# Using soft computing and leaf dimensions to determine sex in immature Pistacia vera genotypes

**Article** *in* Measurement · January 2021

DOI: 10.1016/j.measurement.2021.108988



**4 authors:**

Mehdi Rezaei
University of shahrood, shahrood, Iran


Abbas Rohani
Ferdowsi University Of Mashhad


Parviz Heidari
Shahrood University of Technology


Shaneka Lawson

Measurement 174 (2021) 108988

Contents lists available at ScienceDirect

# Measurement

journal homepage: www.elsevier.com/locate/measurement

# Using soft computing and leaf dimensions to determine sex in immature *Pistacia vera* genotypes

Mehdi Rezaei, Abbas Rohani *, Parviz Heidari, Shaneka Lawson

*Department of Horticultural Sciences, Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran*
*Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran*
*Department of Agronomy and Plant Breeding, Faculty of Agriculture, Shahrood University of Technology, Shahrood, Iran*
*USDA Forest Service, Northern Research Station, Hardwood Tree Improvement and Regeneration Center (HTIRC), Purdue University, PFEN226C, Department of Forestry and Natural Resources, 715 West State Street, West Lafayette, IN 47907, USA*

ARTICLE INFO

*Keywords:*
Classifier
Leaf characteristics
*Pistacia vera*
Pistachio
Sex determination

ABSTRACT

Pistachio (*Pistacia vera*) is a dioecious tree species in the cashew family originating from Central Asia and the Middle East. The sex of pistachio genotypes cannot be established until inflorescence growth during the reproductive stage. The ability to determine sex before flowering would be extremely beneficial for pistachio growers and breeders. In this study, different methods of soft computing, using classifiers (i.e. KNN, SVM, SOM, ANFIS and RBF) were employed to determine pistachio sex based on geometric features of mature leaves. Data were collected from more than 900 leaves representing 45 pistachio genotypes before sexual maturation. Separate training data set sizes (20, 40, 60 and 80%) were used to evaluate the generalizability of each classifier. Results indicated the KNN classifier was the most accurate (>95%) for sex determination during training phases however, accuracy during testing phases varied from 52 to 60% for the four training data set sizes. Training phase accuracy with the 80% data set ranged from 83 to 64% with the respective arrangement of RBF, ANFIS, SVM, and the SOM classifiers. In the test phase, classifier accuracy varied from 88.33% (RBF) to 60% (KNN). Our results indicated RBF could reliably be used to differentiate between male and female pistachio genotypes. Thus, soft computing models are useful tools for predicting sex in pistachio based on leaf dimensions.

## 1. Introduction

Costs of fruit tree breeding are prohibitive due to land requirements, long-term upkeep, and an extended juvenile phase. Additional expenses incurred during removal of male trees prior to fruit evaluation makes dioecious species more expensive to maintain than hermaphroditic tree species. *Pistacia vera* is a dioecious member of the Anacardiaceae family. Early sex detection in offspring of crosses among selected parents in pistachio breeding programs would be invaluable to the industry as pistachio tree sex cannot be determined until they reach sexual maturity and begin inflorescence growth [1]. Use of soft modeling techniques and basic morphological traits to verify pistachio seedling sex prior to maturation would greatly benefit researchers and significantly decrease costs for breeders.

Previous reports describe sex-specific morphological differences within *Pistacia* spp. such as differentiation in *P. atlantica* leaf morphology [2], distinctive main branch angles, leaf indices, leaf areas [3], and distinguishing stoma attributes in *P. chinensis* [4], and leaf longevity and eco-physiological differences in *P. lentiscus* [5,6]. Morphological differences exist in mature pistachio trees as male trees are usually larger in size than female counterparts in commercial orchards with fruit production limiting female tree size [7]. Prior to maturation, males and females of some species can be identified in other ways such as microscopic inspection of the winter bud [3,8,9]. Alternatively, published literature reports pistachio sex determination can be performed at the DNA level with PCR-based DNA amplification using

*Abbreviations:* ANN, Artificial neural network; ANOVA, Analysis of variance; ARB, Adaptive rule-based; AUC, area under a ROC curve; Ens, Ensemble classifier; FIS, fuzzy inference system; FN, Number of positive samples.; FP, Number of negative samples; KNN, K-nearest neighbor; LM, Levenberg-Marquardt; $LW_1$, Ratio of length-to-width of main leaf; $LW_2$, Ratio of length-to-width of terminal leaf; MLP, Multilayer perceptron; Poly2, Polynomial degree 2; Poly3, Polynomial degree 3; RBF, Radial basic function; SOFM, Self-organizing feature map; SOM, Self-organized map; SVM, Support vector machine; TDS, Training data size; TLCF, Two-layer classification framework; TN, Negative samples; TP, Number of positive samples; TSSE, Total sum squared error; YI, Youden's index.

* Corresponding author at: Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran.
*E-mail address:* arohani@um.ac.ir (A. Rohani).

https://doi.org/10.1016/j.measurement.2021.108988
Received 25 August 2020; Received in revised form 10 December 2020; Accepted 4 January 2021
Available online 11 January 2021

sex-specific primers [10–12].

Soft computing such as use of neural networks has hastened our ability to predict physiological processes within living organs [13,14]. Methods like K-nearest neighbor (KNN), support vector machine (SVM), self-organizing map (SOM), artificial neural network fuzzy inference system (ANFIS), and radial basis function (RBF) are being used for unique applications in animal behavior [15], medicine and cancer diagnosis [16,17], plant species hierarchical classification [18], disease diagnosis [19], image classification [20,21], regression studies [22], prediction of plant yield and performance [21,23,24], tree phenology [25], and predictions of superior genotype [26]. Machine learning based on visual assessment of biometric data has been used for sex determination in humans, fish and other animals [27–29], but has yet to be used in plants. In this research, KNN, SOM, SVMs, ANFIS and RBF from deep learning models were used to sex identification in pistachio. The KNN decision rule is a famous classification technique used in pattern recognition and is one of the most useful non-parametric algorithms in soft computing and machine learning [30–33]. Compared to other neural networks, KNN is simple, stable, and demonstrates good grouping performance [34]. A SOM is a type of artificial neural network (ANN) tool used for categorizing pattern in n-dimensional observation data to produce a low-dimensional (typically two-dimensional) output called a map [35]. SVMs are supervised learning models with associated learning algorithms that analyze the data used for classification and regression analysis. Given a set of training examples, each allocated to one of two categories, a SVM training algorithm builds a model that assigns new examples a category, making it a non-probabilistic binary linear classifier. ANFIS is a type of ANN based on the Takagi–Sugeno fuzzy inference system [36] that allows for identification of two network structure parts, namely the premise and consequence. A RBF network is an ANN that typically has three layers: an input layer, a hidden layer with a non-linear RBF activation function, and a linear output layer. RBF uses radial basis functions as activation functions, realizes a globally optimal solution to the adjustable weights in the minimum mean square error sense by using the linear optimization method, and has general approximation and regularization capabilities [37].

Although DNA markers for early sex determination exist, there are very few sex-specific gene loci in pistachio genotypes [12]. Few DNA markers can distinguish between sexes with 100% accuracy. For example, Kamiab, Ebadi, Panahi and Tajabadi [38] reported a 15% error rate in the sex determination of *P. vera* using SCAR (sequence characterized amplified region) markers. DNA testing is expensive and requires expertise and laboratory facilities. The deep learning method described here is a new simple and cost-effective technique for predicting different biological phenomena. It could be employed as an alternative or complementary method for early sex determination in pistachio studies.

A comprehensive literature search revealed no published research in pistachio using soft computing methods for sex-specific studies thus this work is the first to use machine learning classifiers in pistachio sex determination based on leaf characteristics.

The remainder of this manuscript is organized into sections. Input data set measurements, a brief description of the various soft computing classifiers assigned to the data, and a classification performance assessment are provided in Section 2. The results and analysis of each classifier within the different data sets, a direct comparison of the performance classifiers, and the RBF sensitivity analysis are presented in Section 3. Section 4 consists only of the Conclusions.

