International Journal of
Intelligent Systems and Applications in Engineering

ISSN:2147-6799 www.ijisae.org Original Research Paper

Different Apple Varieties Classification Using kNN and MLP Algorithms

Kadir SABANCI*[1], Muhammed Fahri ÜNLERŞEN[2]

Accepted 3rd September 2016

Abstract:In this study, three different apple varieties grown in Karaman province are classified using kNN and MLP algorithms. 90 apples in total, 30 Golden Delicious, 30 Granny Smith and 30 Starking Delicious have been used in the study. DFK 23U445 USB 3.0 (with Fujinon C Mount Lens) industrial camera has been used to capture apple images. 4 size properties (diameter, area, perimeter and fullness) and 3 color properties (red, green, blue) have been decided using image processing techniques through analyzing each apple image. A data set which contains 7 physical features for each apple has been obtained. Classification success rates and error rates have been decided changing the neuron numbers in the hidden layers in the classification using MLP model and in different neighbor values in the classification made using kNN algorithm. It is seen that the classification using MLP model is much higher. While the success rate of classification made according to apple type is 98.8889%.

Keywords:*Image processing, Apple classification, kNN, MLP*

1. Introduction

Apple production is an important economic activity in Karaman. The dominant apple varieties produced in Karaman, are Golden Delicious, Granny Smith and Starking Delicious. During marketing these apples, the product is wanted to be consisted of just one species of apples. The mixture rate of other apple varieties in the product is wanted to be at minimum level. For this purpose many selection system have been proposed in the literature.

He Yong et al. (2007) have investigated the visible/near-infrared (Vis/NIR) spectroscopy method for its capacity to non-destructively distinguish apple varieties. The chemo metrics steps employed to the Vis/NIR data were artificial neural network (ANN), wavelet transform (WT) and principal component analysis (PCA). The best result of classification success rate of the apple types have been found as 100% with WT-ANN model [1].

Ronald et al. (2016) have proposed to investigate the applicability and performance of Naive Bayes algorithm in the classification of apple fruit varieties. Apple classification system prototype was built using MATLAB R2015a development platform environment. The results showed that the averaged values of the estimated accuracy, sensitivity, precision and specificity were 91%, 77%, 100% and 80% respectively [2].

Wu et al. (2016) introduced a fuzzy clustering method, termed fuzzy discriminant c-means (FCM) clustering, to distinguish varieties of apples by using NIR spectroscopy. The dataset consists of four types of apple with 200 samples. Experiments have been done to gather the near infrared reflectance (NIR) spectra. FDCM accuracy have achieved 97%. FDCM was better than FCM, possibility c-means, and Gustafson–Kessel (GK) during clustering operation [3].

Shahin et al. (2013) have proposed to detect bruises in apples by using line–scan x–ray imaging method. Transform and spatial characteristics were investigated about their segregating capability of classification of fruit by using bruise defects. In this study Golden delicious (GD) and Red delicious (RD) apples have been used. To distinguish one month old and a few hour old bruises, several artificial neural network (ANN) classifiers were advanced. The ANN classifier accomplished the classification success rate for RD apples as 90% and for GD apples as 83% when it was employed to classify the apples that have one month old bruises. The accuracy of classifier was 60% for both RD and GD apples when it was employed to classify the apples that have a few hour old bruise [4].

In this study, 90 apples grown in Karaman have been used. 7 morphological attributes have been obtained from their images by using image processing techniques. The apples have been classified into 3 group (Golden Delicious, Granny Smith and Starking Delicious) with kNN and MLP algorithms by using the obtained attributes.

2. Materials and Methods

An interface presented in Figure 1, have been designed by using GUIDE (Graphical User Interface Design) in MATLAB. The image of the apple under the camera is taken when the Take Button pressed on the GUI.

The mean of R (red), G (green) and B (blue) pixel values of the area where the apple is in the image have been calculated. These values have been used for classification of the apples for their colors. Then the colored images have been converted to grey level image by using image processing techniques. Morphological operations have been applied onto bitwise images. The image processing steps have been presented in Figure 2.

[1] *KaramanogluMehmetbey University Electrical and Electronics Engineering Engineering Department Karaman Turkey*
[2] *Necmettin Erbakan University Electrical and Electronics Engineering Engineering Department Konya, Turkey*
** Corresponding Author:Email: kadirsabanci@kmu.edu.tr*
Note: This paper has been presented at the 3rd International Conference on Advanced Technology & Sciences (ICAT'16) held in Konya (Turkey), September 01-03, 2016.

