TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 22, No. 4, August 2024, pp. 887~896
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v22i4.25812

Automated classification of diseased cauliflower: a feature-driven machine learning approach

Mala Rani Barman[1], Al Amin Biswas[2], Marjia Sultana[3], Aditya Rajbongshi[4], Md. Sabab Zulfiker[2], Tasnim Tabassum[5]
[1]Department of Computer Science and Engineering, Sheikh Hasina University, Netrokona, Bangladesh
[2]Department of Computer Science and Engineering, Bangabandhu Sheikh Mujibur Rahman University, Kishoreganj, Bangladesh
[3]Department of Computer Science and Engineering, Begum Rokeya University, Rangpur, Bangladesh
[4]Department of Educational Technology, Bangabandhu Sheikh Mujibur Rahman Digital University, Dhaka, Bangladesh
[5]Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

Article Info

Article history:

Received Oct 21, 2023
Revised Nov 8, 2023
Accepted Dec 27, 2023

Keywords:

Analysis of variance
Bagging classifier
Cauliflower
Image processing
Machine learning

ABSTRACT

Cauliflower is a popular winter crop in Bangladesh. However, cauliflower plants are vulnerable to several diseases that can reduce the cauliflowers' productivity and degrade their quality. The manual monitoring of these diseases takes a lot of effort and time. Therefore, automatic classification of the diseased cauliflower through computer vision techniques is essential. This study has retrieved ten different statistical and gray-level co-occurrence matrix (GLCM)-based features from the cauliflower image dataset by implementing a variety of image processing techniques. Afterwards, the SelectKBest method with the analysis of variance f-value (ANOVA F-value) has been used to identify the most important attributes for classification of the diseased cauliflower. Based on the ANOVA F-value, the top N ($5 \leq N \leq 9$) most dominant attributes is used to train and test five machine learning (ML) models for classification of diseased cauliflower. Finally, different performance metrics have been used for evaluating the effectiveness of the employed ML models. The bagging classifier achieved the highest accuracy of 82.35%. Moreover, this model has outperformed other ML classifiers in terms of other performance metrics also.

Corresponding Author:

Aditya Rajbongshi
Department of Educational Technology, Bangabandhu Sheikh Mujibur Rahman Digital University
Dhaka, Bangladesh
Email: aditya0001@bdu.ac.bd

1. INTRODUCTION

Agriculture significantly contribute in the economic advancement of Bangladesh, which is fuelled by a variety of factors [1]. According to the Bangladesh Bureau of statistics, agriculture products, which include crops, livestock, fisheries, and forest products, made up 12.5% of the country's gross domestic product (GDP) and employed around 40% of the population [2]. The majority of agriculture in Bangladesh is based on traditional subsistence farming. Therefore, with limited resources and arable land, the development of new technologies can boost productivity and income levels.

The realm of machine learning (ML) and computer vision is experiencing a substantial surge in agricultural research to ensure a better production by introducing automatic disease classification, automatic measurement of seed, and soil quality. At present, computer vision technology is employed for various agricultural activities, including the classification of plant diseases, predicting crop outputs, species identification, irrigation management, and soil preservation. Developed nations have already made substantial use of modern agricultural technology, but it is unfortunate that the new technology has received little attention in Bangladesh's agricultural sector [3], [4].

Bangladesh produces a variety of agricultural products, such as rice, fruits, vegetables, and wheat. Among the agricultural goods produced, vegetables and fruit play a key role in maintaining a healthy body. A study examining the dietary habits of over 469,000 participants found that consuming more fruits and vegetables is linked to a decreased risk of death from cardiovascular disease, with a 4% average decrease in risk for every additional serving per day [5].

Bangladesh is naturally blessed with highly fertile land where a variety of crops such as food grains, cash crops, vegetables, and fruits are easily grown. Cauliflower is one of the most popular winter vegetable crops in Bangladesh and in the other countries of the world. It is cultivated in almost all areas of Bangladesh during the winter season, but now cauliflower is cultivated throughout the year due to its popularity and nutritional value. Although, the white flower-like part of cauliflower is usually used for cooking purposes, the stalk and surrounding thick, green leaves are also used in vegetable broths or can be fed as fodder for animals [6], [7]. Cauliflower is a low-fat vegetable that is abundant in folate, water, vitamins A and C which can be very helpful in meeting the high nutritional needs of the human body [8]. Cauliflower, like others in the cabbage family, contains a variety of phytochemicals, which can protect against cardiovascular disease and cancer [9]. It is also helpful for controlling blood cholesterol levels and maintaining a healthy immune system. Considering the health benefits of cauliflower, attention should be paid to growing cauliflower in large quantities to meet the demand. In Bangladesh, the cultivation area for cauliflower spans approximately 9,400 hectares, with an annual production of roughly 73,000 metric tons [10]. The current production falls short of meeting the nutritional requirements of the country’s sizable populace. There are several factors in the low production of cauliflower in Bangladesh such as seed quality, lack of irrigation, improper fertilizers and pesticides uses and disease attack. Moreover, the presence of diseases in cauliflower leaves and flowers negatively affects the yield and quality of the crop.