## 2. Materials and methods

### 2.1. Data set

Forty-five pistachio genotypes were selected for data collection and more than 900 leaves per genotype were measured. Leaves were collected from sexually immature trees during the third year of growth (before flowering) in 2005. Twenty male and twenty female genotypes were randomly selected after flowering in 2019. Calculated data from male (400 leaves) and female (500 leaves) genotypes were used to generate the data sets for modeling. Pistachio seedling trees were chosen from a native population at Damghan (36°14′74″ N, 54°41′07″E; altitude 1170 m) in the Semnan province of Iran. Genotypes within this population were obtained from the open-pollinated seeds from two local pistachio cultivars 'Shah-Pasand' and 'Kalleh-Ghuchi'. These trees are irrigated regularly and maintained by a professional horticulturalist. Leaves from five female Iranian pistachio cultivars (Ahmad Aghaei, Kalleh-Ghuchi, Fandaghi, Shah-Pasand, and Akbari) were also sampled. Roughly 20 mature pinnate leaves from each genotype were randomly selected from early season shoots. A digital caliper with ±0.01 mm accuracy was used to measure length and width of the main leaf and terminal leaflet (Fig. 1). The length: width ratio in leaves is rarely affected by abiotic conditions. Thus, main leaf ($LW_1$) and terminal leaflet ($LW_2$) ratios were used as model inputs.

Features were normalized to remove differences among the feature ranges. The feature normalization range was considered between −1 and 1.

$$F_n = 1 + \frac{2(F - F_{min})}{(F_{max} - F_{min})} \tag{1}$$

where $F = [LW_1, LW_2]$ is the original feature value, $F_n$ is the normalized feature value, and $F_{max}$ and $F_{\min}$ represent maximum and minimum feature values.

### 2.2. Classifier models

#### 2.2.1. K-Nearest Neighbor classifier (KNN)

KNN, the simplest machine learning classifier algorithm, was applied to the training and testing phases to determine the label of each pattern based on the most similar K-neighbor. In this study, the KNN test begins with selection of a K-neighbor from pistachio trees that have the closest Euclidean distance to the characteristics of the tree being tested. The sex class of K-neighbors is determined by majority "vote" thus the class with the largest number of votes becomes the assigned sex of the tree [16,39].

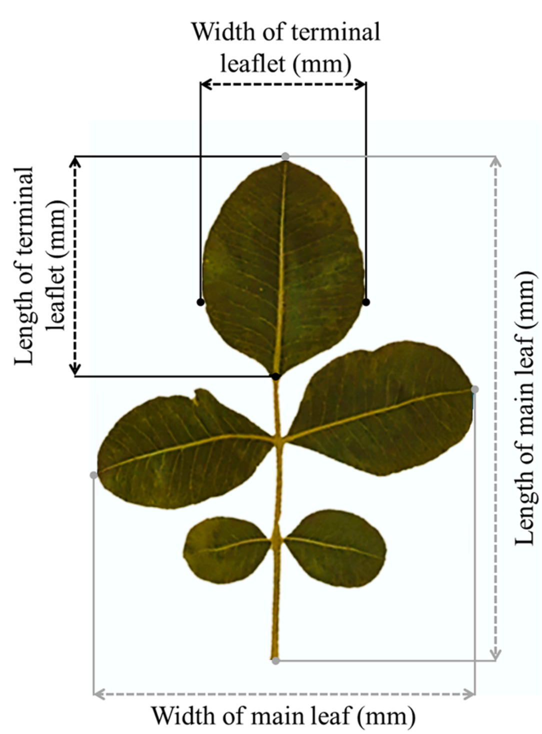


**Fig. 1.** Measurement locations. Depiction of where length and width data were taken for main leaf and terminal leaflets in pistachio.

The voting standard is the short feature distance between the testing tree and its K-neighboring trees. The weighted Euclidean distance best describes this:

$$D_t = \sqrt[2]{\left(\sum_{i=1}^{N} W_i |X_i - X_{im}|^2\right)}, i = 1,2,...,N, t = 1,2,..,T \quad (2)$$

where $X_{im}$ and $X_i$ are, respectively, the training and testing pattern, $W$ and $N$ are the weight of the $i^{th}$ feature, the numbers of training patterns, respectively.

For example, when $K = 4$ is and $X_n$ is an unknown sex pistachio tree; the Euclidian distance is shortest to the three neighboring female pistachio trees rather than the single male tree. Therefore, the unknown sex tree is assigned a female designation (Fig. 2).

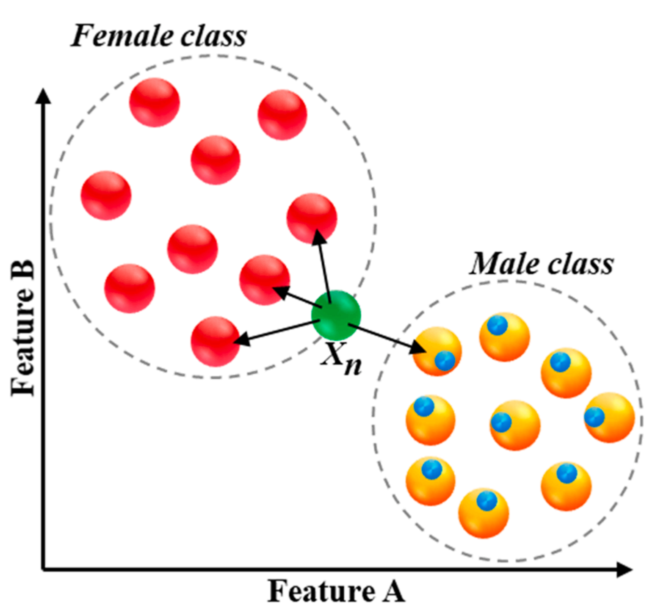


**Fig. 2.** KNN classifier. Example of KNN showing Euclidian distances between an unknown ($X_n$) and its four closest neighbors.

### 2.2.2. Support vector machine (SVM)

The SVM was firstly described as a classifier by Vapnik [40] and has the potential to aid in identification of sex in pistachio. We also selected SVM because it overcomes local minima, requires very little training time, and uses kernel functions to generate a variety of solutions [41,42]. The SVM classifier components for pistachio sex classification are illustrated in Fig. 3. The SVM operates by using a separating hyperplane to simultaneously maximize the margin and minimize misclassification between two female and male classes from the training data set. For pistachio sex detection as a two-class classification problem, the input system is a 2-dimentional vector $x(x = [LW_1, LW_2])$ where male pistachios have a value of $+1$, and females $-1$ ($y \in \{-1, +1\}$). So that:

$$\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \cdots, \boldsymbol{x}_m\} \in R^2 \quad (3)$$

$$y = \{y_1, y_2, y_3, \cdots, y_m\} \in \{-1, +1\} \quad (4)$$

Therefore, the objective of the SVM classifier is to find a decision function that can correctly identify the sex ($y$) of each pistachio ($x$). If the hyperplane created by the SVM is assumed to be linear, then we have:

$$f(x) = \langle w.x \rangle + b \quad (5)$$

where w is the weight vector and b is a scalar quantity. Therefore, the hyperplane ($f(x)$) for the training set could be defined:

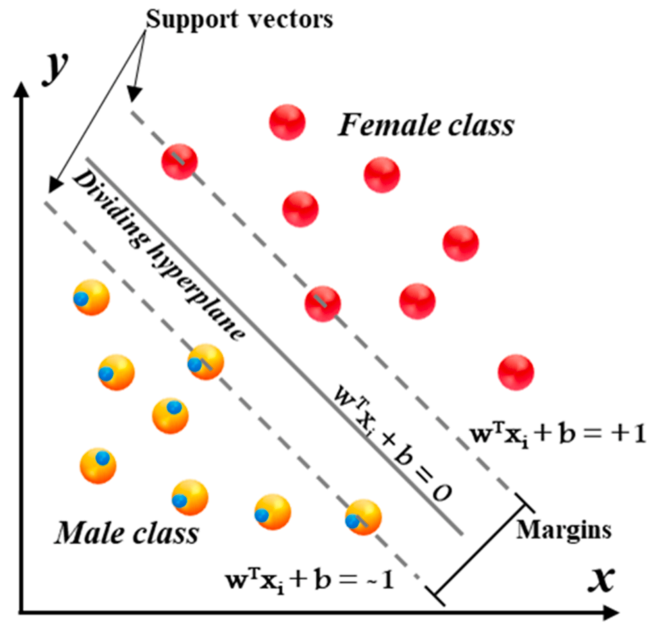


**Fig. 3.** SVM classifier. Illustration of linear support vector classifier components.

$$\boldsymbol{w}.x_i + b \geq 1, if \quad y_i = +1$$

$$\boldsymbol{w}.x_i + b \leq -1, if \quad y_i = -1 \quad (6)$$

The maximum distance between the two margin lines should be indicated by $w$ and $b$ with the goal of solving the minimized equation (8).

$$Minimize \frac{1}{2} ||w||^2 + C\sum_{i=1}^{n} \xi_i$$

$$Subject \ to \quad y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (7)$$

where $\xi$ is the slake variable and C is the penalty parameter.

