International Journal of Knowledge Based Computer Systems
10 (1) June 2022, 27-32
http://www.publishingindia.com/ijkbcs/

Classification of Spices using Machine Learning Techniques

Yukti Gupta[1*], Haneet Kour[2], Jatinder Manhas[3] and Vinod Sharma[4]
[1]M.Tech. Student, Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, India. Email: yuktigupta19@gmail.com
[2]PhD Research Scholar, Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, India. Email: haneetkour9@gmail.com
[3]Sr. Assistant Professor, Department of Computer Science and IT, Bhaderwah Campus, University of Jammu, Jammu and Kashmir, India. Email: manhas.jatinder@gmail.com
[4]Professor, Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, India. Email: vnodshrma@gmail.com
*Corresponding Author

Abstract: Machine learning (ML) has played a significant role in pattern recognition including fruits and vegetables classification. In this paper, comparative analysis of various ML techniques have been carried out for the identification of Spices. For the current work, ML techniques namely Naïve Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RF) and Support Vector Machine (SVM) have been undertaken. The main aim of the current study is to find out the most appropriate ML approach for Spices recognition. The experimental study has been performed on primary dataset of Spices. This dataset consists of 1000 images of five different Spices including clove, green cardamom, cinnamon, black pepper and curry leaf. The performance of the ML techniques have been analyzed on the basis of four parameters i.e. accuracy, precision, recall and f1-score. Out of five implemented ML models, best performance has been predicted by SVM approach with accuracy of 94.5%, precision of 95%, and recall of 94% with f1-score of 0.95.

Keywords: Decision tree, K-Nearest neighbor, Machine learning, Spices recognition, Support vector machine.

I. Introduction

India is referred to the Land of Spices and it is the world's largest manufacturer, consumer and exporter of Spices. It commands a formidable position in the world Spice trade. The total value of Spices produced in India is about 7000 crores annually. The country produces about 75 out of 109 types of Spices listed by the International Organization for Standardization (ISO) [1].

Every Spice has its own flavor and essence, and its addition or omission can genuinely make or destroy a dish. Spices include remarkable nutritional values and have numerous health benefits including antioxidant and anti-inflammatory properties, glucose-lowering effects, appetite control, regulates metabolism, aids in weight loss, improves memory and brain function. Spices are being used by several medical industries such as cosmetic, pharmaceutical and aromatic as perfumery [2].

It is quite difficult to distinguish between Spices as many Spices look similar. The recognition of Spices used in our daily basis food is gaining more importance in our daily life. Research on Spices recognition and classification is vital for several economic sectors, both for the wholesale and retail markets, as well as for the processing industries.

Consuming healthy and good quality fruits and vegetables are the utmost necessity of the purchaser. Hence, automation in food industries are developing nowadays because it is improbable for humans to manually audit the fruits and vegetables as it requires a large number of workers along with a lot of time and effort. In recent years, numerous machine learning (ML) algorithms have been applied with various different feature description methods for fruits and vegetables classification in several real-world applications [3].

In this study, comparative analysis of several ML techniques including Naive Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Random Forest (RF) and SVM (Support Vector Machine) have been carried out for the recognition of Spices. The main motive of the current research work is to implement ML based system for the detection of Spices for food image researchers and dietician to make an appropriate analysis of nutrition and other type of health hazards. Classifying a particular kind of Spice will enable us to distinguish it from another kind. For the current study, five different classes of

Spices including clove, green cardamom, cinnamon, black pepper and curry leaf have been considered.

This paper is divided into five sections: Section II illustrates the literature related to the current work, Section III demonstrates materials and methodology adopted in the current work, Section IV shows experimental results and analysis, and Section V explains the conclusion and future scope.

II. Literature Review

In the recent past, ML techniques have been applied in agricultural domain. These approaches have been found to be effective for the recognition of fruits and vegetables.

In the year 2012, Zhang and Wu [4] used multi-class kernel support vector machine (kSVM) for fruit classification. The SVM approach with max-wins-voting strategy and gaussian radial basis kernel achieved the classification accuracy of 88.2%.