Many farmers in Bangladesh do not have enough knowledge or resources to correctly classify diseased cauliflower in time and because of this, farmers often use the wrong pesticides to treat their crops, which are more harmful to the crop as well as the soil. Early identification of diseased cauliflower can aid in preventing the spread of the disease to other fresh plants. Moreover, precise detection of the diseased cauliflower can help minimize crop losses by ensuring timely interventions. Conventional methods for categorizing diseased cauliflower are sluggish, expensive, demanding of labor, and time-intensive, especially in extensive crop fields. Additionally, farmers in remote rural areas of Bangladesh may be required to undertake long travels to seek advice from agricultural experts. Agriculture professionals, on the other hand, are unable to visit the local area at the proper time. Therefore, the development of an automated classification system to separate the diseased cauliflower would greatly help the farmers of rural areas to promptly detection of the diseased cauliflower and ensure a greater cauliflower yield.

Considering above situation, this paper focuses on developing a computer vision based automated classification system that can accurately classify the diseased cauliflower. The implemented system in this study utilizes image processing techniques to analyze images of the cauliflower crops, extract image features and lastly utilize machine learning models for classification of diseased cauliflower with high accuracy. Though we have applied the proposed approaches on cauliflower dataset only, this approach will be also workable for other vegetables addressing broader agricultural challenges in Bangladesh. The contribution of the study mainly includes the followings: i) collection of cauliflower crop images dataset for the study area of Bangladesh and ii) finding out the most important features for classifying the diseased cauliflower and ML model with promising accuracy.

The subsequent segments are arranged as follows: section 2 offers a brief summary of the related work. Section 3 illustrates the methodology adopted in this research work. Section 4 explains the experimental findings and analyze the performance of several ML techniques. Lastly, the article is concluded with mentioning the future work in section 5.

2. RELATED WORK

Numerous research have been carried out to categorize various diseases affecting plants, fruits, vegetables, and fish. This section discussed and reviewed some of the relevant studies. Durmus *et al.* [11] applied deep learning techniques, utilizing AlexNet and SqueezeNet as network architectures, to identify tomato leaf diseases. The study utilized the PlantVillage dataset for training purposes. The obtained results showed that SqueezeNet is a suitable option for mobile deep-learning classification due to its efficiency and low computational requirements. Kurniawati *et al.* [12] presented a methodology for diagnosing three types of diseases of paddy. The researchers tested two different thresholding methods on 94 paddy leaf images to determine which provided the best accuracy. The highest accuracy, which was achieved through the use of the local entropy threshold, was 94.7%. Ferentinos [13] designed convolutional neural network (CNN) models for identifying and diagnosing plant diseases. A total of 87,848 images were utilized to train the model whereas the visual geometry group (VGG) model obtained an impressive success rate which is 99.53%. Lu *et al.* [14] put forth a new method for identifying rice diseases (10 diseases) utilizing CNN. The method was based on 500 natural images and achieved an accuracy of 95.48%. Youwen *et al.* [15] utilized support vector machine (SVM) classifier to identify diseases in cucumber leaves. The results showed that using a combination of shape and texture features was more effective than using only shape features. Ma *et al.* [16] identified four types of cucumber diseases employing a deep CNN (DCNN). The authors also evaluated the performance of applied models namely random forest (RF), SVM, and AlexNet. The DCNN performed well in recognizing symptoms using an augmented dataset comprising 14,208 symptom images, achieving an accuracy score of 93.4%. Jiang *et al.* [17] used deep learning to recognize three common diseases. Training accuracy of the Resnet-50 model was 98.3% and test accuracy was 98.0%. Mia *et al.* [18] conducted a thorough examination to identify fish diseases. They implemented eight classification models, and used performance evaluation metrics to assess their performance. Among the models, the accuracy of random forest (RF) model was the highest accuracy which was 88.87%. Saad *et al.* [19] employed different transfer learning approach to identify diseases in eggplants and the best accuracy was 99.06% obtained by DenseNet201. Habib *et al.* [20] conducted a comprehensive study for diagnosing jackfruit diseases. The study employed various classification models, and among them, RF obtained the highest accuracy that was 90%, whereas logistic regression had the lowest accuracy. Biswas *et al.* [21] presented an approach for identifying local birds of Bangladesh through the use of six CNN models with transfer learning technique. The method involved training on a dataset of 2,800 images and testing on 700 images. The best results were achieved with MobileNetV2. Mia *et al.* [22] identified the rare local fruits of Bangladesh. The study used a dataset of 480 images and applied the SVM algorithm for recognition. The SVM model demonstrated a high accuracy that is 94.79%. Biswas *et al.* [23] put forward a solution to classify disease affected carrots. They evaluated the performance of five machine learning classifiers where RF classifier achieved the accuracy of 94.17% which was the highest. Habib *et al.* [24] presented a methodology for the papaya diseases detection and classification. The system uses images captured through mobile devices or handheld devices. To segment the images, k-means clustering is used. The SVM classifier has been used and the accuracy was 90.15%. Habib *et al.* [25] conducted a comparison of nine well-known classifiers to identify diseases in papayas. Among all the models, SVM attained the highest accuracy of 95.2%.