### 2.2.3. SOM classifier

Kohonen's self-organizing map (SOM) classifier is an unsupervised neural network that creates a rendering of the training sample input space [43]. The SOM generates a lower-dimensional output layer from high-dimensional feature inputs and shows each input layer neuron is associated with the Kohonen layer neurons. The SOM structure is shown in Fig. 4. During the training phase, the weight vector of the neuron in the Kohonen layer is corrected based on the distance from the input feature vector. The neuron closest to the selected feature, is identified as a winner neuron using equation (8). Winner neuron weights are updated by equation (9):

$$N_i = \underset{j}{\operatorname{argmin}} ||x_i - w_j||, i = 1,2,3,\cdots j = 1,2,3,\cdots \quad (8)$$

$$w_i(t+1) = w_i(t) + \alpha(t) h_{ci}(t) ||x_i(t) - w_i(t)|| \quad (9)$$

where $t$ is the training iteration, $\alpha(t)$ is the learning rate, $h_{ci}(t)$ is the neighborhood function, and c is the winner neuron in the Kohonen layer.

At the beginning of the SOM, the weight of the neurons in the Kohonen layer is measured then the input vector is randomly selected and their distance to all the neurons is calculated in the Kohonen layer. The weight of the winning neurons is then updated based on Equation (9). Learning rate and neighborhood function change with each iteration of training. Training continues until the change is not observed in SOM parameters. After training, the label of each of the winning neurons is identified in the output layer.

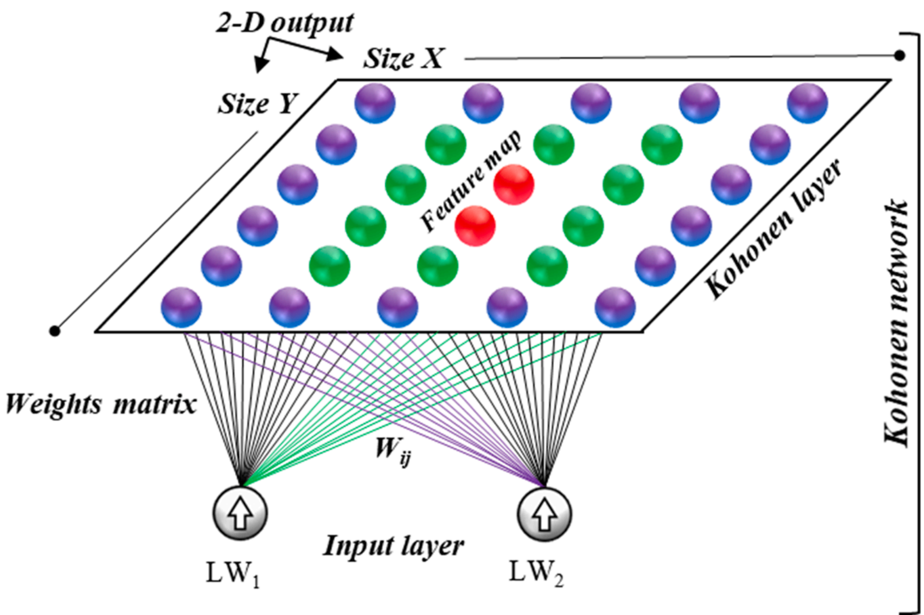


**Fig. 4.** SOM classifier. Schematic of a SOM two-layer neural network.

### 2.2.4. ANFIS classifier

The adaptive neuro-fuzzy inference system (ANFIS) combines an artificial neural network (ANN) with fuzzy inference system (FIS) principles to take advantage of both methods. The ANFIS structure (Fig. 5) was used in this study to aid pistachio sex identification efforts. Fuzzy rules in ANFIS are extracted during the training phase to ensure correct pistachio sex classification. ANFIS can create fuzzy *if-then* rules based on first-order Sugeno models (Equation (10)).

$$\begin{aligned} &Rule\ \ 1: If\ \ LW_1\ \ is\ \ A_1\ \ and\ \ LW_2\ \ is\ \ B_1, then\ \ f_1 = p_1LW_1 + q_1LW_2 + r_1 \\ &Rule\ \ 2: If\ \ LW_1\ \ is\ \ A_2\ \ and\ \ LW_2\ \ is\ \ B_2, then\ \ f_2 = p_2LW_1 + q_2LW_2 + r_2 \end{aligned} \tag{10}$$

where $LW_1$ and $LW_2$ are the leave geometric features, $A_i$ and $B_i$ are the ANFIS fuzzy sets, $fi$ is the output, and $p_i$, $q_i$, and $ri$ are the parameters achieved during the training phase.

The first layer of ANFIS is called the fuzzification layer. Fuzzification output is calculated using the membership function and the number of fuzzy rules are determined with the second layer of FIS. Second layer output is normalized for each rule after entering the third layer. In the fourth layer, the normalized values of the previous layer are multiplied by the 1st-order polynomial. ANFIS output (Equation (11)) is calculated for each sex class in the fifth layer (Ci):

$$C_i = \frac{\sum_j \overline{w}_j (p_jLW_1 + q_jLW_2 + r_j)}{\sum_j w_j} \tag{11}$$

The most efficient and accurate ANFIS design has the least number of FIS rules. The FIS is made using three approaches: grid partition (GP), subtractive clustering (SC) and fuzzy c-means (FCM).

### 2.2.5. RBF neural network classifier

The radial basis function (RBF) artificial neural network was also used to verify pistachio sex classifications. RBF is composed of three layers; input, hidden, and output (Fig. 6). The input layer contains the geometric features of pistachio leaves ($LW_1$, $LW_2$). The hidden layer is made up of neurons with a nonlinear function called the radial basis or Gaussian function. RBF produces a linear output layer and the relationship between input and output is computed in Equation (12).

$$Out_k(\boldsymbol{x}) = \sum_{i=1}^{L} w_{kj} h_j(\boldsymbol{x}) + b_{k0} \tag{12}$$

$$h_j(\boldsymbol{x}) = exp\left( - \frac{\|\boldsymbol{x} - \boldsymbol{\mu}_j\|^2}{2\sigma_j^2} \right) \tag{13}$$

where $\boldsymbol{x}$ is the input vector of features, $h_j(\boldsymbol{x})$ is a radial basis function

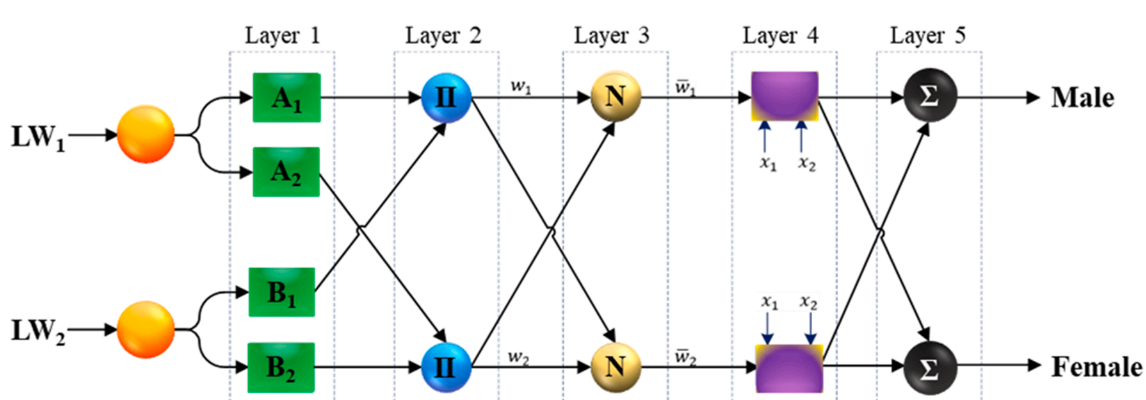


**Fig. 5.** ANFIS structure. This configuration was used to establish sex (male, female) in pistachio genotypes. (circle, fixed node; square, adaptive node).

