Measurement: Sensors 24 (2022) 100518

Contents lists available at ScienceDirect

Measurement: Sensors

journal homepage: www.sciencedirect.com/journal/measurement-sensors

Feature selection and classification improvement of Kinnow using SVM classifier

Sukhpreet Singh *, Kamal Malik

Department of Computer Science and Engineering, CT University, Ferozepur Road, Sidhwan Khurd, 142024, India

ARTICLE INFO

Keywords:
Classification
Feature extraction
Feature selection
Kinnow
Normalization

ABSTRACT

An approach to enhance the classification of the kinnow is proposed in this paper. The fruit images are captured in the proposed approach, and texture, color, shape, and size features are extracted and merged to generate a dataset. To cope with outliers and dominant features, the Pareto normalization method is used. A hybrid feature selection approach that combines neighborhood component analysis and ReliefF methods to select the optimal features is proposed to eliminate redundant and irrelevant features. The dataset is classified using the SVM machine learning algorithm following the feature selection. Utilizing the SVM classifier, the proposed approach chooses 54.20% of the features with an accuracy of 94.67%. This proves that the proposed approach is efficient and can be used for the classification of the other fruits.

1. Introduction

In the agriculture field, more than three-fourths of Indians work directly or indirectly, which makes the foundation of the country's economy [1]. India is the world's second-largest fruit grower, producing 81285 thousand MT in 2012-13. By 2020–2021, it is expected to reach 98 million tones, making it the world's top producer of citrus fruits, mangoes and bananas. India is the second-largest producer of citrus fruits after Mexico (H. [2]. The most widely used and grown citrus fruit in India is kinnow. About half of the entire citrus producing regions are under cultivation. The kinnow ranks first among citrus crops in terms of yield, juice content, and fruit quality. The kinnow citrus fruit type is mostly grown in the Punjab area. It is round in form and orange in color which becomes ripe in the month of January or February. It is a significant economic factor since it is a high source of calcium and vitamin C [3]. The grading of the kinnow is an important task for the farmers as the post-harvest procedures such as handling, packing etc. depends on the grading methods. The manual grading of the fruits may have errors thus require more time and quite expensive.

Recently, automatic inspection of the fruits based on the computer vision techniques are used for the classification of the fruits into different categories which can lower labor costs while improving product quality [4]. The automatic classification of the fruits is highly significant since it might increase the grower's profit. Classification methods can enhance the post-harvesting methods of the fruits. Different researchers have proposed different methods for the grading, defect classification of the kinnow. Singh and Nill proposed a technique for the kinnow classification based on color-based features using the k-means segmentation method and artificial neural network classifier (H. [2]. Hadimani and Mittal introduced a computer vision methodology to estimate the maturity of the kinnow based on the color and physiochemical values [5]. Yadav et al. investigated the physical attributed of the kinnow for the grading task [6]. Hadimani and Garg proposed an approach to classify different defects of the kinnow based on the different features [7]. Akhter et al. proposed classification technique for the classification of the citrus fruits including kinnow using the color and shape features. The principal component analysis method used for the feature extraction of the citrus fruits [8].

The existing methods are for the classifications of the kinnow are not considered all the features of the fruits. There may be irrelevant and redundant feature exists in the extracted features which affect the classification task. The other problem in the classification task is the presence of the outliers and dominant feature which degrade the classification accuracy.

These problems motivate us to propose an approach for the classification of the kinnow. In the proposed approach different features of the kinnow are extracted and optimal features are selecting using the proposed feature selection method. A normalization method is utilized in the proposed approach to cope with the outliers and the dominant features.

