engineering proceedings

MDPI

Proceeding Paper

Identification of Turmeric Rhizomes Using Image Processing and Machine Learning †

Shubhangi Patil [1,*] and Gouri Patil [2]

1 Department of Computer Science, Dhanaji Nana Mahavidyalaya, Faizpur 425503, India
2 Department of Computer Science, Bhusawal Arts, Science and P. O. Nahata Commerce College, Bhusawal 425201, India; gourimpatil@gmail.com
* Correspondence: shubhangipatil14@gmail.com
† Presented at the International Conference on Recent Advances on Science and Engineering, Dubai, United Arab Emirates, 4–5 October 2023.

Abstract: India is the world's leading producer and exporter of turmeric. Indian turmeric is known as the best in the world because of its natural medicinal properties. Different turmeric varieties have different amounts of nutritional value, which results in variations in their cost and quality. The quality assessment of turmeric aids in evaluating and determining its quality, and it helps to promote its marketing. Hence, the identification of turmeric cultivars is of great importance. But it requires manual inspection by human experts, generates subjective results and is time-consuming. Machine vision will provide a more accurate and faster way to identify different agricultural products and their varieties. This study presents an automated system to identify turmeric rhizome varieties by extracting morphological, color and texture features. The classification of different rhizome types is carried out by using image processing techniques followed by K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Naïve Bayes (NB), Random Forest (RF) and Linear Discriminant Analysis (LDA) classifiers. The proposed work shows promising results for the identification of turmeric rhizome varieties.

Keywords: turmeric rhizome; segmentation; feature extraction; GLCM; K-Nearest Neighbor

Citation: Patil, S.; Patil, G. Identification of Turmeric Rhizomes Using Image Processing and Machine Learning. *Eng. Proc.* **2023**, *59*, 34. https://doi.org/10.3390/engproc2023059034

Academic Editors: Nithesh Naik, Rajiv Selvam, Pavan Hiremath, Suhas Kowshik CS and Ritesh Ramakrishna Bhat

Published: 12 December 2023

1. Introduction

The automated identification and classification of agricultural products is an important process in the area of agricultural automation [1]. India has produced and supplied a wide variety of premium-quality spices to the entire world since ancient times, and hence, it is known as the 'Spice Bowl of the World' [2]. India contributes 75% of the production of whole spices in the world [3,4]. Turmeric is one of the important cash crops of India, and is immensely tolerant to a variety of environmental changes, including high temperatures and moderate drought. Turmeric cultivation is extremely profitable to farmers as they can obtain a high yield by selecting pure-variety and good-quality seeds. But visual inspection for variety identification and quality assessment needs human experts, and is a time-consuming process with subjective results [4]. Hence, it requires automation in turmeric rhizome identification that can be used during harvest for varietal identification and during post-harvest for quality assessment [5]. The challenges for the automated identification of turmeric rhizomes are the similarity in color, shape and variation among the same category of turmeric, and the availability of large datasets, which is a major issue.

An appropriate combination of various handcrafted features and the selection of appropriate machine learning classifiers can give precise and accurate classification results for small datasets [6,7]. This research work presents a performance evaluation of different machine learning classifiers, including K-Nearest Neighbor, Support Vector Machine, Naïve Bayes, Linear Discriminant Analysis and Random Forest classifiers, for analyzing the extracted color, texture and morphological features. We selected some RGB and LAB features after analyzing the results of the classifiers for all extracted features [8]. These selected features were applied for the classification of five different turmeric varieties, including Phule Swarupa, Ambe Halad, Kullu Local, Pitambar, and R-Sonia. The proposed system can be implemented for identifying the variety of turmeric at the time of cultivation for higher production with the pure variety, as well as on the market to obtain economic benefits for dealers and farmers.