Figure 1. Graphical user interface designed in MATLAB GUIDE.

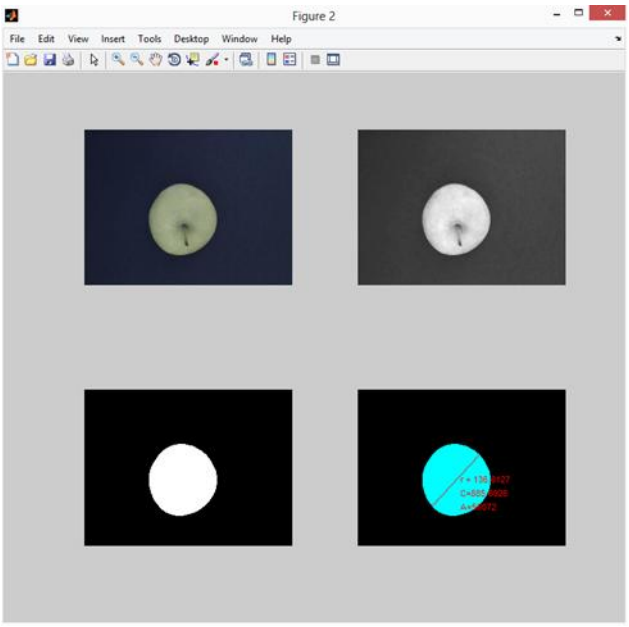

Figure 2. The image processing steps

The physical properties of the apple like radius, perimeter, area, volume, mass and eccentricity have been extracted. So that creates 7 attributes by gathering 4 morphological and 3 colour.

2.1. Software-WEKA

Developed by Waikato University in New Zealand, WEKA is an open-source data mining software with a functional graphical interface which incorporates machine learning algorithms [5]. WEKA includes various classification,data pre-processing, regression, association rules, clustering and visualization tools. The algorithms can be applied on the data cluster either directly or by calling via Java code [6,7]. They are also suitable for developing new machine learning algorithms.

2.2. Multilayer Perceptron

It is a feed forward type artificial neural network model which corresponds input sets onto propersets of output. A multilayer perceptron (MLP) is composed of multiple layers where each layer is connected to the other one. Each node is a processing element or a neuron that has a nonlinear activation function exceptthe input nodes. It uses a supervised learning technique named back propagation and it is used for training the network. The alteration of the standard linear perceptron, MLP is capable of distinguishing data which are not linearly separable [7].

2.3. K-Nearest Neighbour Algorithm

A supervised learning algorithm, k-NN solves classification problems. Classification is the examination of the attributes of an image and the designation of this image to a predefined class. The critical point is the determination of the features of each category previously [8]. Conforming to the used classifcation algorithm k-NN based on the attributes drawn from the classification stage, the distance of the new individual that is wanted to be classified to all previous individuals is considered and the nearest k class is used. As an outcome of this procedure, the belonging of the test data is determined due to the k-nearest neighbour category which contains more exactly determined classes.In k-NN, the determination of the algorithm used for distance calculation and neighbour number are the critical optimization points.In the study, the optimum k number is appointed with experiments. In the calculation of distance, the Euclidean Distance is performed. Euclidean calculation method [9]:

$$d(x_i, x_j) = \left(\sum_{s=1}^{p} (x_{is} + x_{js})^2 \right)^2$$

xi and xj are two points that is wanted to be learnt the distance between them.

In classification procedure, inputs are diameter, perimeter, area, fullness, mean red, mean green and mean blue values. The outputs are Golden Delicious, Granny Smith and Starking Delicious. The whole classification system have been presented in Figure 3.