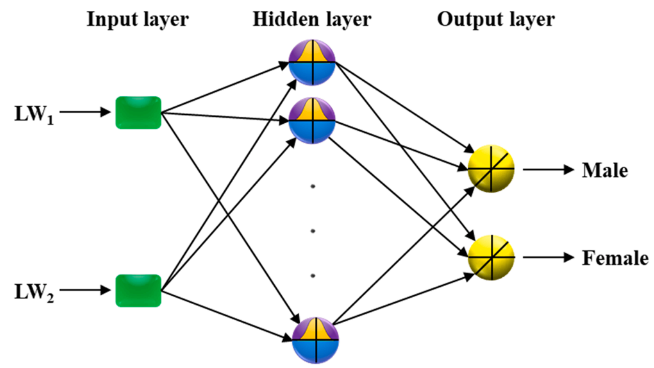


**Fig. 6.** RBF classifier. Structure of the RBF neural network when establishing pistachio sex from leaf data.

(activation function), $b_{k0}$ is the bias, L is the number of neurons in the hidden layer, and $\boldsymbol{\mu}_j$ and $\sigma_j$ are, respectively, the center and spread of the activation function.

During the training phase and based on the correct class, pistachio sex patterns were calculated using the Levenberg-Marquardt (LM) training algorithm of optimal weight values between neurons, μ and σ. The LM training algorithm optimizes RBF classifier parameters based on minimization of the total sum squared error (TSSE).

### 2.3. Evaluation of classification performance

The aim of this study was to identify the sex of pistachio trees in two male and female groups so that binary classification would result in a positive or negative output. A confusion matrix was used to estimate classifier performance (Table 1). The matrix consisted of four components; true sex positive (TP), false sex positive (FP), true sex negative (TN), and false sex negative (FN). The TP and TN were the number of sex patterns correctly identified and correctly rejected, respectively. The FP and FN were incorrectly identified and incorrectly rejected, respectively.

Classifier performance was estimated for TP, TN, FP and FN using Equation (14).

$$Accuracy(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{14}$$

Percent accuracy refers to the ratio of correct predictions over the total number of evaluated patterns. The best classifier performance is achieved when the accuracy is close to a hundred [44,45].

The functioning stages of machine learning methods in KNN, SVM, SOM, ANFIS, and RBF models are divided into two phases, training and testing. The data set in this study, which includes the geometric characteristics of 900 pistachio tree leaves (400 male, 500 female), was randomly divided into two data sets for training and testing. During the training phase, classifier parameters that can be changed are adjusted using the training data set. Then, in the testing phase, classification performance is evaluated using a data set other than the training data set to assess its generalizability in real conditions. In the larger the training data set, the classifier can use more patterns to adjust its parameters, and increases the possibility of learning. Furthermore, the larger the size of the testing data set, the more acceptable the model performance ergo generalizability of the classifier is higher [43]. To meet this goal, four randomly generated different sized training data sets, including 80, 60, 40 and 20% of total patterns, were used (Table 2). Model performance was evaluated by accuracy estimates based on different training data set sizes. A flowchart was created to describe the objectives of this article (Fig. 7). The penultimate objective of this work is to isolate the best classifier with the highest accuracy in distinguishing between male and female pistachio trees based on geometric characteristics (LW1, LW2) of leaves. The study began with optimization of important parameter values for each classifier. Five classifier types (KNN, SVM, SOM, ANFIS, and RBF) were assessed using different data sets (Table 2). Ultimately, the classifier with the highest accuracy is selected for two classes (male, female) in pistachio.

**Table 1**
A classifier confusion matrix.

| | | Predicted sex | |
|---|---|---|---|
| | | Positive (Female) | Negative (Male) |
| Actual sex | Positive (Female) | True positive (TP) | False negative (FN) |
| | Negative (Male) | False positive (FP) | True negative (TN) |

## 3. Results and discussion

Optimum values are calculated based on output from the various data sets before comparisons are made of classifier performances.

### 3.1. Adjusting parameters of the classifiers

#### 3.1.1. KNN classifier

KNN can categorize pistachios with the highest probability into a K group based on the nearest-neighbors classification rules in the training phase. Results acquired from the KNN classifier with K-neighborhood values that differ from one to eleven for the training and testing phases are shown in Fig. 8. These results indicate size of the training and testing data sets can drastically impact performance and accuracy of the classifier. Different data sets were used in two training and testing phases to evaluate reliability and generalization (Fig. 8). The minimum classification accuracy in the training phase with 20, 40, 60 and 80% data set sizes was 95%. The KNN classifier was able to learn and apply classification rules, however, testing phase results indicated KNN classifier accuracy varied from 52 to 60%. While the KNN classifier is able to be trained, it is unable to accurately classify sex in pistachio genotypes. Further use of KNN for generalization of pistachio sex must be carefully pondered. KNN's best performance was obtained for K = 1 for the 80% training data set and K = 11 for other sizes of the training data set. KNN classifier performance is dependent upon K-neighborhood value [46,47]. The KNN classifier correctly identified 340 (of 400 male) and 452 (of 500 female) pistachio genotypes within the 80% training data set, a classification accuracy of 88%. A reduction in training data set size, led to decreased classifier accuracy. For example, 185 of 400 males and 117 of 500 females were incorrectly identified in the 20% training data set. KNN classifier performance in the training phase was generally more accurate than the testing phase [43] however the subpart test results, which are vital, can call into doubt the ability of KNN to correctly determine pistachio sex. Additional classifier options must be evaluated.

#### 3.1.2. SVM classifier

SVM was the second classifier selected for use for pistachio sex classification using female and male leaf morphological traits. SVM classification using the 20, 40, 60 and 80% training data sets was presented for the two training and testing phases (Fig. 9). Kernel function type is one of the most important SVM design factors. The data presented here highlight the inconsistencies that can result when using various design factors and sizes of the training data set. The RBF function type of SVM produced better results than other kernel function types in the 20, 40 and 80% data set sizes. However, a greater accuracy was achieved with polynomial function (Poly2) in the 60% data set size. SVM classification performance was similar in both the training and testing phases. The lowest accuracy percentages for the SVM training and testing phases equaled 63% and 59%, respectively.

Comparison of KNN and SWM results indicated that, despite the poor performance of SVM in the training phase, it outperformed KNN during the testing phase. The lowest percent accuracy of classification for all training data set sizes during the testing phase with SVM was 61%. For example, 222 (of 400 male) and 75 (of 500 female) were incorrectly

**Table 2**
Training and testing set sizes for each class.

| Category | Train 80% | Test 20% | Train 60% | Test 40% | Train 40% | Test 60% | Train 20% | Test 80% | Total |
|---|---|---|---|---|---|---|---|---|---|
| Class I (Males) | 320 | 80 | 240 | 160 | 160 | 240 | 80 | 320 | 400 |
| Class II (Females) | 400 | 100 | 300 | 200 | 200 | 300 | 100 | 400 | 500 |
| Total | 720 | 180 | 540 | 360 | 360 | 540 | 180 | 720 | 900 |