2. Related Work

Machine vision has a wide area of application in the agricultural domain [9]. It has been used for many purposes like crop detection, yield prediction, the recognition of fruits and vegetables and their quality analysis, disease identification, the varietal identification of different agricultural products, and many more [10]. A. Aznan et al. presented an approach for the classification of fifteen different types of commercial rice grain samples. The morpho-colorimetric features were extracted from the rice sample images and classified by an ANN model with ten hundred neurons and by using the Bayesian Regularization algorithm. The highest accuracy obtained by this model was 91.6% [11]. S. Mawaddah et al. suggested a way to classify rhizome images by using a Support Vector Machine classifier. Three different types of rhizomes, turmeric, ginger and galangal, were identified based on their extracted shape and texture features. The texture feature and combination of shape and texture features demonstrated a good accuracy rate with the SVM classifier [12]. K. Sarode et al. used a Graphical User Interface to calculate the different values for the texture feature of given images. This work reduced the overhead of calculation and analyzed the features of the images using the GLCM approach [13]. M. Mohd Ali et al. worked on infrared thermal images to categorize three different cultivars of pineapple. Principal Component Analysis is used to reduce the number of features out of the fourteen different extracted features from the thermal images. In this work, a number of different machine learning algorithms, such as KNN, SVM, decision tree, LDA and Naïve Bayes, were compared to obtain the highest classification accuracy [14]. E. Ropelewska et al. proposed a model to distinguish three different types of plum cultivars. In this study, 2200 texture features were extracted for each individual color channel of the LAB, UVS, RGB and XYZ colorspace. The Kstar classifier provides high performance for the recognition of plum cultivars based on the selected texture features [15].

3. Methodology

An automated turmeric rhizome identification system is proposed to identify turmeric rhizomes in an efficient and effective way. This system provides a more accurate and faster way for turmeric rhizome identification than using traditional manual identification.

According to the proposed architecture of the system, the workflow steps are listed below:

Step 1: Acquire the rhizome images from fields or agriculture research centers to form the dataset.

Step 2: Read each image from the dataset.

Step 3: Preprocess the original image by resizing, denoising, and performing colorspace conversion.

Step 4: Segment the image using the threshold technique.

Step 5: Extract the color and texture features of the rhizome.

Step 6: Train the model using the different machine learning classifiers.

Step 7: Classify the testing set using the trained model.

The following Figure 1 shows the architecture of the proposed turmeric identification system.

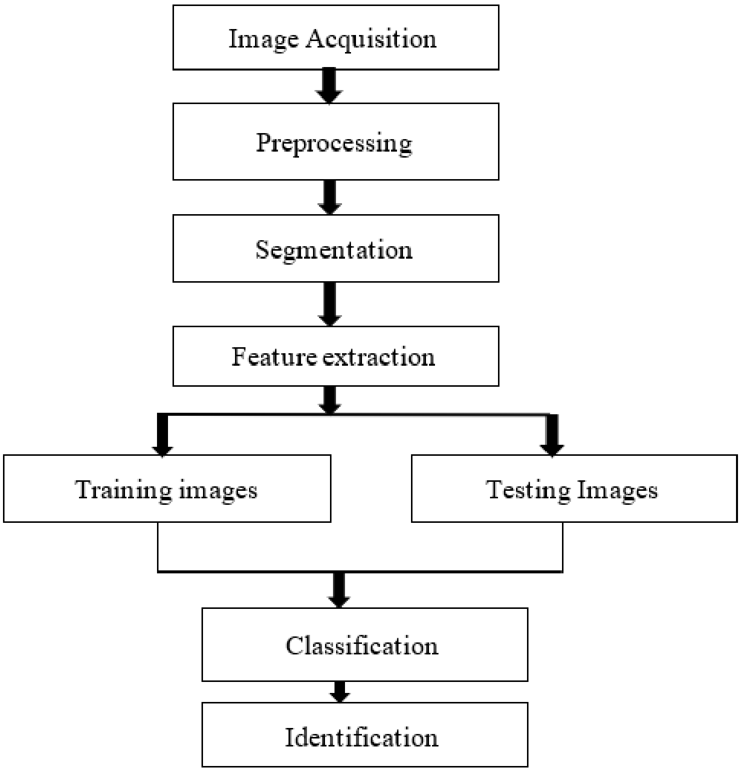


Figure 1. Proposed turmeric rhizome identification system's architecture.