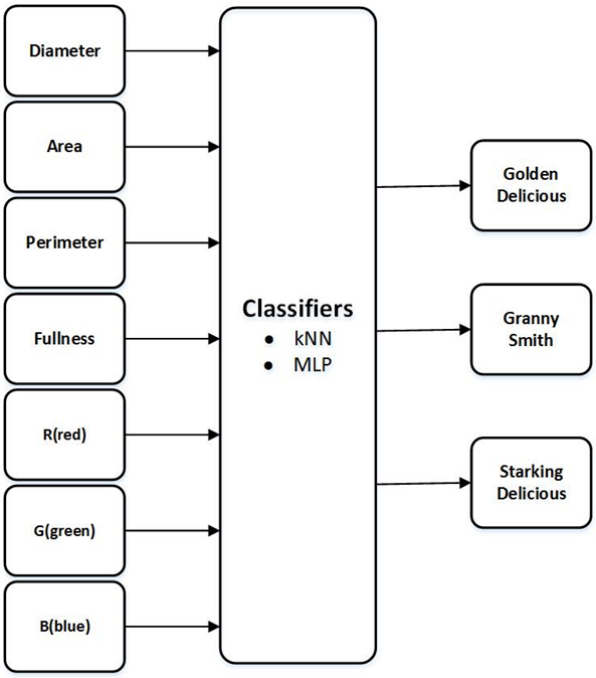

Figure 3. The block diagram of the classification process

3. Results and Discussion

In the study, WEKA software was used in order to classify apples by species (Golden Delicious, Granny Smith and Starking Delicious). Using the kNN algorithm, the classification success rates, the mean absolute error (MAE) and the root mean squared error (RMSE) were obtained for various k-neighbour values. The classification accuracieswere obtained with k-NN algorithm, and MAE and RMSE values were presented in Table 1. The success rate change versus number of neighbours for kNN method is presented in Figure 4.

Table 1. The success rate and error values obtained by using kNN classifier

Neighbourliness number (k)	*Classification accuracy (%)*	*MAE*	*RMSE*
1	93.3333	0.0414	0.1886
2	96.6667	0.2296	0.2854
3	97.7778	0.0364	0.1283
4	90.0000	0.0512	0.1537
5	94.4444	0.0543	0.1541
6	93.3333	0.0539	0.1500
7	95.5556	0.0546	0.1503
8	94.4444	0.0644	0.1573
9	94.4444	0.0728	0.1670
10	91.1111	0.0803	0.1735

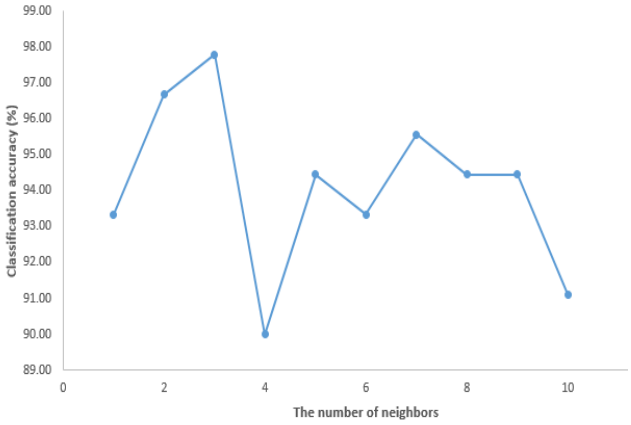

Figure 4. Success rate versus number of neighbours.

The data in the same dataset were processed using the multilayer perceptron model, and the classification accuracy, MAE and RMSE values were obtained while the hidden layer neuron number were changing from 1 to 20. In the MLP model, the training was performed by taking the learning rate value as 0.3, momentum value as 0.2 and iteration number as 500. The classification accuracy, MAE and RMSE values were collectedby using the MLP model have been presented in Table 2.The diagram demonstrating the changes in classification success rate based on the number of neuron in the hidden layer is demonstrated in Figure 5.

Table 2. The success rate and error values obtained by using MLP classifier

The number of neurons in the hidden layer	*Classification accuracy (%)*	*MAE*	*RMSE*
1	94.4444	0.1975	0.2513
2	97.7778	0.033	0.0985
3	95.5556	0.0375	0.1350
4	96.6667	0.0352	0.1152
5	98.8889	0.1698	0.2225
6	97.7778	0.0355	0.1089
7	97.7778	0.0269	0.1049
8	97.7778	0.0223	0.1073
9	96.6667	0.0479	0.1234
10	96.6667	0.0441	0.1214
20	95.5556	0.0746	0.1476

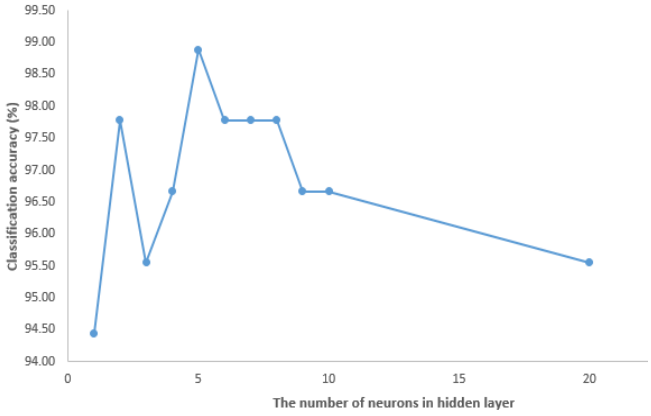

Figure 5. Success rates versus number of neurons in the hidden layer

The model of MLP whose hidden layer consists of 5 neurons which creates the highest classification accuracy is shown in Figure 6.