* Corresponding author.
E-mail addresses: Sukhpreet.manshahia@gmail.com (S. Singh), kamal.malik91@gmail.com (K. Malik).

https://doi.org/10.1016/j.measen.2022.100518
Received 13 September 2022; Received in revised form 24 September 2022; Accepted 9 October 2022
Available online 17 October 2022

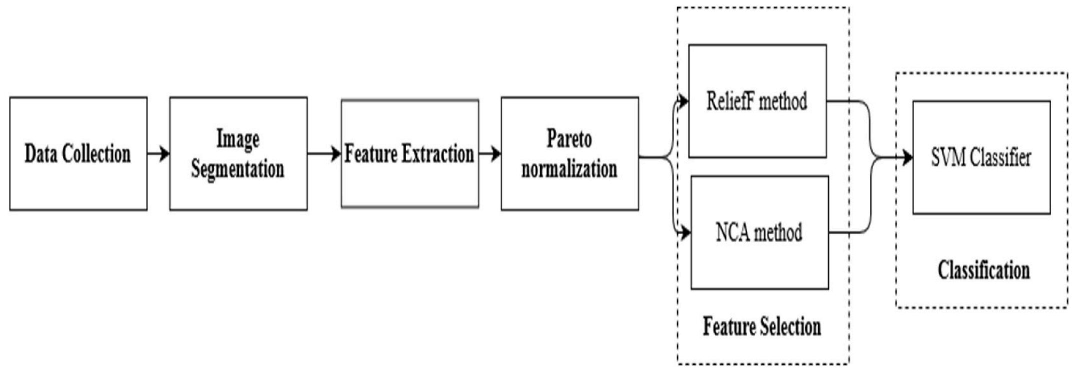

Fig. 1. Block diagram of proposed approach.

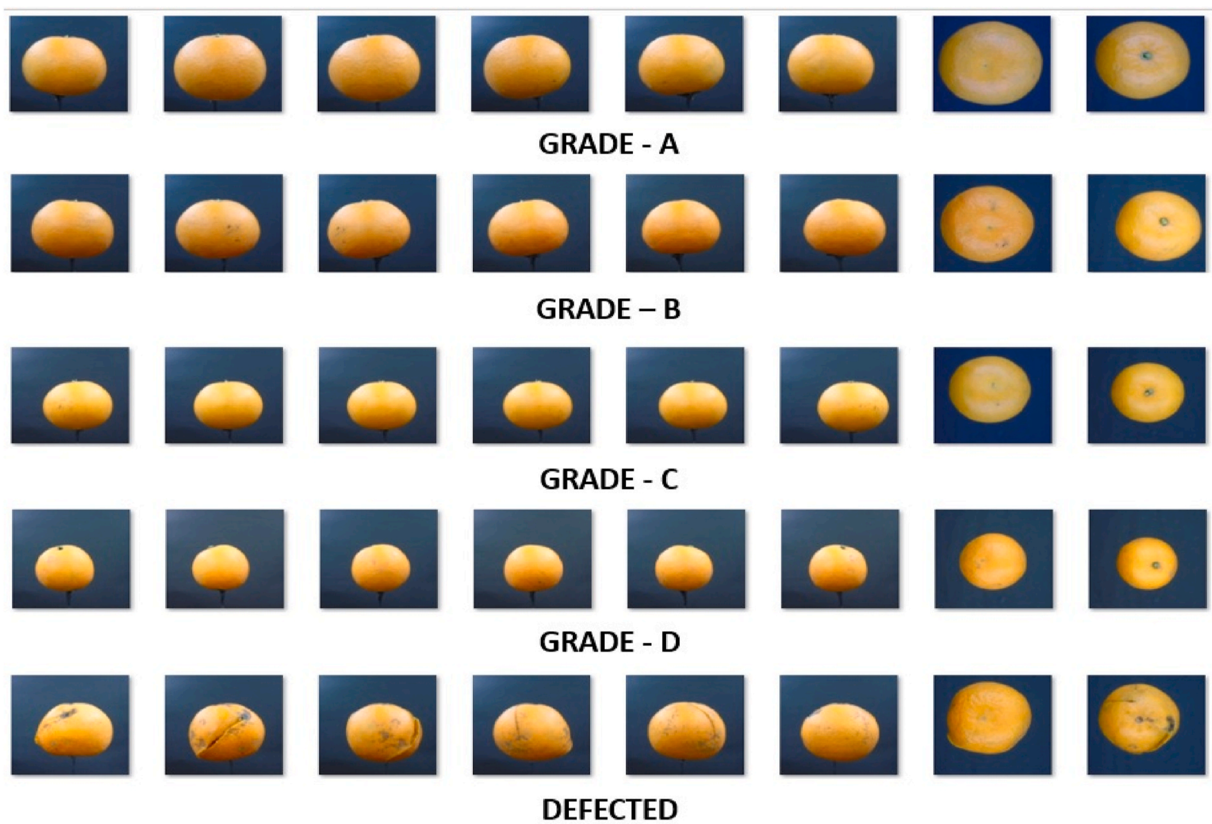

Fig. 2. Images of kinnow of different grade.