3. METHODOLOGY

The methodology section is comprised of two main parts. In the first part, we have employed a variety of image-processing strategies to extract the statistical and GLCM-based features. Additionally, in the second part, we have employed different machine learning algorithms to identify the diseased cauliflowers effectively.

3.1. Feature extraction procedure

Extraction of features from the collected cauliflower images is given in the Figure 1. In this section, we discussed all the steps in detail that are involved into the feature extraction process. This procedure comprises steps like image resizing, contrast enhancement, color space conversion, and image segmentation by K-Means clustering.

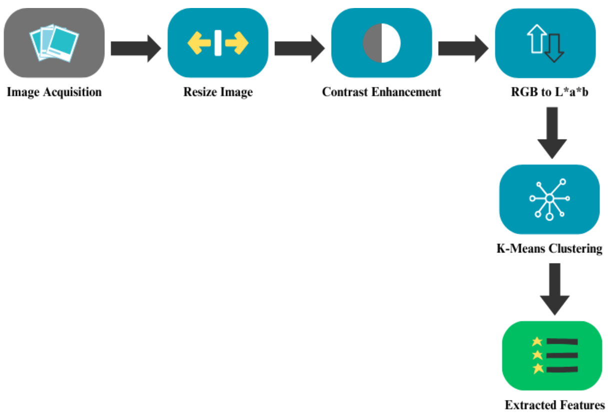

Figure 1. Feature extraction procedure from the collected cauliflower images

3.1.1. Acquisition of image

This study does not require any ethical permission, as no human subjects were involved in the experiments. The process of obtaining images for the study began with a collection of 501 cauliflower images from the field. These images were then divided into two categories: diseased and healthy cauliflower images. Out of these images, 24.59% belong to disease-free cauliflower, while the remaining 75.41% represent diseased cauliflowers.

3.1.2. Resizing image into fixed size

The size of the collected data varies due to the fact that it was acquired by users with various devices from varied perspectives. To standardize the size of the collected images, images have been resized to 300×300 using bicubic interpolation. When we possess data concerning intensity values I as well as the derivatives f_x, f_y, and f_{xy} at the four corners situated at (0, 0), (0, 1), (1, 0), and (1, 1) of a unit square, we can represent the interpolated intensity surface [24] using (1).

$$f\ (x, y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij}\ x^i y^j \tag{1}$$

Here, a_{ij} are coefficients.

3.1.3. Contrast enhancement

After resizing the images, we enhanced their contrast by employing the histogram equalization technique. This method is employed to reveal more information within the images and enhance their overall quality. Let's assume that we have X rows (representing height) and Y columns (representing width) measured in pixels, with each pixel represented by a color intensity (C_k) and an associated pixel count (P_k), and the image's intensity levels denoted as I. In this context, the processed image [24] is generated by transforming each pixel's C_k value into a corresponding color intensity S_k as defined by (2).

$$S_k = \mathrm{T}(C_k) = \frac{\mathrm{I-1}}{\mathrm{XY}} \sum_{\mathrm{j=0}}^{\mathrm{k}} \mathrm{n_j} \tag{2}$$

Where, k=0, 1, 2, 3… I-1.