Next year, Sanjayan *et al.* (2013) [5] employed ML techniques for the evaluation of the nutritional value of ripened fruits of a commonly known medicinal plant Morinda tinctoria. The work targeted on the estimation of Ash content, protein, carbohydrate, vitamins and mineral content of Morinda tinctoria fruit. The difference in nutritional content between fresh and dried fruits had been also evaluated. The ash content for fresh and dry fruits are 4% and 1.6% respectively.

After two years, Surya and Priya (2015) [6] proposed a food image recognition system for calculating the calorie and nutrition values using SVM classifier. In this approach, the area, volume and size of the undertaken food item has been used to measure the calorie and nutrition precisely.

Femling *et al.* (2018) [7] depicted MobileNet based approach for recognizing fruit and vegetables within the retail market using photographs captured with a video camera connected to the system. The proposed system achieved maximum accuracy of 97%.

Dheir *et al.* (2019) [8] introduced a system for classifying nuts types. This study applied Convolutional Neural Network (CNN) algorithm to the dataset of five classes of Nuts that contains 2868 images. This method achieves excellent results with accuracy of 98%.

Bhattacharya and Mukherjee (2020) [9] implemented Neural Network based spices image classification system for Indian food images. The system is based on segmentation and recognition of spices images that identifies spice types. The dataset contains 4 types of spices containing a total of 80 spices images. The experimental study achieved the accuracy of 96.7%.

In the same year, Rojas-Aranda *et al.* (2020) [10] developed an image classification method based on lightweight CNN. A new dataset of images is designed belonging to three different classes of fruits, within or without plastic bags. Various input features have been added to the CNN architecture to improve classification accuracy. The experimental results predicted the 95% classification accuracy for fruits without plastic bag, and 93% for fruits inside plastic bag.

Next year, Shaikh *et al.* (2021) [11] applied faster Region-based Convolutional Neural Network (RCNN) model to detect and classify fruits as affected or unaffected based on their surface. The experimental study achieved accuracy of 60-75% for healthy apple, 60-70% for bad apple, 85-99% for healthy pear, 80-98% for bad pear, 80-97% for healthy banana and 70-80% for bad banana.

In the same year, Patil *et al.* (2021) [12] created grading and classifying techniques for dragon fruits with ML and Deep Learning (DL) algorithms i.e. SVM, CNN and Artificial Neural Network (ANN) based on a thorough review of algorithms available for fruit quality detection and classification using numerous characteristics of fruits and vegetables.

Zakeri *et al.* (2021) [13] presented a computer vision-based approach for grading jujube fruits using ML techniques. This method takes into account most of the important pricing factors and can be used to increase farmers' profits. They used various classifiers and training methods to get the best results, and with the help of DT, they were able to achieve a classification accuracy of 98.8%.

III. Materials and Methods

The goal of the current research is to identify Spices using ML techniques. This paper presents a comparative study of different ML techniques for the classification of Spices in order to find out the most suitable technique for Spices classification. The experimental study has been done on Python platform using *Spyder Interface* [14]. Fig. 1 represents the methodology for the current study.

A. Data Collection

The proposed study has been done on primary dataset as the domain of work is new and no dataset is available on the internet. For the current study, five different classes of Spices including *clove*, *green cardamom*, *cinnamon*, *black pepper* and *curry leaf* have been collected.