The rest of the paper is organized as follows: Section 2, describes the proposed approach in detail. The analysis of the obtained results is discussed in the Section 3. The conclusions and findings are given in the last section.

2. Proposed approach

In this paper, an approach is proposed for the classification of the kinnow. Fig. 1 depicts the block diagram of the proposed approach. The proposed approach is divided into various stages which are: dataset collection, image segmentation, feature extraction, feature normalization, feature selection, and Classification.

The details description of the various stages of the proposed approach is explain in the subsections.

2.1. Dataset collection

Two charged coupled devices (CCD) color cameras are used for the image acquisition system. Color images of size 2492 $\times$ 2492 are generated using the system. Two T-shaped fluorescent tubes at the top of the square chamber used for the lighting system. Two camera holes are located in the chamber: one at the chamber's top and one at the left wall. To avoid direct light, a reflective surface cover has been positioned between the fluorescent bulb and kinnow. With the kinnow rotated thirty degrees horizontally, six images from the side camera and two images from the top camera are captured. There are total five categories of grades of kinnow (Grade-A, Grade-B, Grade-C, Grade-D and Defected) as shown in Fig. 2. These grades are classified according to size and texture of Kinnow. For each grade, 30 kinnow are selected. Kinnow Mandarin variety is collected from Malwa belt of Punjab and manual grading is done by us and experts on the basis on size of kinnow. Total

Table 1
Total images collected for dataset preparation.

Categories of Grades	No. of kinnow	No. of images
Grade-A	30	240
Grade-B	30	240
Grade-C	30	240
Grade-D	30	240
Defected	30	240
Total	**150**	**1200**

images collected are shown in Table 1.

Sample images collected are shown in Fig. 2.

2.2. Image segmentation

Segmentation methods are used for the identification of the different regions of an image(G. [9]. In this approach two regions are selected in which the first region specifies the fruit background and other region is fruit. To achieve the segmentation of the image, Expectation-Maximization (EM) algorithm is used [10].

Fig. 3 shows the block diagram of the EM algorithm; i represents the number of the iteration, ε and i_{max} are the termination criteria for the iteration, and L_i is the likelihood $L_i = \sum_{t=1}^{N_t} p_i\ (x_t)$.

In this method the pixels of the image are clustered into different groups for the image segmentation. The algorithm is based on the Gaussian mixture model. The algorithm is divided into two parts: expectation and maximization step. In the expectation step the expected value of the log-likelihood of the complete data is calculated while in the maximization step new parameters are calculated which is used to maximize the expected value for the image segmentation [11].

Segmented sample image is shown in Fig. 3.

2.3. Feature extraction

In the proposed approach, different features are retrieved from the Kinnow fruit images to divide fruits into different categories. To discover the patterns among fruit samples belonging to the specific category, the features are extracted. According to this method, characteristics are derived based on the Kinnow texture, color, and shape and size. Texture features of the images are a collection of two-dimensional arrays that are used to visualize texture of an object [12]. These features are used to analyze the distribution of color intensities of an image. The texture features of the fruit is using linear binary patterns (LBP) [13], Gray-Level Co-occurrence Matrices (GLCM) [14] feature extraction methods. The formulation of the LBP extraction method is given as follow:

$$LBP_{PX} = \sum_{P_x=0}^{P_x=1} d\left(g_{p_x} - g_a\right)$$

$$d(n) = \begin{cases} 1, n \geq 0 \\ 0, n < 0 \end{cases}$$

Where P_x is considered as the index for the image pixel as $(g_0, \ldots\ldots\ldots\ldots, g_{P_{x-1}})$.