3.1.4. Color space conversion

After improving contrast of the images, RGB was converted to L*a*b. This modification was necessary for improved performance in the segmentation technique using k-means clustering (3) from [24] was used to convert RGB to CIE (XYZ) and subsequently to L*a*b.

$$\begin{bmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{bmatrix} = \begin{bmatrix} 3.240479 & -1.537150 & -0.498535 \\ -0.969256 & 1.875992 & 0.041556 \\ 0.055648 & -0.204043 & 1.057311 \end{bmatrix} \begin{bmatrix} \mathrm{R} \\ \mathrm{G} \\ \mathrm{B} \end{bmatrix} \tag{3}$$

In order to obtain the L*a*b color space, we calculate the tri-stimulus values for the reference white as X_n, Y_n, and Z_n. Additionally, we assume more that in (4).

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{if } t > 0.008856 \\ 7.787\mathrm{t} + \frac{16}{116} & \text{if } t \leq 0.008856 \end{cases} \tag{4}$$

The computation of L*a*b values relies on the use of the equations provided in [24], specifically (5) to (7).

$$L^* = \begin{cases} 116\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16 & \text{if} \frac{Y}{Y_n} > 0.008856 \\ 903.3\left(\frac{Y}{Y_n}\right) & \text{if} \frac{Y}{Y_n} \leq 0.008856 \end{cases} \tag{5}$$

$$a^* = 500(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)) \tag{6}$$

$$b^* = 200(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)) \tag{7}$$

3.1.5. Image segmentation

Image segmentation with k-means clustering is a method that involves dividing an image into distinct segments based on color or feature similarity using the k-means clustering algorithm. This involves the following steps.

– Assigning each pixel to the cluster that has the closest center value.
– Segmenting the image based on color.
– Selecting the cluster that contains only the region of interest (ROI).

3.1.6. Feature extraction

To classify the diseased cauliflower, various statistical and gray-level co-occurrence matrix (GLCM) features have been derived. Statistical features include mean (MEAN), variance (VAR), standard deviation (SD), kurtosis (KURT), and skewness (SKEW). This features are described as follows. In the following from (8)-(12) concerning statistical features, x_i is the ith gray level (out of a total of L possible value), whose probability of occurrence is $P(x_i)$ [23], [24]. The moment and mode is represented by μ and M_0, respectively.

$$MEAN: m = \sum_{i=0}^{L-1} x_i P(x_i) \tag{8}$$

$$VAR: \sum_{i=0}^{L-1}[(x_i - m)^2 P(x_i)] \tag{9}$$

$$SD: s = \sqrt{\{\sum_{i=0}^{L-1}[(x_i - m)^2 P(x_i)]\}} \tag{10}$$

$$KURT: \frac{\mu_4}{{\mu_2}^2} - 3 \tag{11}$$

$$SKEW: \frac{m - M_0}{s} \tag{12}$$

Contrast (CON), correlation (CORR), energy (ENRG), and entropy (ENTR), and homogeneity (HOMO) have been used as GLCM features. Gray-level spatial dependency matrix elements (i, j) indicate the number of times a pixel with value i appeared horizontally next to a pixel with value j [23] [24]. The above features are described by the following (13) to (17):

$$CON: \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i - j)^2 \, P(i,j) \tag{13}$$

$$CORR: \frac{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (i-\mu_i)(j-\mu_j) \, P(i,j)}{\sigma_i \sigma_j} \tag{14}$$

$$ENRG: \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j)^2 \tag{15}$$

$$HOMO: \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{P(i,j)}{1+(i-j)^2} \tag{16}$$

$$ENTR: -\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} P(i,j) \log P(i,j) \tag{17}$$

The segmentation quality has an impact on feature extraction. From the segmented cauliflower image, the 10 features are retrieved. A sample of extracted features values from the acquired image is shown in Table 1. The study utilized these features to classify images into either the disease-free or diseased category.

Table 1. Feature extraction procedure from collected dataset

Name of class	Acquired image	Contrast enhancement	Segmented image	Extracted features values
Diseased cauliflower				0.388, 0.855, 0.458, 0.930, 25.593, 42.950, 3.472, 1652.382, 5.406, 11.750
Healthy cauliflower				0.552, 0.969, 0.283, 0.900, 97.805, 103.922, 4.772, 5272.607, 1.286, 0.273

3.2. Diseased Cauliflower classification using ML model

In this section, the process of using extracted features to classify diseased cauliflower is explained. Figure 2 illustrates the implementation procedure to classify diseased cauliflower. The implementation procedure starts with splitting the inputted dataset. Then, we have performed feature selection to identify the most important features. After that we have employed different machine learning classifiers to identify the diseased cauliflowers. Finally, the performances of the machine learning models have been compared using different performance metrics.