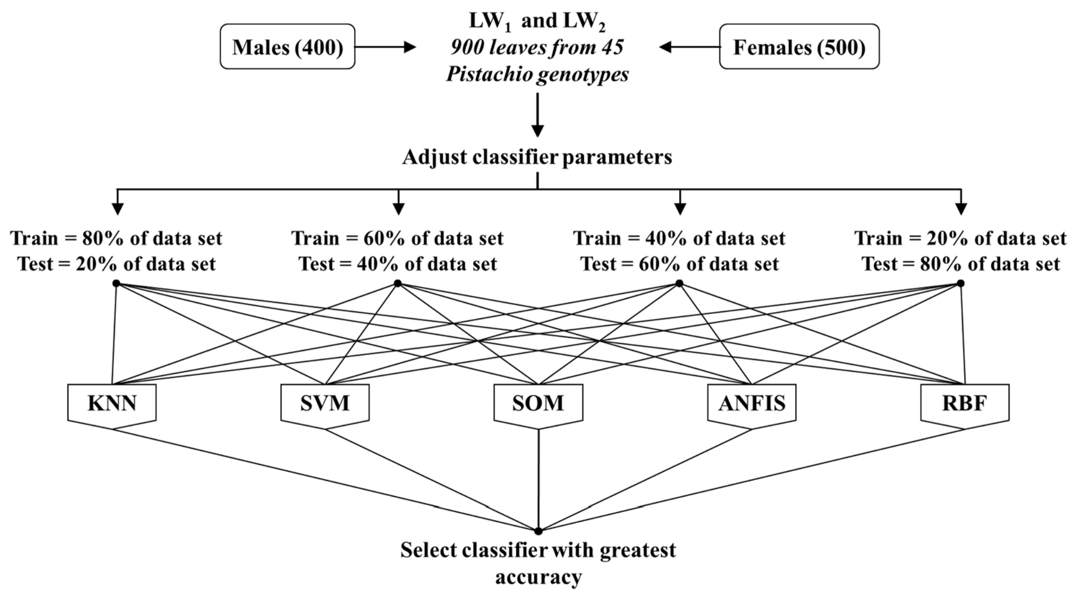


**Fig. 7.** Project flowchart. Representation of action steps taken to complete this work.

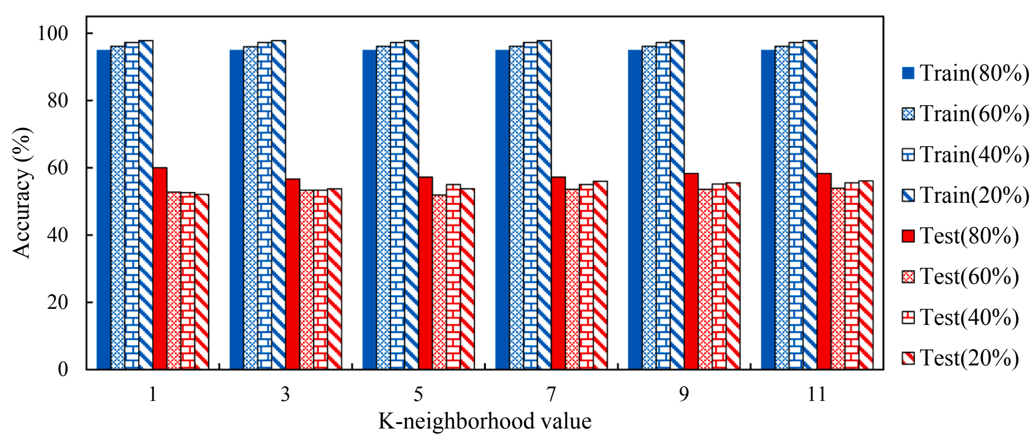


**Fig. 8.** Mean accuracy of KNN for K with different training data set sizes during the training and testing phases of sex classification in pistachio genotypes.

identified in the 80% training data set.

*3.1.3. SOM classifier*

The KNN and SVM classifiers had advantages and disadvantages but neither were reliable for pistachio sex classification. Therefore, the third classifier tested for suitability was SOM. The SOM classifier is an artificial neural network method that can classify input features based on how they grouped in the input space. SOM size can substantially alter classification performance [48]. Therefore, several different SOM topology dimensions ($3 \times 20$, $5 \times 12$, $6 \times 10$, $6 \times 20$, and $12 \times 10$) were chosen for assessment of accuracy in pistachio sex classification (Fig. 10). The $6 \times 10$ topology was more accurate during the testing phase than the other selected topologies. The generated data also showed that 1500 iterations was an appropriate number of epochs for training the SOM. Also, topology is affected by updating SOM weights in the output layer [49] and controls how neurons communicate.

Three distinct topologies hexagonal, rectangular, and random were used in this study. Results were obtained from four distance functions (Link, Box, Euclidean, and Manhattan) for the four different training data set sizes (Fig. 11). Topology and function performance can be affected by training data set size. The Link distance function was superior in the 20% and 80% training data set while the Box function was most accurate in 40% and 60% training data set sizes (Fig. 11). The type of topology to use is dependent upon the problem being studied [50]. Here, the best classification accuracy in the two training and testing phases among topologies was noted for random topology (Fig. 11). Minimum and maximum accuracy of SOM classification in the training phase varied from 64% to 72% across the different training data sets and

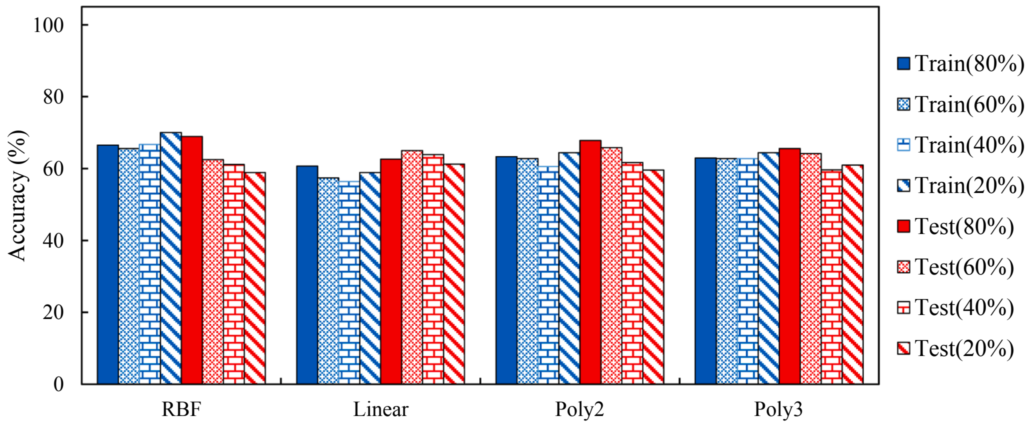


**Fig. 9.** Evaluation of specific SVM kernel functions with different training data set sizes during the training and testing phases of sex classification in pistachio genotypes.

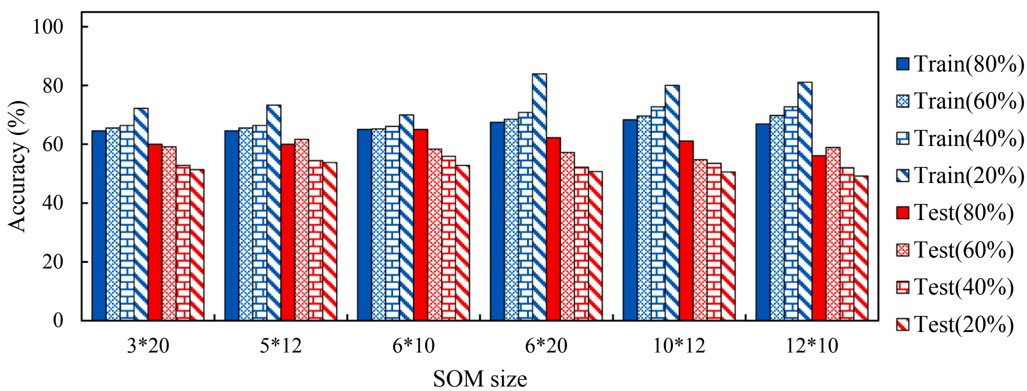


**Fig 10.** Use of selected SOM topologies with different training data set sizes during the training and testing phases of sex classification in pistachio genotypes.

from 52% to 66% during the testing phase. The SOM classifier correctly classified 196 (of 400 male) and 389 (of 500 female) pistachio genotypes within the 80% training data set. Despite optimization, the SOM neural network generated a classification accuracy of only 65%.

### 3.1.4. ANFIS classifier

The ANFIS classifier was applied to this pistachio sex classification study as this method can simultaneously combine artificial neural network (ANN) principles and a fuzzy inference system (FIS). Preparation and construction of FIS can influence classifier performance. Accuracy of pistachio sex classification based on three FIS construction methods including grid partition (GP), subtractive clustering (SC) and fuzzy c-means (FCM) was shown in Fig. 12 for the four training data set sizes in the training and testing phases. Maximum accuracy in the 20, 40 and 80% data set were obtained with the GP method (Fig. 12) while the FCM method displayed the highest accuracy in the 60% training data set size. Minimum and maximum accuracy of sex classification in the training phase were 63% and 68%, respectively. Accuracy within the testing phase ranged from 59 to 66%. These results indicated ANFIS was able to correctly classify 180 (of 400 male) and 424 (of 500 female) pistachio sex class patterns. Thus, this method was 67% accurate in the best conditions.