3.1. Image Acquisition

The images were acquired to create a dataset of five different varieties of turmeric rhizome by using a high-resolution digital camera. All the images were captured in natural sunlight and while maintaining a constant distance between the camera and the turmeric rhizome. The samples of different turmeric variety rhizomes were collected from the research center as well as from fields to create a dataset. The dataset contains a total of 1350 images of five different turmeric rhizome varieties, including Phule Swarupa (295), Ambe Halad (285), Kullu Local (200), Pitambar (280) and R-Sonia (290). The images were acquired during two different turmeric cultivation seasons. The sample images of the five varieties of turmeric rhizome are shown in Figure 2.

Figure 2. Five different categories of turmeric rhizome.

3.2. Image Preprocessing

Image pre-processing improves the quality of an image to better analyze it. In preprocessing, all the acquired images are first annotated manually, and then, resized to 150 × 300. After that, a median filter is applied to remove unwanted noise from the image. To study the variations in the color of the turmeric variety, the images are then converted into different colorspaces.

3.3. Segmentation

Image segmentation is the process of extracting objects or regions of the image used to obtain some useful information from it [8]. The segmentation phase is most important for extracting the rhizome from the image. Three steps are involved in the segmentation process here. Firstly, the image is converted into the HSI colorspace, and a hue channel image is selected for further processing as it separates the background and the rhizome region. After that, by using Otsu thresholding, a binary image is obtained for the extraction of morphological features. The third step involves the use of a canny edge detection operator and morphological operation to obtain the boundary of the turmeric rhizome. The result of the segmentation process is shown in Figure 3.

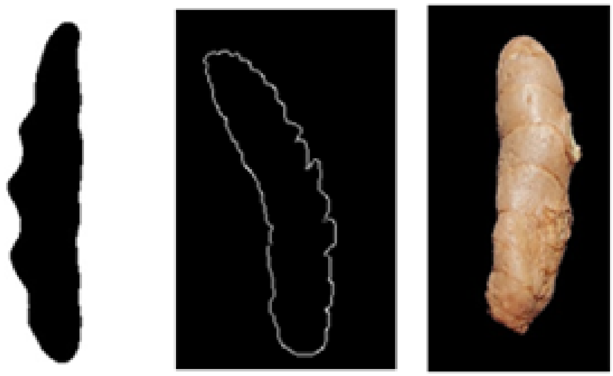


Figure 3. Segmented images.

3.4. Feature Extraction

In the field of computer vision and machine learning, feature extraction plays a crucial role in the analysis of an object in an image. It is observed from the literature that the morphological, color and texture features of turmeric play an important role in varietal identification and quality analysis, as every turmeric variety's rhizome has a unique color, size and texture. Hence, a turmeric rhizome can be identified by using extracted features and can be classified according to its particular class. The color, texture and morphological features were extracted by using the most popular feature extraction techniques. A total of 48 features, including 6 morphological, 15 texture, and 27 color features, were extracted.

3.4.1. Color

Color is the most important visual feature to identify the turmeric variety, as each of them has a slight difference in color [15]. RGB color histograms of each of the five varieties' images are shown in Figure 4. The image color is represented using a different color model. RGB, LAB and HSI are frequently used models that provide unique images that correspond to each colorspace. Hence, 27 color features of the turmeric image area were extracted, like mean, range and variance, for the separate color channels of RGB, HSI and LAB images. The RGB color histograms of the five turmeric varieties are shown in Figure 4.