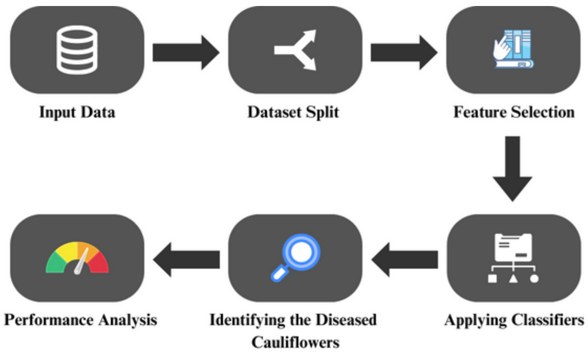

Figure 2. Implementation procedure for applying machine learning techniques to classify the diseased cauliflowers

3.2.1. Dataset splitting

Data splitting is required to train and test the ML models. After extracting the features, training data and testing data have been separated. During the testing phase, 20% of the dataset was utilized, whereas 80% of the dataset was allocated to train the ML models.

3.2.2. Feature selection

In this step, the ANOVA F-values of the features have been measured. Considering the ANOVA F-values, we have chosen the top N ($5 \leq N \leq 9$) attributes having the highest ANOVA F-scores. Based on these top N features, the new training and testing datasets have been prepared to train and test the ML models. Table 2 shows the ANOVA F-scores for the features used in this study. The features are arranged in descending sequence based on their ANOVA F-scores. By employing feature selection techniques, we have eliminated less important features, which has enhanced the scalability of the proposed system by reducing training and testing times.

Table 2. ANOVA F-scores of the features

Feature name	ANOVA F-Score	Feature name	ANOVA F-score
Contrast	78.916679	Mean	1.013293
Correlation	44.150215	Energy	0.391681
Homogeneity	38.739379	Skewness	0.209609
Variance	16.117210	Kurtosis	0.116511
Standard deviation	6.593958	Entropy	0.031020

3.2.3. Applying ML classifier

To classify the diseased cauliflower, five machine learning models have been utilized. Each machine learning model that has been considered has its own unique properties and characteristics. Therefore, in the study, five ML models are RF, SVM, AdaBoost, bagging, and gradient boosting.

3.2.4. Performance evaluation

After applying the classifier, we have measured the performance (accuracy, recall, precision, F1-scores) of the different applied ML models using the following (18) to (21).

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \tag{18}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad (19)$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad (20)$$

$$F1 - score = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (21)$$

4. FINDINGS AND RESULT ANALYSIS

The performance of ML models in categorizing diseased cauliflower has been evaluated. For comparing the applied ML models performance, different metrics like accuracy, recall, precision, F1-scores have been calculated. Based on the ANOVA F-scores, the top N features ($5 \leq N \leq 9$) is selected to train and test the machine learning models. To evaluate the actual effects of the employed feature selection techniques (ANOVA F-scores), we have also trained and tested the proposed ML models with the dataset having all the features. The applied ML models performance of this study is illustrated in Table 3. It can be found that the bagging classifier achieves the highest accuracy of 82.35% while considering the top 9 features (contrast, correlation, homogeneity, variance, standard deviation, mean, energy, skewness, and kurtosis) according to the ANOVA F-scores. It has outperformed other applied ML models in terms of recall (sensitivity) and F1-scores also. Here, the recall, precision, and F1-score of the bagging classifiers are 87.96%, 89.62% and 88.78%, respectively.

Table 3. Different machine learning model's performance for different features set

Features set	Model name	Accuracy (%)	Recall (%)	Precision (%)	F1-score (%)
Top 05 features	SVM	70.59	74.07	86.96	80
	Random forest	73.53	77.78	87.5	82.35
	AdaBoost	72.79	78.7	85.86	82.12
	Bagging	78.68	85.19	87.62	86.39
	Gradient boosting	81.62	87.96	88.79	88.37
Top 06 features	SVM	71.32	75.00	87.1	80.6
	Random forest	74.26	78.7	87.63	82.93
	AdaBoost	72.06	75.93	87.23	81.19
	Bagging	80.15	84.26	90.1	87.08
	Gradient boosting	78.68	82.41	89.9	85.99
Top 07 features	SVM	71.32	75.00	87.1	80.6
	Random forest	75.74	82.41	86.41	84.36
	AdaBoost	73.53	77.78	87.5	82.35
	Bagging	80.88	86.11	89.42	87.73
	Gradient boosting	80.15	85.19	89.32	87.21
Top 08 features	SVM	70.59	74.07	86.96	80.00
	Random forest	77.94	80.56	90.63	85.3
	AdaBoost	75.00	81.48	86.27	83.81
	Bagging	80.15	86.11	88.57	87.32
	Gradient boosting	80.15	86.11	88.57	87.32
Top 09 features	SVM	70.59	74.07	86.96	80.00
	Random forest	81.62	86.11	90.29	88.15
	AdaBoost	75.74	81.48	87.13	84.21
	Bagging	82.35	87.96	89.62	88.78
	Gradient boosting	81.62	87.96	88.79	88.37
Without feature selection	SVM	70.59	74.07	86.96	80.00
	Random forest	79.41	83.33	90.00	86.54
	AdaBoost	77.21	83.33	87.38	85.31
	Bagging	82.35	87.96	89.62	88.78
	Gradient boosting	81.62	87.96	88.79	88.37