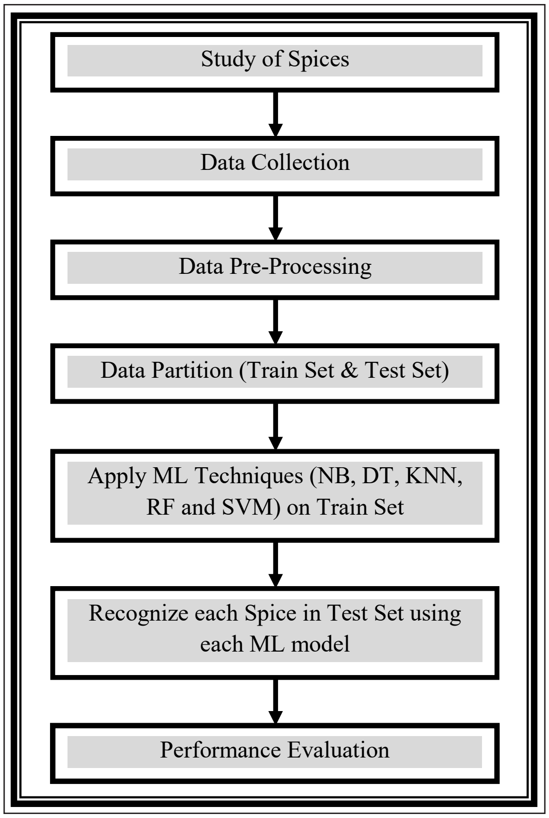

Fig. 1: Work Flow of Current Research

Each Spice image has been captured by using high quality camera and these images have been taken from different angles in order to introduce sufficient variations. White sheet of paper has been used as background of each image. The collected dataset consists of 1000 images, belonging to five different classes of undertaken Spices. There are 200 images of each Spice taken in this study. Fig. 2 depicts Sample of Spices in the collected dataset.

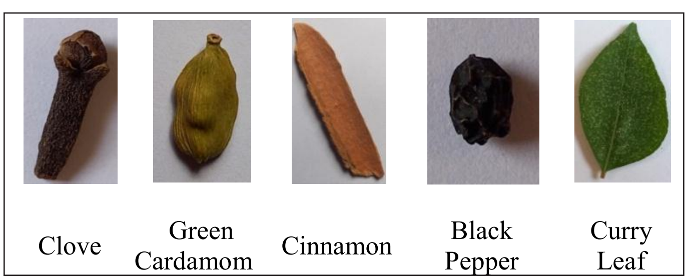

Fig. 2: Sample of Spices Dataset

B. Data Pre-Processing and Data Partition

The data pre-processing steps include cropping, renaming and resizing of the images to the standard dimensions, in order to reduce computational load. The collected images of Spices are in the RGB format and these images have been converted to gray-scale format.

After data pre-processing, data has been partitioned in the ratio 80:20 into two sets: *Train Set* and *Test Set*. Train Set consists of 800 images whereas Test Set contains 200 images.

C. Implementation using Machine Learning Techniques

After data partition, ML techniques have been implemented on Train Set to build the model for Spices recognition. For the current work, five ML techniques namely NB [15], DT [16], KNN [17], RF [18] and SVM [19] have been undertaken. For the implementation of each ML model, feature extraction approach i.e. Discrete Wavelet Transform (DWT) [20] has been applied. DWT has been implemented at Level-1 using *Haar* approach; which computes four coefficients i.e. cA (approximation coefficient), cH (horizontal coefficient), cV (vertical coefficient) and cD (diagonal coefficient). For each coefficient, five features i.e. contrast, energy, dissimilarity, homogeneity, and correlation have been computed. Thus, for each Spice image, total 20 features have been extracted and these features have been saved in *.csv* file. The final size of *.csv* file becomes 1000*21 since the dataset has 1000 images and 20 are the input extracted features and 01 is the output class label.

Each undertaken ML technique has been applied to build the Spices recognition model using Train Set. For each ML technique, the extracted features have been fed as input feature vector and the class label has been used as predictors.

In this study, NB algorithm has been predicted on the basis of the probability of Spices by applying Bayes theorem. It assumes that the occurrence of certain feature of Spice image is independent of the occurrence of other features.

In DT model, decision nodes are used to make any decision and leaf nodes are the output of these decisions. Decisions have been taken on the basis of features of the Spices dataset using criterion parameter with entropy value as a metric for impurity.

KNN model has been implemented using 'feature similarity' for predicting the values of new data points and the value of k = 7 with *Euclidean* distance metric has been taken.

For the current work, RF technique has been constructed using 20 small decision trees known as estimators on data samples with max depth 5 and gets the prediction from each of them. By means of voting best solution has been selected.

SVM approach has created best decision boundary (hyperplane) that segregated the 5 different classes of Spices in multidimensional space using ECOC (error correcting code output) with kernel function.