Energy, contrast, homogeneity, mean, correlation, maximum probability, entropy, etc. based features are extracted in the GLCM method. Based on the various orientations, statistical features are calculated in this feature extraction method. On the other hand the color space features are used to represent the image data into distinct spaces to visualization of colors [15]. In the feature extraction method, HSV color space features are extracted [16]. The formulation of the HSV features set is as follows:

$$mx_{(i,j)} = max\left(I_{R(i,j)}, I_{G(i,j)}, I_{B(i,j)}\right);\ mn_{(i,j)} = min\left(I_{R(i,j)}, I_{G(i,j)}, I_{B(i,j)}\right)$$

$$H(i,j) = \begin{pmatrix} \dfrac{60 * \left(I_{G(i,j)} - I_{B(i,j)}\right)}{mx - mn} I_{R(i,j)} > \max\left(I_{G(i,j)}, I_{B(i,j)}\right) \\ \dfrac{180 * \left(I_{B(i,j)} - I_{R(i,j)}\right)}{mx - mn} I_{G(i,j)} > \max\left(I_{R(i,j)}, I_{B(i,j)}\right) \\ \dfrac{300 * \left(I_{R(i,j)} - I_{G(i,j)}\right)}{mx - mn} I_{B(i,j)} > \max\left(I_{R(i,j)}, I_{G(i,j)}\right) \end{pmatrix}$$

$$S(i,j) = \left[\frac{mx - mn}{mx}\right]$$

$$V(i,j) = [mx]$$

where I denotes the input RGB image of the Kinnow fruit.

The size and shape features are also extracted of the Kinnow fruit as this feature is widely used by the farmers for the grading the different categories of the fruits.

2.4. Feature normalization

The conversion of features into a common range, known as normalization, is a crucial preprocessing procedure that prevents larger numeric feature values from dominating smaller numeric feature values. The major objective of the normalization methods is to reduce the bias of the features which are contributing numerically more than others. In the proposed approach pareto normalization method is used as the performance of pareto is better for the feature selection process (D. [17].

2.5. Feature selection

High dimensional datasets may have irrelevant and redundant features. The quality of the feature set affects the performance of the machine learning algorithms. Thus, Feature selection is important task to enhance the performance of the machine learning algorithms. In the

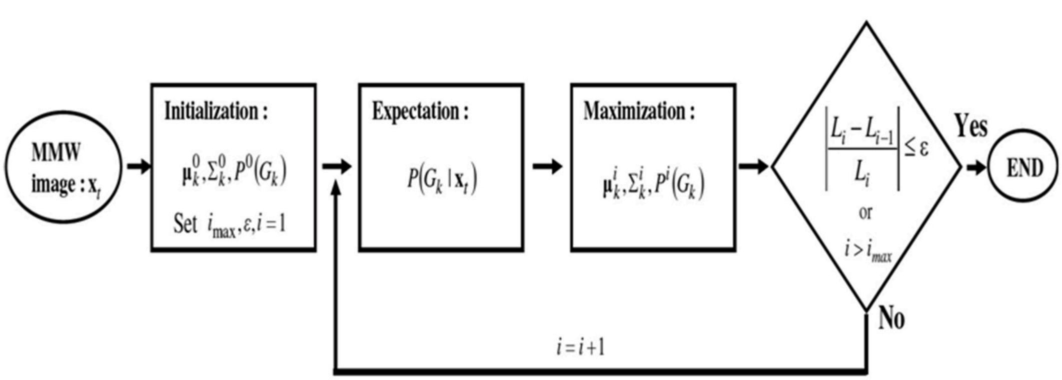

Fig. 3. EM algorithm.

Table 2
Performance metrics comparisons.