In terms of using all the 10 features, the bagging classifier showed similar performance to the earlier model trained with top 9 features. But there was a significant enhancement in the performance of the RF classifier after selecting the top 9 features. By taking all the features, the RF classifier achieved an accuracy of 79.41%. On the other hand, by considering the top 9 features, the accuracy of the RF classifier increased by more than 2%. There were also significant increases in the other metrics.

In terms of using the top 5, 6, 7, and 8 features, the effectiveness of the applied models degraded due to the elimination of important features. Additionally, the ROC curves of the proposed machine learning models have been plotted. Figure 3 represents the ROC curves of the models considering the top 9 features. The area under curve (AUC) value of the bagging classifier is 0.83 which is the highest among other applied

ML models. The AUC value of RF, gradient boosting, and AdaBoost classifiers are 0.81, 0.78, and 0.75, respectively. Conversely, the SVM classifier demonstrated the lowest AUC value of 0.69.

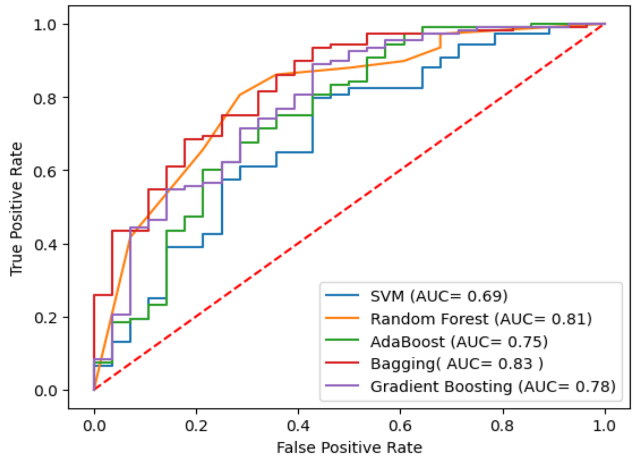

Figure 3. The ROC curves of the five applied ML classifiers (taking into account the top nine dominant features)

5. CONCLUSION

Our proposed solution has the potential to support the agriculture field in Bangladesh by increasing the effectiveness and productivity of cauliflower through automated classification of diseased cauliflower. A total of ten features (GLCM and statistical) extracted from the collected image dataset that were grouped (based on the ANOVA F-value) and utilized to train and test the five ML models to classify cauliflower. The bagging classifier considering the top 09 most dominant features (based on the ANOVA F-value) have been achieved the best performance among all of our investigated classifiers. The accuracy, recall, precision, and F1-score of the bagging classifier are 82.35%, 87.96%, 89.62% and 88.78%, respectively. Finally, it is concluded with a suggestion that appropriate features in the ML models' training and testing has significant impact on the performance of models. However, this approach can be used later to generate end to end solution which will be helpful for the farmers in real-time farming. In addition, we can employ multiclass classification approach to identify diverse diseases of cauliflowers, as well as presenting the explainability of the decision making process of the ML models.

ACKNOWLEDGEMENTS

This research was funded under the Centennial Research Grant (CRG), which was allocated by the Ministry of Education of the Government of the People's Republic of Bangladesh through the University Grants Commission at Sheikh Hasina University.