IV. Result and Analysis

The main objective of this study is to develop a Spices Recognition framework based on ML techniques in order to recognize different Spices. To test the performance of the proposed system, experiments were performed on a dataset of 1000 images of five Spices that includes: clove, green cardamom, cinnamon, black pepper and curry leaf. The training dataset contains 800 images whereas the testing dataset contains 200 images. DWT feature extractor has been employed to compute the texture features: correlation, energy, contrast, homogeneity, and dissimilarity.

All the undertaken ML approaches have been evaluated based on four performance measures i.e. accuracy, recall, precision and f1-score. Accuracy depicts about how accurately these ML techniques can recognize the given spice. In addition to it, it is required to avoid the misclassification of the given spice by undertaken ML approach. To deal with such issues, recall and precision parameters have been considered. Recall tells about *false negative* whereas precision tells about *false positive*. It is needed that the undertaken ML technique should predict lowest false positive and false negative. If one technique depicts *high false positive and low false negative*; and the other one predicts *low false positive and high false negative*, then f1-score is employed to evaluate the model performance as it computes the harmonic mean of precision and recall.

From the experimental results, it has been found out that all the ML techniques can be applied for the Spices classification. Fig. 3 presents experiment results for NB. The experiment results for DT are shown in Fig. 4. Fig. 5 represents the experiment results for KNN. Fig. 6 depicts experiment results for RF. Fig. 7 shows the experiment results for SVM.

```
Confusion Matrix:
[[25  1  0 12  0]
 [ 4 33  1  1  0]
 [ 1  0 31  0  8]
 [ 0  2  0 38  0]
 [ 6  0  3  0 34]]
Classification Report:
              precision    recall  f1-score   support

         0.0       0.69      0.66      0.68        38
         1.0       0.92      0.85      0.88        39
         2.0       0.89      0.78      0.83        40
         3.0       0.75      0.95      0.84        40
         4.0       0.81      0.79      0.80        43

    accuracy                           0.81       200
   macro avg       0.81      0.80      0.80       200
weighted avg       0.81      0.81      0.80       200

Accuracy: 0.805
```

Fig. 3: Experimental Results for NB

```
Confusion Matrix:
[[30  1  0  1  1]
 [ 1 43  2  2  0]
 [ 1  1 30  1  4]
 [ 4  1  0 33  0]
 [ 0  0  5  0 39]]
Classification Report:
              precision    recall  f1-score   support

         0.0       0.83      0.91      0.87        33
         1.0       0.93      0.90      0.91        48
         2.0       0.81      0.81      0.81        37
         3.0       0.89      0.87      0.88        38
         4.0       0.89      0.89      0.89        44

    accuracy                           0.88       200
   macro avg       0.87      0.87      0.87       200
weighted avg       0.88      0.88      0.88       200

Accuracy: 0.875
```

Fig. 4: Experimental Results for DT

```
Confusion Matrix:
[[33  0  0  6  0]
 [ 3 34  0  0  0]
 [ 0  1 32  0  7]
 [ 1  0  0 40  0]
 [ 0  0  0  0 43]]
Classification Report:
              precision    recall  f1-score   support

         0.0       0.89      0.85      0.87        39
         1.0       0.97      0.92      0.94        37
         2.0       1.00      0.80      0.89        40
         3.0       0.87      0.98      0.92        41
         4.0       0.86      1.00      0.92        43

    accuracy                           0.91       200
   macro avg       0.92      0.91      0.91       200
weighted avg       0.92      0.91      0.91       200

Accuracy: 0.91
```

Fig. 5: Experimental Results for KNN

```
Confusion Matrix:
[[22  1  0  3  0]
 [ 1 47  0  1  0]
 [ 0  2 38  0  1]
 [ 4  0  0 40  0]
 [ 0  0  1  0 39]]
Classification Report:
              precision    recall  f1-score   support

         0.0       0.81      0.85      0.83        26
         1.0       0.94      0.96      0.95        49
         2.0       0.97      0.93      0.95        41
         3.0       0.91      0.91      0.91        44
         4.0       0.97      0.97      0.97        40

    accuracy                           0.93       200
   macro avg       0.92      0.92      0.92       200
weighted avg       0.93      0.93      0.93       200

Accuracy: 0.93
```