	Accuracy	Sensitivity	Specificity	F-score	Precision	Kappa value	Matthews Correlation Coefficient
Without feature selection and normalization	90.84	90.84	97.71	90.54	92.44	89.09	71.39
With feature selection and without normalization	92.47	92.47	98.12	92.17	93.64	90.95	76.46
With feature selection and normalization **(Proposed Approach)**	94.67	94.67	98.67	94.50	95.62	93.67	83.33

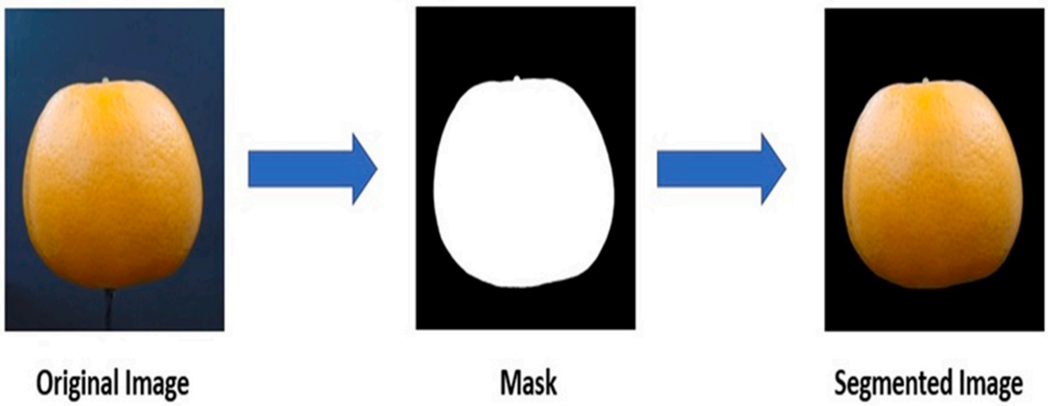

Fig. 4. Image of segmented kinnow.

proposed approach, Neighborhood component analysis [18] and ReliefF [19] feature selection techniques are utilized to improve the classification accuracy and speeding up calculation. NCA is a non-parametric utilized to increase the predictive accuracy of the classification algorithms by optimizing the anticipated classification accuracy with a regularization term, it learns a feature weighting vector [20]. NCA has the benefit of providing information on important features together with feature rating. More significantly, throughout the dimensionality reduction process, NCA doesn't lose any information. Whereas, the ReliefF determines the feature weights based on the Manhattan distance and assign both positive and negative weights [21]. ReliefF is an improved version of Relief feature selection method [22]. In the proposed approach both feature selection methods are combined to select optimal feature having highest weights for classification improvement.

2.6. Classification

Support Vector Machine (SVM) classifier is used as a learning algorithm in the proposed approach. As, the SVM searches for the best hyperplane to divide the categories using the feature values. SVM is a supervised learning model with associated learning calculations that dissect data used for regression and classification. The base spacing from the isolating hyper plane to the nearest instance must be increased for support vector machines. In the basic SVM, only binary classification is supported, however case extension is also possible for multi-class classification. In this work, SVM is used for the multi-class classification of