REFERENCES

[1] "Agriculture growth reduces poverty in Bangladesh," *The World Bank*, 2016. https://www.worldbank.org/en/news/feature/2016/05/17/bangladeshs-agriculture-a-poverty-reducer-in-need-of-modernization#:~:text=Agriculture has played a key,some income from agricultural activities (accessed Oct. 03, 2023).
[2] "Bangladesh-country commercial guide," *International Trade Administration*. https://www.trade.gov/country-commercial-guides/bangladesh-agriculture-sectors (accessed Oct. 03, 2023).
[3] H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li, "Computer vision technology in agricultural automation-A review," *Information Processing in Agriculture*, vol. 7, no. 1, pp. 1–19, 2020, doi: 10.1016/j.inpa.2019.09.006.
[4] E. Mavridou, E. Vrochidou, G. A. Papakostas, T. Pachidis, and V. G. Kaburlasos, "Machine vision systems in precision agriculture for crop farming," *Journal of Imaging*, vol. 5, no. 12, 2019, doi: 10.3390/jimaging5120089.
[5] X. Wang *et al.*, "Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: Systematic review and dose-response meta-analysis of prospective cohort studies," *BMJ (Online)*, vol. 349, 2014, doi: 10.1136/bmj.g4490.
[6] T. Y. Orin, M. U. Mojumdar, S. M. T. Siddiquee, and N. R. Chakraborty, "Cauliflower leaf disease detection using computerized techniques," in *2021 IEEE 6th International Conference on Computing, Communication and Automation, ICCCA 2021*, 2021, pp. 730–733, doi: 10.1109/ICCCA52192.2021.9666437.
[7] "Agriculture learning," 2020. https://www.en.agriculturelearning.com/crops/cauliflower-modern-cultivation-process-and-its-health-benefits/ (accessed Oct. 03, 2023).

[8] "Report on the productivity survey of Cauliflower crop," 2014. https://bbs.portal.gov.bd/sites/default/files/files/bbs.portal.gov.bd/page/16d38ef2_2163_4252_a28b_e65f60dab8a9/cauliflower2014.pdf (accessed Oct. 03, 2023).
[9] C. Leitzmann, "Characteristics and health benefits of phytochemicals," *Forschende Komplementarmedizin*, vol. 23, no. 2, pp. 69–74, 2016, doi: 10.1159/000444063.
[10] "Cauliflower." http://en.banglapedia.org/index.php?title=Cauliflower (accessed Oct. 03, 2023).
[11] H. Durmus, E. O. Gunes, and M. Kirci, "Disease detection on the leaves of the tomato plants by using deep learning," 2017, doi: 10.1109/Agro-Geoinformatics.2017.8047016.
[12] N. N. Kurniawati, S. N. H. S. Abdullah, S. Abdullah, and S. Abdullah, "Investigation on image processing techniques for diagnosing paddy diseases," in *SoCPaR 2009-Soft Computing and Pattern Recognition*, 2009, pp. 272–277, doi: 10.1109/SoCPaR.2009.62.
[13] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018, doi: 10.1016/j.compag.2018.01.009.
[14] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, "Identification of rice diseases using deep convolutional neural networks," *Neurocomputing*, vol. 267, pp. 378–384, 2017, doi: 10.1016/j.neucom.2017.06.023.
[15] T. Youwen, L. Tianlai, and N. Yan, "The recognition of cucumber disease based on image processing and support vector machine," in *Proceedings-1st International Congress on Image and Signal Processing, CISP 2008*, 2008, vol. 2, pp. 262–267, doi: 10.1109/CISP.2008.29.
[16] J. Ma, K. Du, F. Zheng, L. Zhang, Z. Gong, and Z. Sun, "A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network," *Computers and Electronics in Agriculture*, vol. 154, pp. 18–24, 2018, doi: 10.1016/j.compag.2018.08.048.
[17] D. Jiang, F. Li, Y. Yang, and S. Yu, "A tomato leaf diseases classification method based on deep learning," in *Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020*, 2020, pp. 1446–1450, doi: 10.1109/CCDC49329.2020.9164457.
[18] M. J. Mia, R. B. Mahmud, M. S. Sadad, H. Al Asad, and R. Hossain, "An in-depth automated approach for fish disease recognition," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 9, pp. 7174–7183, 2022, doi: 10.1016/j.jksuci.2022.02.023.
[19] I. H. Saad, M. M. Islam, I. K. Himel, and M. J. Mia, "An automated approach for eggplant disease recognition using transfer learning," *Bulletin of Electrical Engineering and Informatics*, vol. 11, no. 5, pp. 2789–2798, 2022, doi: 10.11591/eei.v11i5.3575.
[20] M. T. Habib, M. J. Mia, M. S. Uddin, and F. Ahmed, "An in-depth exploration of automated jackfruit disease recognition," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1200–1209, 2022, doi: 10.1016/j.jksuci.2020.04.018.
[21] A. A. Biswas, M. M. Rahman, A. Rajbongshi, and A. Majumder, "Recognition of local birds using different CNN architectures with transfer learning," *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021, doi: 10.1109/ICCCI50826.2021.9402686.
[22] M. R. Mia, M. J. Mia, A. Majumder, S. Supriya, and M. T. Habib, "Computer vision based local fruit recognition," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 2810–2820, 2019, doi: 10.35940/ijeat.A9789.109119.
[23] A. A. Biswas, M. S. Zulfiker, A. Rajbongshi, M. J. Mia, and A. Majumder, "Feature ranking based carrot disease recognition using MIFS method," in *Lecture Notes in Networks and Systems*, 2022, vol. 420, pp. 56–68, doi: 10.1007/978-3-030-96305-7_6.
[24] M. T. Habib, A. Majumder, A. Jakaria, M. Akter, M. S. Uddin, and F. Ahmed, "Machine vision-based papaya disease recognition," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 3, pp. 300–309, 2020.
[25] M. T. Habib, A. Majumder, R. N. Nandi, F. Ahmed, and M. S. Uddin, "A comparative study of classifiers in the context of papaya disease recognition," in *Proceedings of International Joint Conference on Computational Intelligence*, 2020, pp. 417–429, doi: 10.1007/978-981-13-7564-4_36.