Fig. 6: Experimental Results for RF

```
Confusion Matrix:
[[23  0  0  3  0]
 [ 1 47  0  1  0]
 [ 1  1 39  0  0]
 [ 4  0  0 40  0]
 [ 0  0  0  0 40]]
Classification Report:
              precision    recall  f1-score   support

         0.0       0.79      0.88      0.84        26
         1.0       0.98      0.96      0.97        49
         2.0       1.00      0.95      0.97        41
         3.0       0.91      0.91      0.91        44
         4.0       1.00      1.00      1.00        40

    accuracy                           0.94       200
   macro avg       0.94      0.94      0.94       200
weighted avg       0.95      0.94      0.95       200

Accuracy: 0.945
```

Fig. 7: Experimental Results for SVM

Table I presents experimental results of five ML techniques for Spices classification. Fig. 8 shows the comparative analysis of all the undertaken ML techniques. NB predicted lowest value for accuracy of 80.5%, precision of 81%, recall and f1-score of 81% and 0.80. DT model achieved an accuracy of 87.5%, precision of 88%, recall of 88% and f1-score of 0.88. KNN model attained an accuracy of 91%, precision of 92%, recall of 91% and f1-score of 0.91. RF model obtained an accuracy of 93%, precision of 93% and recall of 93% with f1-score of 0.93. SVM model showed best results as it achieved higher value for accuracy, precision, recall and f1-score being 94.5%, 95%, 94% and 0.95 respectively.

Table I: Experimental Results of Five ML Techniques for Spices Recognition

Traditional ML Technique	Accuracy	Precision	Recall	F1-Score
NB	80.5%	81%	81%	0.80
DT	87.5%	88%	88%	0.88
KNN	91%	92%	91%	0.91
RF	93%	93%	93%	0.93
SVM	94.5%	95%	94%	0.95

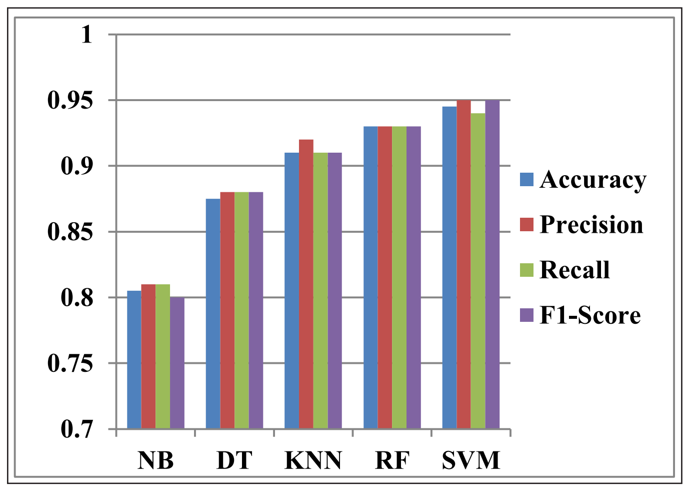

Fig. 8: Comparative Analysis of All Undertaken ML Techniques

V. Conclusion and Future Scope

From experimental results, it has been concluded that ML approaches can be used for the detection of Spices. The main purpose of the current work is to implement various ML techniques for the recognition of different Spices in order to reveal out the better approach for Spices classification. Out of all the ML techniques taken in this study, SVM has been found to be more effective technique for the Spices recognition with an accuracy of 94.5%, precision of 95% and recall of 94% with f1-score of 0.95. This study is limited to only five Spices. Future work will involve recognition of other different varieties of Spices using combined features more efficiently.

References

[1] K. Ponmozhi, and R. Jose, “A review on spices classification and recognition,” *International Journal of Mathematics and Computer*, vol. 12, no. 1, pp. 25-32, 2022.

[2] A. Jiang, “Health benefits of culinary herbs and spices,” *Journal of AOAC International*, vol. 102, no. 2, pp. 395-411, 2019, doi: 10.5740/jaoacint.18-0418.