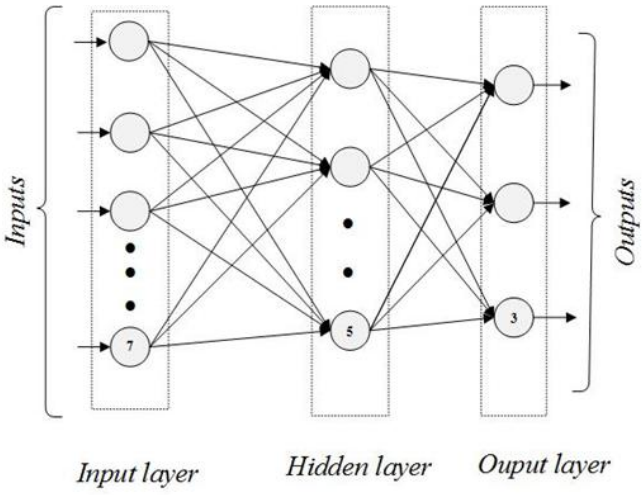

Figure 6. The structure of Multilayer Perceptron

4. Conclusions

In this study, classification of apples as Golden Delicious, Granny Smith and Starking Delicious have been proposed by using machine learning algorithms. The classification success rates and error values of data mining algorithms were calculated for kNN and MLP. It was observed that the success rate was higher for the classifications performed using the MLP algorithm. The best classification result was achieved when the number of neighbourhood was 3 and the success rate was 97.7778%. The MAE and RMSE error value were 0.0364 and 0.1283 respectively. For the classification accuracy obtained using MLP algorithm, the highest classification accuracy and minimum MAE value was achieved for 5 neurons in the hidden layer and it was 98.8889%. For this MLP structure, the MAE and RMSE error values were 0.1698 and 0.2225 respectively.

Acknowledgment

This study was supported by KaramanogluMehmetbey University scientific research project coordination with 01-M-15 project number.

References

[1] He, Y., Li, X., Shao, Y., (2007). Fast Discrimination of Apple Varieties Using Vis/NIR Spectroscopy, Vol. 10(1). Pages. 9-18.

[2] Ronald, M., Evans, M., (2016). Classification of Selected Apple Fruit Varieties Using Naive Bayes, Indian Journal of Computer Science and Engineering (IJCSE), Vol. 7(1) Pages. 13-19.

[3] Wu, X., Wu, B., Yang, N.,(2016). Classification of Apple Varieties Using Near Infrared Reflectance Spectroscopy and Fuzzy Discriminant C-Means Clustering Model, Journal of Food Process Engineering, doi: 10.1111/jfpe.12355.

[4] Shahin, M.A., Tollner, E.W., McClendon, R.W., Arabnia, H.R.,(2013). Apple Classification Based on Surface Bruises Using Image Processing and Neural Networks, Transactions of the ASAE, Vol. 45(5). Pages. 1619–1627.

[5] Witten I.H., Frank E., Hall M.A., (2011). Data mining: practical machine learning tools and techniques. Elsevier, London.

[6] Patterson, D., Liu, F., Turner, D., Concepcion, A., Lynch, R., (2008). Performance Comparison of the Data Reduction System. Proceedings of the SPIE Symposium on Defense and Security, Mart, Orlando, FL, pp. 27-34.

[7] Hall, M., Frank, E., Holmes, G., Pfahringer, B.,Reutemann, P. & Witten, I. H., (2009). The WEKA Data Mining Software: An Update, SIGKDD Explorations, Vol. 11(1). Pages. 10-18.

[8] Wang, J., Neskovic, P., & Cooper, L. N., (2007). Improving nearest neighbour rule with a simple adaptive distance measure, Pattern Recognition Letters, Vol. 28(2). Pages. 207-213.

[9] Zhou, Y., Li, Y. & Xia, S., (2009). An improved KNN text classification algorithm based on clustering, Journal of computers, Vol. 4(3). Pages. 230-237.