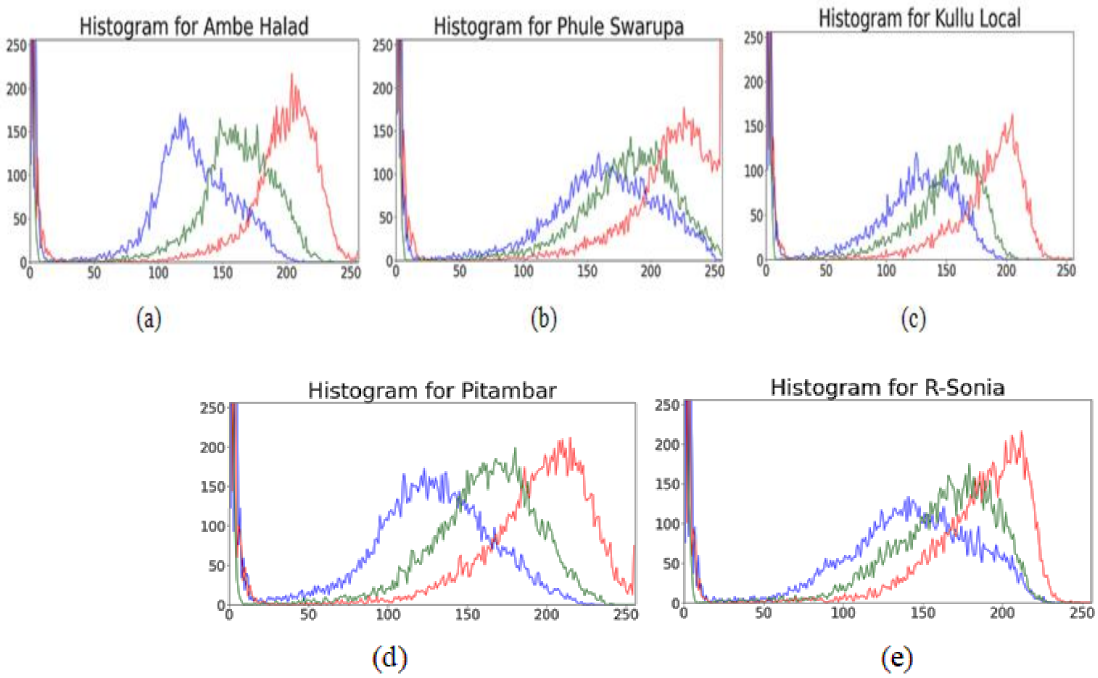


Figure 4. RGB color histogram of (**a**) Ambe Halad, (**b**) Phule Swarupa, (**c**) Kullu Local, (**d**) Pitambar and (**e**) R-Sonia varieties.

3.4.2. Morphology

Each turmeric variety has rhizomes of different sizes and shapes. Hence, morphological features are also important in their identification. A number of morphological features, including, area, width, height, perimeter, circularity and aspect ratio, were extracted [3,4]. The feature descriptions are as follows:

1. Area: the number of pixels in the region of the rhizome.
2. Width: width of a minimum rectangle enclosing the rhizome.
3. Height: height of a minimum rectangle enclosing the rhizome.
4. Perimeter: total number of boundary pixels.
5. Circularity: 4pi(Area)/(Perimeter^2)
6. Aspect ratio: Width * Length/Area

3.4.3. Texture

The texture feature represents the visual patterns of the image; it is the best component to study when assessing structural patterns from image surfaces. Grayscale images are used to calculate the GLCM features [16]. They enable the computation of how recurrently a pixel and its contiguous pixel arise vertically, horizontally and diagonally [13]. The GLCM approach was used to obtain five statistical features, including homogeneity [13], correlation [17], energy [17], contrast [13] and dissimilarity [3].