[3] Machine Learning. [Online]. Available: https://en.m.wikipedia.org/wiki/Machine_learning#:~:text=Machine%20learning%20(ML)%20is%20the,a%20part%20of%20artificial%20intelligence

[4] Y.-D. Zhang, and L. Wu, “Classification of fruits using computer vision and a multiclass support vector machine,” *Sensors*, vol. 12, no. 9, pp. 12489-12505, 2012, doi: 10.3390/s120912489.

[5] V. Anuradha, A. Praveena, and K. Sanjayan, “Nutritive analysis of fresh and dry fruits of Morinda tinctoria,” *International Journal of Current Microbiology and Applied Sciences*, vol. 2, no. 3, pp. 65-74, 2013.

[6] R. Surya, and S. S. Priya, “Food image recognition using SVM classifier for measuring calorie and nutrition values,” *International Journal of Scientific & Engineering Research*, vol. 6, no. 4, pp. 324-328, 2015.

[7] F. Femling, A. Olsson, and F. Fernandez, “Fruit and vegetable identification using machine learning for retail applications,” *International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2018, doi: 10.1109/SITIS.2018.00013.

[8] I. M. Dheir, A. S. A. Mettleq, and A. A. Elsharif, “Nuts types classification using deep learning,” *International Journal of Academic Information Systems Research (IJAISR)*, vol. 3, no. 12, pp. 12-17, 2019.

[9] S. Bhattacharya, and J. Mukherjee, “Recognisation of Indian spices based on the combination of features and comparison using neural network,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 7, no. 6, pp. 2418-2423, 2020.

[10] J. Aranda, J. Varela, J. Tello, and G. Ramirez, “Fruit classification for retail stores using deep learning,” *in K.* F. Mora, J. A. Marín, J. Cerda, J. Carrasco-Ochoa, J. Martínez-Trinidad, and J. Olvera-López (Eds.) Pattern Recognition, MCPR 2020, *Lecture Notes in Computer Science,* vol. 12088, Springer Nature Switzerland AG, , 2020, doi: 10.1007/978-3-030-49076-8_1.

[11] H. Shaikh, Y. Wagh, S. Shinde, and S. Patil, “Classification of affected fruits using machine learning,” *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 3, 2021.

[12] P. Patil, S. Lande, V. Nagalkar, S. Nikam, and G. Wakchaure, “Grading and sorting technique of dragon fruits using machine learning algorithms,” *Journal of Agriculture and Food Research,* vol. 4, no. 3, 2021. [Online]. Available: https://doi.org/10.1016/j.jafr.2021.100118

[13] A. Zakeri, R. Hedayati, M. Khedmati, and M. Taghipour-Gorjikolaie, “Classification of jujube fruit based on several pricing factors using machine learning methods,” 2021. [Online]. Available: https://doi.org/10.48550/arXiv.2111.00112

[14] Spyder Interface. [Online]. Available: https://anaconda.org/anaconda/spyder

[15] C. Prasad, “An efficient heart disease detection system utilizing Naïve Bayes classification,” *Journal of Applied Information Science*, vol. 9, no. 1, pp. 22-25, 2021.

[16] B. Poornima, “An improved decision tree classification for breast cancer detection with optimal parameters,” *Journal of Applied Information Science*, vol. 9, no. 1, pp. 19-21, 2021.

[17] G. Guo, H. Wang, D. Bell, Y. Bi, and K. Greer, “KNN model-based approach in classification,” *OTM Confederated International Conferences*, 2003, doi: 10.1007/978-3-540-39964-3_62.

[18] Gokul S., and N. Banu P. K., “Analysis of phishing detection using logistic regression and random forest,” *Journal of Applied Information Science*, vol. 8, no. 1&2, pp. 7-13, 2020.

[19] A. MS, and Manjesh R., “Fruit recognition using SVM technique,” *International Journal of Science and Research (IJSR)*, vol. 9, no. 2, pp. 1210-1214, 2020, doi: 10.21275/SR20212155833.

[20] S. Thakral, and P. Manhas, “Image processing by using different types of discrete wavelet transform,” *Advanced Informatics for Computing Research*, 2018.