

[56] Sekhar Nandi, C.; Tudu, B.; Koley, C. Computer Vision Based Mango Fruit Grading System, in *International Conference on Innovative Engineering Technologies (ICIE'T 2014)*, Bangkok, Thiland, December, 2014, pp. 1–4.