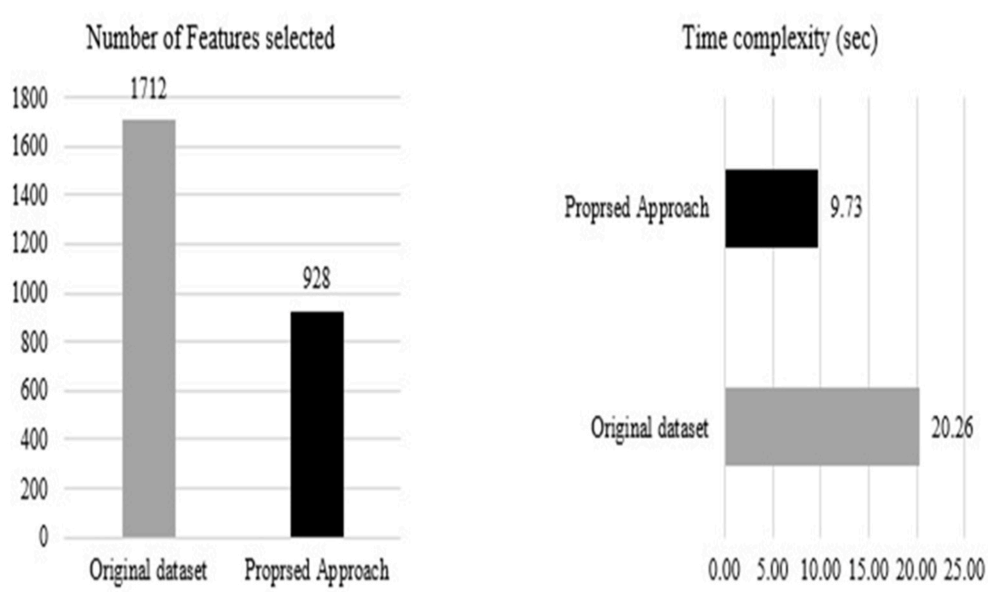

Fig. 5. Features selected and Time complexity of the proposed approach.

the kinnow.

3. Results and discussions

In this paper, an approach for classification improvement of the kinnow is proposed. The experiments are performed on a Windows 10 Pro operating system running on the computer having the Intel® Xenon® CPU E5-2650 v3 (2.30 GHz) and 8 GB of RAM using the MATLAB 2019a platform. Five-fold cross-validation is used for the classification performance evaluation.

In Table 2, the impact of the feature selection and normalization methods on the classification of the kinnow have been presented. Accuracy, sensitivity, specificity, F-score, precision, Kappa value, and Matthew's correlation coefficient metrics are used for the analysis purpose.

From the Table 1, it can be observed that the performance of proposed approach is better as compared to the other cases considered in this work. This proves that the proposed approach is very efficient in the kinnow classification.

In Fig. 4, numbers of features selected, and computational time of the proposed approach is shown with comparison to the original dataset (see Fig. 5).

From the figure, it has been observed that the 928 features are selected using the proposed feature selection approach from the total of 1712 features. Also, the execution time of the proposed algorithm is less in comparison to the time required for the classification of the whole datasets. This proved the significance of the proposed approach in the kinnow classification task.

4. Conclusion

An approach to enhance the classification of kinnow is proposed in this paper. By employing pareto normalization and a hybrid feature selection strategy, the proposed approach has enhanced the classification performance of the SVM classifier by eliminating the issue of outliers, dominating features, redundant, and irrelevant features. The proposed approach involves capturing kinnow images and 1712 features are extracted using various feature extraction methods. The effectiveness of the data normalization and features selection approaches has been analyzed. The proposed approach accurately selected 928 features with a 94.67% accuracy rate. Additionally, the computation time of the proposed approach is 10.53 s quicker, demonstrating its efficiency. The proposed approach can eventually be used for the classification of the various fruits.

CRediT authorship contribution statement

Sukhpreet Singh: Conceptualization, Methodology, Data curation, Writing – original draft, Investigation. **Kamal Malik:** Visualization, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] D. Londhe, S. Nalawade, G. Pawar, V. Atkari, S. Wandkar, Grader: a review of different methods of grading for fruits and vegetables, Agric. Eng. Int.: CIGR Journal 15 (3) (2013) 217–230.

[2] H. Singh, N. Gill, Machine vision based color grading of kinnow Mandarin, Int. J. Adv. Res. Comput. Sci. Software Eng. 5 (5) (2015) 1253–1259.

[3] Samiksha, Citrus Fruit Cultivation in India-Production Area, Climate, Harvesting and Fruit Handling Citrus Fruit Cultivation in India-Production Area, Climate, Harvesting and Fruit Handling!, 2021. https://www.yourarticlelibrary.com/fruits/citrus-fruit-cultivation-in-india-production-area-climate-harvesting-and-fruit-handling/24698.

[4] V. Kavitha, M.R. Devi, Predicting the diseases by Graphcut method for citrus fruit, Int. Res. J. Manag. Sci. Technol. 7 (12) (2016) 465–470.

[5] L. Hadimani, N. Mittal, Development of a computer vision system to estimate the colour indices of Kinnow mandarins, J. Food Sci. Technol. 56 (4) (2019) 2305–2311, https://doi.org/10.1007/s13197-019-03641-9.