### 3.1.5. RBF classifier

The RBF neural network was the fifth and final classifier selected for use in identification of male and female pistachio genotypes based on leaf characteristics. The number of neurons in the RBF hidden layer can have a major impact on the accuracy of pistachio sex detection. Accuracy changes in terms of the neuron numbers in the RBF hidden layer during the training and testing phases with the four training data set sizes are shown in Fig. 13. As neuron numbers within the hidden layer increased, overall accuracy increased (Fig. 13). However, accuracy for the two training and testing phases varied with data set size. In the 60% and 80% training data set sizes, an increase in hidden layer neuron numbers to 100 and 150 neurons, increased accuracy before leveling off. Selecting 100 hidden layer neurons for the 20% and 40% training data set sizes yielded the most promising results despite a slight reduction in testing phase classification accuracy. Minimum and maximum accuracy within the training phase was 80% and 84% and 77% and 88% within the testing phase, respectively. Compared to the other four classifier models, these results are superior. In the 80% training data set size, the RBF correctly classified 298 (of 400 male) and 442 (of 500 female) pistachio sex class patterns. This corresponded to an overall accuracy of 82%.

## 3.2. Comparison of classifier performance

The various classifiers used in this study and the values obtained for optimal performance in accurately determining pistachio sex are discussed in the opening results section. Therefore, based on these optimal values for classifier parameters, accuracy KNN, SVM, SOM, ANFIS, and RBF for the different data set sizes (80, 60, 40, and 20%) are shown in Table 3. The results show a greater accuracy for KNN and RBF in the training phase compared to the other classifiers. However, RBF and SVM have higher detection accuracy and generalizability in the testing phase. It should be noted that the greater accuracy of KNN for all data set sizes primarily stems from a highly accurate training phase. Accuracy of KNN classification in the testing phase is low, leading to a lower

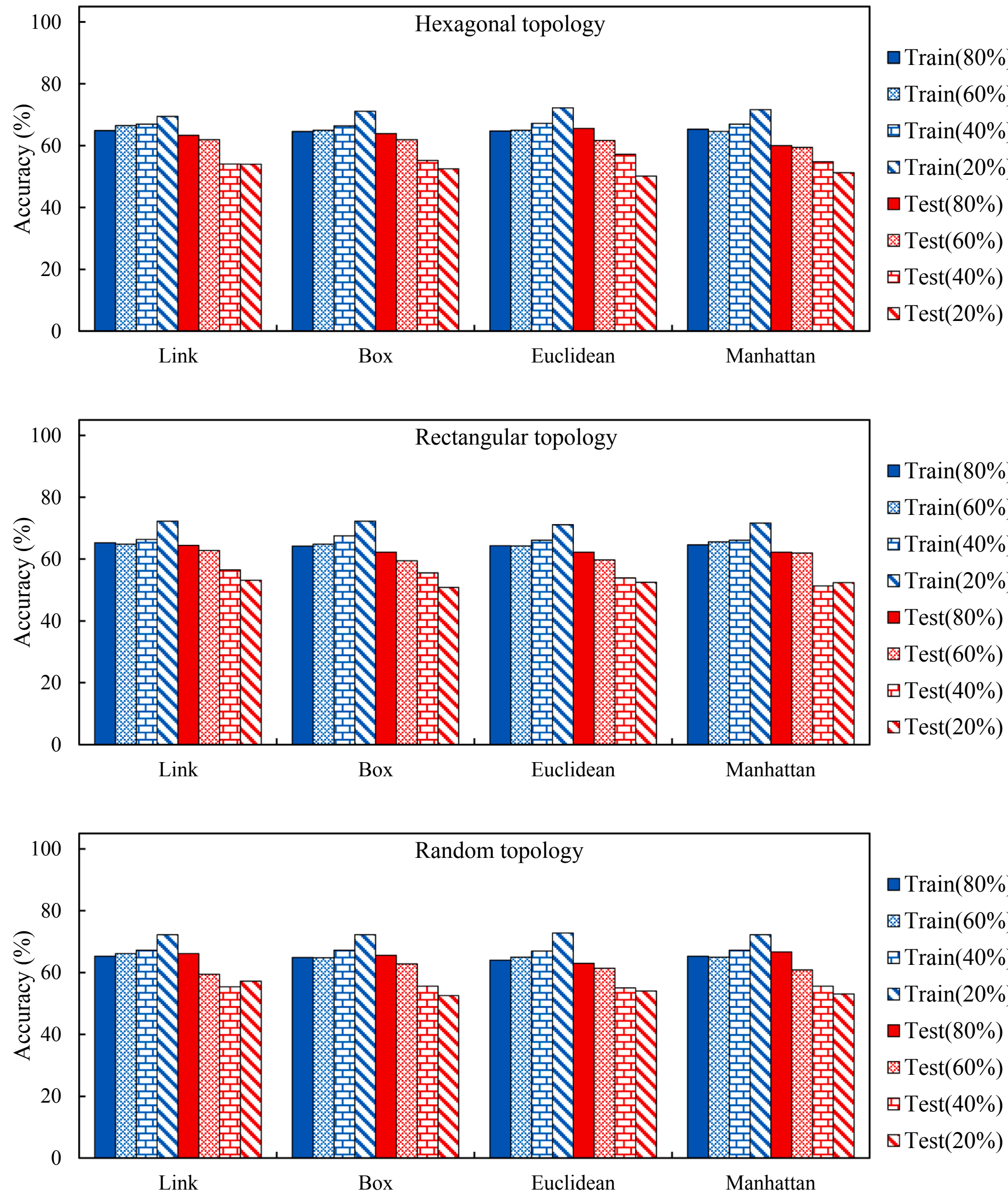


**Fig. 11.** Accuracy of pistachio sex classification using four distance function types (Link, Box, Euclidean, and Manhattan) and three topologies (hexagonal, rectangular and random) for different training data set sizes during the training and testing phases.

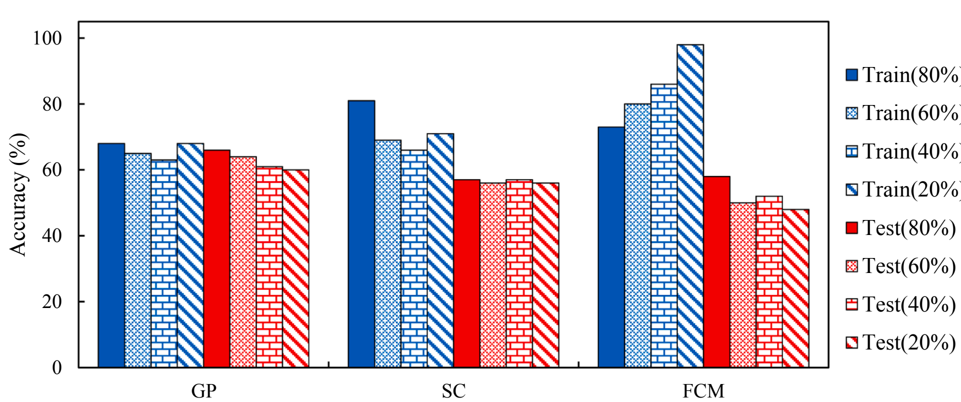


**Fig. 12.** Accuracy of pistachio sex classification using three FIS construction methods (GP, SC, and FCM) different training data set sizes during the training and testing phases. (GP, grid partition; SC, subtractive clustering; and FCM, fuzzy c-means).

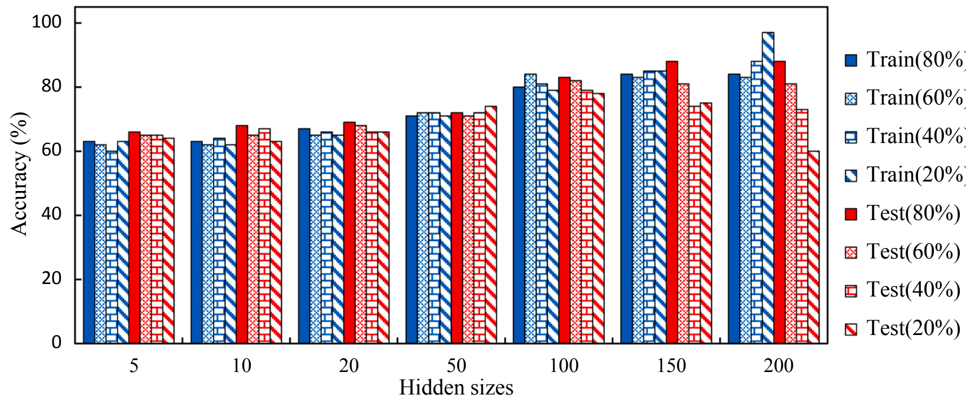


**Fig. 13.** Percent accuracy of neuron numbers in the hidden layer of the RBF neural network with different training data set sizes during the training and testing phases of sex classification in pistachio genotypes.