$$\text{Homogeneity}: \sum_{i,j=1}^{N} \frac{Pi,j}{(1+(i-j)2)} \tag{1}$$

$$\text{Correlation}: \frac{\sum_{i,j=1}^{N}(i-x^{-})(j-y^{-})P(i,j)}{\sigma_x \sigma_y} \tag{2}$$

$$\text{Energy}: \sum_{i,j=1}^{N} P(i,j)^2 \tag{3}$$

$$\text{Contrast}: \sum_{i,j=1}^{N} Pi,j(i-j)^2 \tag{4}$$

$$\text{Dissimilarity}: \sum_{i,j=1}^{N} Pi,j\,|i-j| \tag{5}$$

4. Classification

Classifier selection is an important step because the same set of features may produce different results for different classification techniques. Hence, through the comparative study of different machine learning classifiers used by many researchers to work on similar types of problems, the outperforming classifiers are selected in this work. The five different varieties of turmeric rhizome are considered for classification. Firstly, the dataset was divided into training and testing subsets with an 80% and 20% split. The classification was performed by using KNN, SVM, Naïve Bayes, Random Forest and LDA classifiers based on the extracted color, texture and morphological features. The classification accuracy was also evaluated based on some selected RGB and LAB features after analyzing the classification results of all the extracted features. The proposed classifier was tested with visually similar turmeric rhizome varieties. The performances of classifiers were evaluated with a confusion matrix and an ROC curve [18,19]. According to the confusion matrix, all five varieties of turmeric rhizome were classified with the highest accuracy by using the SVM classifier.

5. Results

The classification of five different categories of turmeric rhizome was performed by using KNN, SVM, Naïve Bayes, Random Forest and LDA classifiers based on the extracted features as shown in Figure 5a. The highest accuracy obtained when RGB features were used is 92.24% when using the SVM classifier. The LDA classifier shows 89.01%, 88.4%, 81.1% and 67.7% accuracy based on LAB, HSI and texture and morphological features, respectively as shown in Table 1a. The classification accuracy based on selected RGB and LAB features are presented in Figure 5b. It is found that SVM with the selected features gives better classification result as shown in Table 1b. The confusion matrix and the AUC-ROC curves of SVM classifier with selected features is shown in Figure 6. The classification accuracy with selected features is tested for the visually similar turmeric varieties like Pitambar and R-Sonia as well as Phule swarupa and Kullu Local as shown in Table 2. It can be seen from Figure 7a that the LDA is able to provide better accuracy for the classification of Pitambar and R-Sonia turmeric varieties, and Random Forest classifier is able to yield better classification accuracy for Phule Swarupa and Kullu Local varieties as presented in Figure 7b.

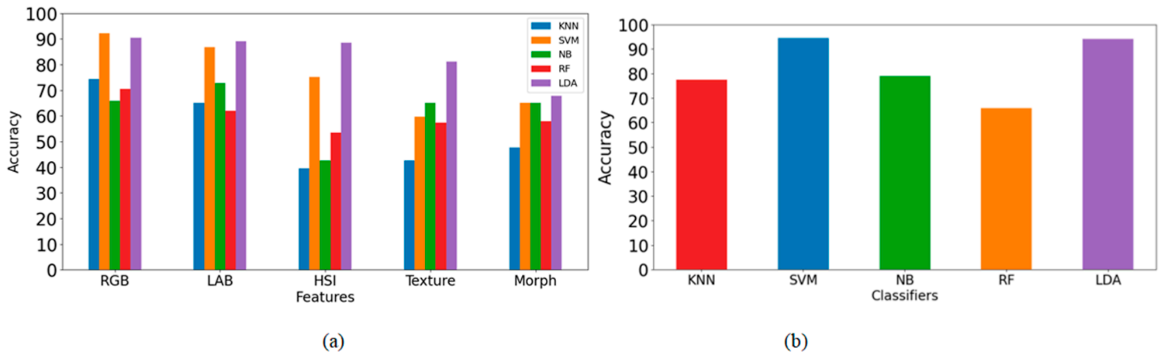


Figure 5. Performance of different classifiers on five varieties of turmeric rhizome data (**a**) using all the extracted features and (**b**) using selected features.