[6] A. Ali, N.K. Yadav, A. Ali, R. Dev, M. Pandurang, Evaluation of physical properties of different grades of kinnow Mandarin, J. Pharmacogn. Phytochem. 8 (1) (2019) 1414–1417.

[7] L. Hadimani, N.M. Garg, Automatic surface defects classification of Kinnow mandarins using combination of multi-feature fusion techniques, J. Food Process. Eng. 44 (1) (2020) 1–15, https://doi.org/10.1111/jfpe.13589.

[8] N. Akhter, M. Idrees, F.U. Rehman, M. Ilyas, Q. Abbas, M. Luqman, Shape and texture based classification of citrus using principal component analysis, Int. J. Agric. Extension 9 (2) (2021) 229–238, https://doi.org/10.33687/ijae.009.02.3525.

[9] G. Singh, M.K. Sachan, An unconstrained and effective approach of script identification for online Bilingual handwritten text, Natl. Acad. Sci. Lett. 43 (5) (2020) 453–456, https://doi.org/10.1007/s40009-020-00889-0.

[10] D.-S. Lee, J.-Y. Son, S. Yeom, S.-H. Kim, Automatic image segmentation for concealed object detection using the expectation-maximization algorithm, Opt Express 18 (10) (2010) 10659–10667, https://doi.org/10.1364/OE.18.010659, 18 (10), 10659–10667.

[11] I. Kokkinos, P. Maragos, Synergy between object recognition and image segmentation using the expectation-maximization algorithm, IEEE Trans. Pattern Anal. Mach. Intell. 31 (8) (2009) 1486–1501, https://doi.org/10.1109/TPAMI.2008.158.

[12] S. Naik, B. Patel, Machine vision based fruit classification and grading - a review, Int. J. Comput. Appl. 170 (9) (2017) 22–34, https://doi.org/10.5120/ijca2017914937.

[13] G. Muhammad, Date fruits classification using texture descriptors and shape-size features, Eng. Appl. Artif. Intell. 37 (2015) 361–367, https://doi.org/10.1016/j.engappai.2014.10.001.

[14] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Transact. Syst. Man Cybernetics 6 (1973) 610–621.

[15] H. Afrisal, M. Faris, P.G. Utomo, L. Grezelda, I. Soesanti, F.M. Andri, Portable smart sorting and grading machine for fruits using computer vision, in: Proceeding - 2013 International Conference on Computer, Control, Informatics and its Applications: "Recent Challenges in Computer, Control and Informatics", vol. 2013, IC3INA, 2013, pp. 71–75, https://doi.org/10.1109/IC3INA.2013.6819151.

[16] R.C. Gonzalez, R.E. Woods, S.L.U. Eddins, Digital Image Processing Using MATLAB, Pearson Prentice Hall, Upper Saddle River, New Jersey, 2004.

[17] D. Singh, B. Singh, Investigating the impact of data normalization on classification performance, Appl. Soft Comput. 97 (2020), 105524, https://doi.org/10.1016/j.asoc.2019.105524.

[18] J. Goldberger, S. Roweis, G. Hinton, R. Salakhutdinov, Neighbourhood components analysis, Adv. Neural Inf. Process. Syst. (2005).

[19] I. Kononenko, Estimating attributes: analysis and extensions of RELIEF, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 784 LNCS, Springer Verlag, 1994, pp. 171–182, https://doi.org/10.1007/3-540-57868-4_57.

[20] W. Yang, K. Wang, W. Zuo, Neighborhood component feature selection for high-dimensional data, J. Comput. 7 (1) (2012) 162–168, https://doi.org/10.4304/jcp.7.1.161-168.

[21] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. 53 (1–2) (2003) 23–69, https://doi.org/10.1023/A:1025667309714.

[22] T. Tuncer, S. Dogan, F. Ozyurt, An automated Residual Exemplar Local Binary Pattern and iterative ReliefF based corona detection method using lung X-ray image, Chemometr. Intell. Lab. Syst. 203 (2020), 104054, https://doi.org/10.1016/j.chemolab.2020.104054.