**Table 3**
Classifier accuracy values for the training and testing phases at all data set sizes.

| Phase | Data Set Size (% of data) | Percent Accuracy for Individual Classifiers | | | | |
|---|---|---|---|---|---|---|
| | | KNN | SVM | SOM | ANFIS | RBF |
| Train | 80 | 95.00 | 66.53 | 64.72 | 67.08 | 83.47 |
| Test | 20 | 60.00 | 68.89 | 66.11 | 66.30 | 88.33 |
| All | 100 | 88.00 | 67.00 | 65.00 | 67.11 | 84.44 |
| Train | 60 | 96.11 | 62.78 | 65.74 | 64.63 | 84.07 |
| Test | 40 | 57.78 | 65.83 | 61.94 | 64.07 | 82.22 |
| All | 100 | 80.78 | 64.00 | 64.22 | 64.11 | 83.33 |
| Train | 40 | 97.22 | 66.67 | 66.11 | 63.33 | 80.83 |
| Test | 60 | 58.70 | 61.11 | 57.96 | 60.62 | 79.44 |
| All | 100 | 74.11 | 63.33 | 61.22 | 61.67 | 80.00 |
| Train | 20 | 97.78 | 70.00 | 71.67 | 68.33 | 79.44 |
| Test | 80 | 58.61 | 58.89 | 51.67 | 58.80 | 77.78 |
| All | 100 | 66.44 | 61.11 | 55.67 | 60.67 | 78.11 |

generalizability than other methods.

Comparing the performance of five classifiers in two training and testing phases in four different sizes of training data sets, the RBF neural network can be considered the best way to determine the sex of pistachios based on leaf characteristics. RBF correctly identified 238 of 320 males and 350 of 400 females at 80% of the training data set. Accuracy for RBF is similar in the training and testing phases for all data set sizes. A data set size reduction from 80% to 40% only decreased accuracy 4% (from 84% to 80%). Remarkably, the RBF neural network was able to distinguish pistachios sex based exclusively on leaf dimensions. In the same vein, should additional morphological leaf traits (leaf shape, shape of tip and base in leaflet, leaf margin) be added to the data, accuracy of this method would further improve. Heidari, Rohani and Rezaei [26] used the RBF neural network to correctly identify pistachio kernel color with 98.95% accuracy by utilizing eight pistachio leaf morphological feature measurements as input data. Several DNA markers were developed for early sex detection in pistachio [11,12,51,52]. The first of these sex-linked DNA markers, developed on RAPD (OPO08$_{945}$), showed error rates of 15 to 20% [12,38]. The error rate reported here for sex determination using RBF modeling was similar with ~88% accuracy. Recently developed SNP and SCAR markers have reported the ability to detect the sex in the early stages of pistachio seedling growth [10,51,52]. However, false negatives in some female individuals and false positives in some male individuals with use of SCAR markers have also been reported [51]. New sex-linked DNA markers are very effective for determination of pistachio sex but require much higher overhead costs to accomplish the feat (lab facilities, equipment, and expertise required for DNA testing).

Machine learning models based on shape, texture, and leaf images have been used for plant classification and identification [53,54], detecting plant diseases [55–57], cultivar and species determination [58] and genetic value [26] however there have been no studies on sex identification in synoecious plants. Several classifier types were used in these studies with the goal of selecting the best choice based on some performance criteria. Identification and classification of apple neutrosophic leaf diseases based on leaf images was investigated with nine different classifiers and the Random forest (RF) classifier model, with 98.4% accuracy, was offered as the best deep learning model [56]. The SVM model, with 97.92% accuracy, provided the best results with leaf characteristics for a study by Lu, Jiang, Ghiassi, Lee and Nitin [54] in *Camellia* (Theaceae). A study in grapevine cultivars obtained 94% accuracy from leaf morpho-colorimetry data [58] with ANN. Citrus classification with different deep learning models and a MLP (multilayer perception) neural network achieved 98.14% accuracy in a multi classifier study of fused leaf features [59]. Classifier performance is often dependent on the specific type of problem being investigated. Classifiers are most often evaluated by incorporating the collected data (100% data set size) into the training and testing phases. Few researchers consider the generalization capability gained by altering training data set size. Moreover, as demonstrated here, training and testing phase accuracy for the assorted classifiers varied and also showed considerable deviation when linked to training and testing data set size.

### 3.3. Sensitivity analysis

Accuracy results showed the RBF neural network was the best choice of the classifiers compared in this study of pistachio sex determination based on leaf characteristics. A sensitivity analysis of the RBF classifier was created from two very important parameters, number of neurons in the hidden layer (hidden sizes) and the spread parameter for the two phases of training and testing (Fig. 14). The RBF (L) hidden layer neuron numbers can substantially effect predictive function. Increased hidden layer neuron numbers led to improved classification accuracy, network size, computations required in the training phase, and training phase duration (Fig. 14**a**). Increasing neuron numbers to more than 100, had little positive effect on RBF function. Therefore, 100 hidden layer neurons were used as the optimal number for the RBF neural network. Value of the spread parameter (σ) also plays a critical role in determining RBF classification accuracy [60]. If σ increases, classification accuracy will have a downward trend (Fig. 14**b**). The same result is obtained if a very small σ value is used. On the other hand, the two parameters L and σ can interact. Therefore, this issue should be considered before selecting their values. In our experience, if optimal values for these two parameters are found, classification accuracy of the RBF neural network needs to provide more information than the LW1 and LW2 properties for both male and female pistachio genotypes. As such, more accurate neural network classifications with these two features alone are not likely. Heidari,

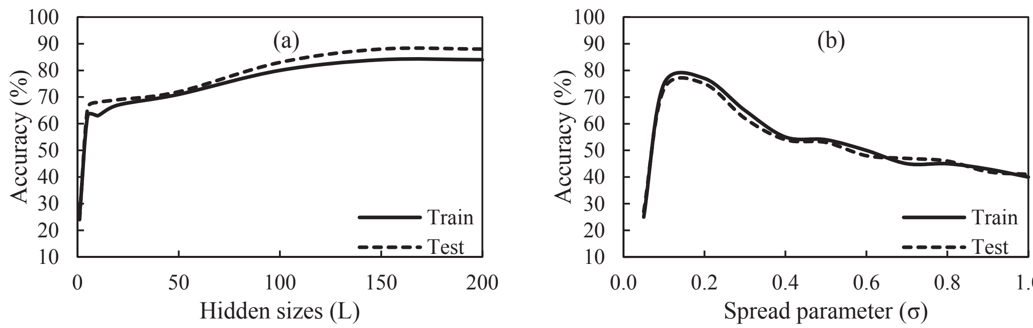


**Fig. 14.** Sensitivity analysis. Classification accuracy with (a) hidden sizes and (b) spread parameters.

Rohani and Rezaei [26] showed that eight morphological features of pistachio leaves used as input data for the RBF neural network resulted in a 98.95% classification accuracy for pistachio kernel color detection.