Table 1. Performance of different classifiers on five different varieties of turmeric rhizome data (**a**) using all the extracted features and (**b**) using selected features.

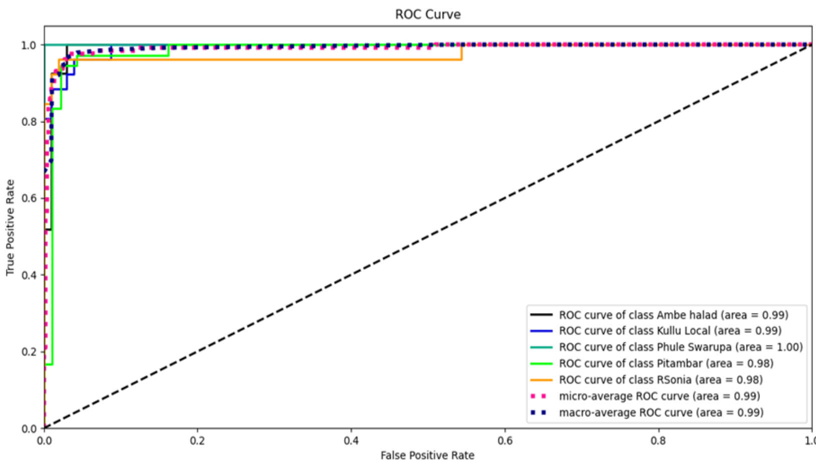
Classifier	Accuracy				
	RGB	LAB	HSI	Texture	Morph
KNN	74.41	65.11	39.53	42.63	47.61
SVM	92.24	86.82	75.19	59.68	65.07
Naïve Bayes	65.89	72.86	42.63	65.11	65.07
Random Forest	70.54	62.01	53.48	57.36	57.93
LDA	90.5	89.01	88.4	81.1	67.7

(**a**)

Classifier	Accuracy
KNN	77.51
SVM	94.57
Naïve Bayes	79.06
Random Forest	65.89
LDA	94.1

(**b**)

(**a**)



(**b**)

Figure 6. (**a**) Confusion matrix and (**b**) AUC-ROC curves of SVM classifier for the turmeric rhizome classification.

Table 2. Accuracy of classifiers when applied to Pitambar and R-Sonia, and Phule Swarupa and Kullu Local, with selected features.

Classifier	Accuracy for Pitambar and R-Sonia	Accuracy for Phule Swarupa and Kullu Local
KNN	90.32	93.16
SVM	96.77	93.18
NB	91.93	95.45
RF	96.77	97.72
LDA	96.9	95.9

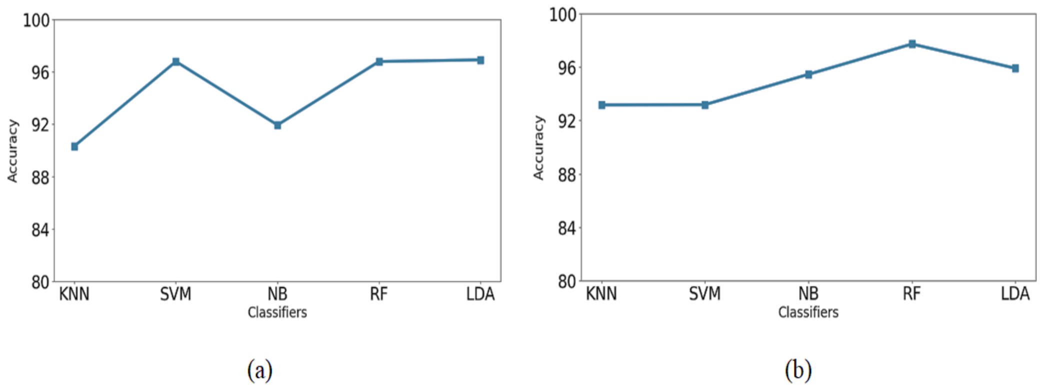


Figure 7. Accuracy of classifiers when applied to (**a**) Pitambar and R-Sonia, and (**b**) Phule Swarupa and Kullu Local, with selected features.