BIOGRAPHIES OF AUTHORS

Mala Rani Barman was born in Dhaka, Bangladesh in 1995. She received her M.Sc. and B.Sc. degrees in Computer Science and Engineering from Jahangirnagar University in 2017 and 2016 respectively. Currently she is an assistant professor and Head of the Computer Science and Engineering Department of Sheikh Hasina University. Her research interest is in machine learning, computer vision and deep learning. She can be contacted at email: malabarman42@gmail.com.

Al Amin Biswas received the Bachelor of Science and Master of Science degrees from the Department of Computer Science and Engineering (CSE), Jahangirnagar University, Dhaka, Bangladesh. He is currently working as a Lecturer with the CSE Department, Bangabandhu Sheikh Mujibur Rahman University (BSMRU), Kishoreganj, Bangladesh. His research interests include machine learning, deep learning, data mining, and computer vision. He has more than 35 research papers published in international conferences and journals. He has served as the Session Chair for the International Conference on Computer Communication and Informatics (ICCCI 2021). He also served as the Co-Convener for the International Conference on Machine Intelligence and Data Science Application (MIDAS 2021). He reviewed research papers for many international conferences and journals. He received the best paper award twice for his outstanding research works from two springer conferences. He can be contacted at email: alaminbiswas.cse@gmail.com.

Marjia Sultana currently working as a lecturer at the department of Computer Science and Engineering of Begum Rokeya University, Rangpur, Bangladesh. She received her B.Sc. (Hons) degree in Computer Science and Engineering from Jahangirnagar University, Savar, Dhaka in 2016 and M.Sc. degree from the same university in 2018. Her research interest focuses on machine learning, deep learning, data mining, computer vision, image processing and computer networking. She has several research papers published in international conferences and journals. She can be contacted at email: marjia.cse@brur.ac.bd.

Aditya Rajbongshi currently working at Bangladesh's Bangabandhu Sheikh Mujibur Rahman Digital University (BDU) as a lecturer in the Department of Educational Technology. In 2017 and 2018, he graduated with a B.Sc. and an M.Sc. from Jahangirnagar University in Dhaka, Bangladesh's Department of Computer Science and Engineering, respectively. He was employed as a lecturer at the National Institute of Textile Engineering and Research in Dhaka, Bangladesh, prior to joining BDU. His areas of interest in research are computer vision, image processing, data mining, machine learning, and deep learning. He has had over 20 research articles accepted for publication in journals and conferences worldwide. He reviewed research articles for several publications and international conferences. He can be contacted at email: aditya0001@bdu.ac.bd.

Md. Sabab Zulfiker was born in Jamalpur, Bangladesh. He received his B.Sc. and M.Sc. degrees from the Department of Computer Science and Engineering, Jahangirnagar University, Dhaka, Bangladesh in 2016 and 2017, respectively. Currently, he is working as a lecturer in the Department of Computer Science and Engineering at Bangabandhu Sheikh Mujibur Rahman University (BSMRU), Kishoreganj, Bangladesh. Before joining BSMRU, he worked as a Senior Lecturer at Daffodil International University (DIU), Dhaka, Bangladesh. Additionally, he has hands-on experience in software development and worked as a Software Engineer at S3 Innovate Pte. Ltd. His research interests include machine learning, deep learning, data mining, natural language processing, image processing, and computer vision. He can be contacted at email: sabab.rumc@gmail.com.

Tasnim Tabassum was born in Bangladesh on October 16, 1997. She received the bachelor and the master degrees in Computer Science and Engineering from Jahangirnagar University. She is currently working as a lecturer at Department of CSE, Daffodil International University. Her research interest is machine learning and computer vision. She can be contacted at email: tasnimtabassum16@gmail.com