6. Conclusions

The identification of turmeric rhizome variety becomes a challenging task as more varieties are added to the dataset. The performance of different machine learning classifiers, KNN, SVM, Naïve Bayes, Random Forest and LDA, are evaluated to identify the five varieties of turmeric rhizome Phule Swarupa, Ambe Halad, Kullu Local, Pitambar and R-Sonia. The results of the classifiers are analyzed for all the extracted color, texture and morphological features [20]. It is observed that the SVM classifier outperforms in the classification of all five varieties by using RGB color features. After analyzing the classifier's result for all the extracted features, it is found that RGB and LAB features give better results; hence, the classifier's performance is again analyzed for a combination of some selected RGB and LAB features. As a result, it is found that for the selected features, SVM achieves the highest accuracy of 94.57%.

When the turmeric rhizome varieties are visually similar, then it becomes really difficult to identify them [21,22]. Hence, the accuracy of all these classifiers with the selected features for the classification of such visually similar varieties is evaluated. It is found that the highest accuracy of 96.9% is obtained by using LDA for the classification of Pitambar and R-Sonia and the highest accuracy of 97.72% is achieved for the classification of Kullu Local and Phule Swarupa by using Random Forest classifier. This system provides faster and more accurate identification of turmeric varieties, which is useful for farmers and merchants as well as customers. The present study can be extended to identify more turmeric cultivars by adding more turmeric varieties.

Author Contributions: Methodology, S.P.; supervision, G.P.; writing—original draft, S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sahidullah Md Nayan, N.M.; Morshed Md, S.; Hossain, M.M.; Islam, M.U. Date Fruit Classification with Machine Learning and Explainable Artificial Intelligence. *Int. J. Comput. Appl.* **2023**, *184*, 1–5. [CrossRef]
2. Kanase, V.; Khan, F. An Overview of Medicinal Value of Curcuma Species. *Asian J. Pharm. Clin. Res.* **2018**, *11*, 40. [CrossRef]
3. Aarthi, S.; Suresh, J.; Prasath, D. Morphological characterization of Indian turmeric (*Curcuma longa* L.) Genotypes using DUS descriptor. *J. Plant. Crops* **2018**, *46*, 173–179.
4. Kaur, A.; Saini, N.; Kaur, R.; Das, A. Automatic classification of turmeric rhizomes using the external morphological characteristics. In Proceedings of the 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 21–24 September 2016; pp. 1507–1510. [CrossRef]
5. Wan Nurazwin Syazwani, R.; Muhammad Asraf, H.; Megat Syahirul Amin, M.A.; Nur Dalila, K.A. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alex. Eng. J.* **2022**, *61*, 1265–1276. [CrossRef]
6. Hassan NM, H.; Ramadan Gamal Elshoky, B.; MMabrouk, A.M. Quality of performance evaluation of ten machine learning algorithms in classifying thirteen types of apple fruits. *Indones. J. Electr. Eng. Comput. Sci.* **2023**, *30*, 102. [CrossRef]
7. Behera, S.K.; Rath, A.K.; Mahapatra, A.; Sethy, P.K. Identification, classification & grading of fruits using machine learning & computer intelligence: A review. *J. Ambient Intell. Humaniz. Comput.* **2020**, *11*, 1–11. [CrossRef]
8. Venkateswaran, D.T.D. Fruits Recognition System Based on Color, Shape, Principal Component and Region Features. *Int. J. Res. Anal. Rev.* **2019**, *6*, 226–231.
9. Saxena, P.; Priya, K.; Goel, S.; Aggarwal, P.K.; Sinha, A.; Jain, P. Rice Varieties Classification using Machine Learning Algorithms. *J. Pharm. Negat. Results* **2022**, *13*, 3762–3772.
10. Yu, F.; Lu, T.; Xue, C. Deep Learning-Based Intelligent Apple Variety Classification System and Model Interpretability Analysis. *Foods* **2023**, *12*, 885. [CrossRef] [PubMed]
11. Aznan, A.; Gonzalez Viejo, C.; Pang, A.; Fuentes, S. Computer Vision and Machine Learning Analysis of Commercial Rice Grains: A Potential Digital Approach for Consumer Perception Studies. *Sensors* **2021**, *21*, 6354. [CrossRef] [PubMed]
12. Mawaddah, S.; Mufid, M.R.; Basofi, A.; Fiyanto, A.; Aditama, D.; Nurlaila, N. Rhizome Image Classification Using Support Vector Machine. In Proceedings of the International Conference on Applied Science and Technology on Social Science 2021 (iCAST-SS 2021), Samarinda, Indonesia, 21–23 October 2022. [CrossRef]
13. Sarode, K.; Savdekar, R.; Chaudhari, T. Texture feature Analysis of an image using Gray Level Co- Occurrence Matrix. *Int. J. Nov. Res. Dev.* **2022**, *7*, 5.
14. Mohd Ali, M.; Hashim, N.; Abd Aziz, S.; Lasekan, O. Characterisation of Pineapple Cultivars under Different Storage Conditions Using Infrared Thermal Imaging Coupled with Machine Learning Algorithms. *Agriculture* **2022**, *12*, 1013. [CrossRef]
15. Ropelewska, E.; Cai, X.; Zhang, Z.; Sabanci, K.; Aslan, M.F. Benchmarking Machine Learning Approaches to Evaluate the Cultivar Differentiation of Plum (*Prunus domestica* L.) Kernels. *Agriculture* **2022**, *12*, 285. [CrossRef]
16. Komal, K.; Sonia, D. GLCM Algorithm and SVM Classification Method for Orange Fruit Quality Assessment. *Int. J. Eng. Res.* **2019**, *8*, 7.
17. Jitanan, S.; Chimlek, P. Quality grading of soybean seeds using image analysis. *Int. J. Electr. Comput. Eng. IJECE* **2019**, *9*, 3495. [CrossRef]
18. Arwatchananukul, S.; Saengrayap, R.; Chaiwong, S.; Aunsri, N. Fast and Efficient Cavendish Banana Grade Classification using Random Forest Classifier with Synthetic Minority Oversampling Technique. *IAENG Int. J. Comput. Sci.* **2022**, *49*, 46–54.
19. Kurade, C.; Meenu, M.; Kalra, S.; Miglani, A.; Neelapu, B.C.; Yu, Y.; Ramaswamy, H.S. An Automated Image Processing Module for Quality Evaluation of Milled Rice. *Foods* **2023**, *12*, 1273. [CrossRef] [PubMed]
20. Panigrahi, J.; Pattnaik, P.; Dash, B.B.; Ranjan Dash, S. Rice Quality Prediction using Computer Vision. In Proceedings of the 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 13–14 March 2020; pp. 1–5. [CrossRef]
21. Ghazal, S.; Qureshi, W.S.; Khan, U.S.; Iqbal, J.; Rashid, N.; Tiwana, M.I. Analysis of visual features and classifiers for Fruit classification problem. *Comput. Electron. Agric.* **2021**, *187*, 106267. [CrossRef]
22. Garcia JA, A.; Arboleda, E.R.; Galas, E.M. Identification of Visually Similar Vegetable Seeds Using Image Processing and Fuzzy Logic. *Int. J. Sci. Technol. Res.* **2020**, *9*, 5.