## 4. Conclusion

Classification of the sex in the early stages of growth for a dioecious tree such as pistachio is exceedingly difficult for pistachio growers and breeders. Males and females are indistinguishable until maturation and the initiation of inflorescence growth or with the aid of costly and time-consuming DNA tests. This study used five different soft computing methods to classify pistachio sex with only leaf morphological traits from early developmental stages as input data. The classification methods included KNN, SVM, SOM, ANFIS and RBF. To obtain a more accurate estimation of classifier performance, the optimal values for the parameters affecting each classifier were computed and evaluated. To gage generalizability of each classifier, different training data set sizes (20, 40, 60 and 80%) were used. The results showed that the accuracy and effectiveness of each classification method can vary depending on the size of the training data set. The performance of each classifier is also significantly affected by parameter selection. The KNN classifier had the greatest capacity for training in pistachios sex determination. Training abilities of the remaining classifiers can be arranged RBF, ANFIS, SVM, SOM, respectively. In addition to showing a good capacity for training, classifiers must demonstrate accuracy in the test phase as well. Reliability of these methods for practical applications is directly related to test phase accuracy. With the 80% training data set as our baseline, our results showed the overall classification accuracy of the five methods as: RBF (88.33%), SVM (68.89%), ANFIS (66.3%) and SOM (66.11%). Therefore, in our experience, KNN and RBF were the best of the five classifiers for resolving the sex of pistachio genotypes. Although KNN has good training ability, RBF was the best overall choice with 84.44% classification accuracy in the testing and 88.33% accuracy in the training phase. Only two morphological leaf traits were used to obtain 84% accuracy in pistachio sex determination. Therefore, soft computing models can be used to identify the sex of pistachios when provided with leaf morphological data. This study provides an efficient and accurate alternative to DNA marker testing, and can greatly increase selection efficiencies for pistachio growers. The ability to identify male and female pistachio genotypes before maturation substantially reduces the costs and time commitment required of breeders. As previously mentioned, only two leaf characteristics were used for pistachio sex determination in this work. Additional traits such as leaflet shape, shape of tip and base in leaflet, and leaf margin be added to the collected. It is suggested that supplementary methods such as deep learning be assessed for usefulness in answering further questions regarding pistachio sex. A more extensive collection of pistachio genotype biodiversity could be compiled and used to corroborate the results presented here.

## CRediT authorship contribution statement

Visualization, Writing - review & editing, Investigation.

## 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.

## Appendix:. A MATLAB source code for classifiers

```
x;% input vector
Fx;% output vector
[xn,xs] = mapminmax(x,-1,1); % Normalization
[trainInd,valInd,testInd] = dividerand(900 ,0.80 ,0 ,0.20);
1. RBF classifier
hiddenSizes = 100;
spread = 0.125;
net = newrb(xn(:,trainInd),Fxn(:,trainInd),spread,hiddenSizes);
class_predict = sim(net,xn);%Network output
2. ANFIS classifier
genOpt = genfisOptions('GridPartition');
genOpt.NumMembershipFunctions = [45];
genopt.InputMembershipFunctionType='trapmf';
inFIS = genfis(xn',Fx',genOpt);
opt = anfisOptions('InitialFIS',inFIS,'EpochNumber',20);
fis = anfis(TrainData,opt);
class_predict = evalfis(xn',fis);
3. SVM classifier
classes = unique(Fx);
for j = 1 : 2
indx = (Targets(:,trainInd)==classes(j))'; SVMModels{j} = fitcsvm
(Features(:,trainInd)',indx,'solver','SMO',
'KernelFunction','rbf','KernelOffset',0.15,'KernelScale',1,'BoxCon-
straint',50);
end
for j = 1 : n_class

[ ~ , score ] = predict(SVMModels{ j } , xn');
Scores(: , j) = score(: , 2);

end
[ ~ , class_predict ] = max(Scores , [ ] , 2);
4. SOM classifier
net = selforgmap([610],50,5,'randtop','linkdist');
net.layers{1}.transferFcn = 'compet';
net = configure(net,xn(:,trainInd));
[ net, resultout] = train(net, xn(:,trainInd));
class_predict = sim(net ,xn(:,trainInd));
5. KNN
```

KNN = fitcknn(xn(:,trainInd)',Fx(:,trainInd)','BreakTies','nearest',' BucketSize', 10,'NSMethod','kdtree','Distance','minkowski','NumNeighbors',25);
class_predict = predict(KNN , xn')';

## References


[1] J.I. Hormaza, V.S. Polito, Pistillate and staminate flower development in dioecious *Pistacia vera (Anacardiaceae)*, Am. J. Bot. 83 (1996) 759–766.
[2] M. Khorasani, H. Nosrati, A. Razbane Haghighi, S. Kelig, Study of morphological diffrentiation of leaf in males and females of *Pistacia atlantica* desf species in Arasbaran forests, Iranian, J. Plant. Res. 27 (2015) 605–612.
[3] Q. Bai, Z. Ma, Y. Zhang, S. Su, P. Leng, The sex expression and sex determining mechanism in *Pistacia* species, Breeding Sci. 69 (2019) 205–214.
[4] X.Y. Wang, The comparative study on biological characteristic between female and male individuals of *Pistacia chinensis* Bunge, Henan Univ. Sci. Technol. Diss, 2013.
[5] O. Correia, M.C.D. Barradas, Ecophysiological differences between male and female plants of *Pistacia lentiscus* L, Plant Ecol. 149 (2000) 131–142.
[6] S. Jonasson, H. Medrano, J. Flexas, Variation in leaf longevity of *Pistacia lentiscus* and its relationship to sex and drought stress inferred from leaf δ13C, Funct. Ecol. 11 (1997) 282–289.
[7] D.M. Minor, R.K. Kobe, Fruit production is influenced by tree size and size-asymmetric crowding in a wet tropical forest, Ecol. Evol. 9 (2019) 1458–1472.
[8] S.C.H. Barrett, J. Hough, Sexual dimorphism in flowering plants, J. Exp. Bot. 64 (2013) 67–82.
[9] M.B. García, R.J. Antor, Sex ratio and sexual dimorphism in the dioecious *Borderea pyrenaica (Dioscoreaceae)*, Oecologia 101 (1995) 59–67.
[10] B. Esfandiyari, G.H. Davarynejad, F. Shahriari, M. Kiani, A. Mathe, Data to the sex determination in *Pistacia* species using molecular markers, Euphytica 185 (2012) 227–231.
[11] S.K. Salih Kafkas, G.D. Selim Cetiner, R. Perl-Treves, A.N. Ada Nissim-Levi, Development of sex-associated RAPD markers in wild Pistacia species, J. Horticult. Sci. Biotechnol. 76 (2001) 242–246.
[12] J.I. Hormaza, L. Dollo, V.S. Polito, Identification of a RAPD marker linked to sex determination in *Pistacia vera* using bulked segregant analysis, Theor. Appl. Genet. 89 (1994) 9–13.
[13] S.-H.M. Ashtiani, A. Rohani, M.H. Aghkhani, Soft computing-based method for estimation of almond kernel mass from its shell features, Sci. Hortic. 262 (2020), 109071.
[14] K.C. Saranya, A. Goyal, Automatic Plant Leaf Classification Based on Back Propagation Networks for Medical Applications, Soft Computing for Problem Solving, Springer, 2020, pp. 981–991.
[15] R. Nathan, O. Spiegel, S. Fortmann-Roe, R. Harel, M. Wikelski, W.M. Getz, Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: general concepts and tools illustrated for griffon vultures, J. Exp. Biol. 215 (2012) 986–996.
[16] H.S. Khamis, K.W. Cheruiyot, S. Kimani, Application of k-nearest neighbour classification in medical data mining, Int. J. Inform. Commun. Technol. Res. (2014).
[17] S.A. Medjahed, T.A. Saadi, A. Benyettou, Breast cancer diagnosis by using k-nearest neighbor with different distances and classification rules, Int. J. Comput. Appl. 62 (2013).
[18] K. Manimekalai, M. Vijaya, Taxonomic classification of Plant species using support vector machine, J. Bioinform. Intell. Control 3 (2014) 65–71.
[19] S. Mohapatra, D. Patra, S. Satpathy, An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images, Neural Comput. Appl. 24 (2014) 1887–1904.
[20] N. Hashim, S.E. Adebayo, K. Abdan, M. Hanafi, Comparative study of transform-based image texture analysis for the evaluation of banana quality using an optical backscattering system, Postharvest Biol. Technol. 135 (2018) 38–50.
[21] D. Chen, R. Shi, J.-M. Pape, K. Neumann, D. Arend, A. Graner, M. Chen, C. Klukas, Predicting plant biomass accumulation from image-derived parameters, GigaScience 7 (2018) giy001.
[22] H.Z. Abyaneh, Evaluation of multivariate linear regression and artificial neural networks in prediction of water quality parameters, J. Environ. Health Sci. Eng. 12 (2014) 40.
[23] M. Ali, R.C. Deo, Modeling wheat yield with data-intelligent algorithms: artificial neural network versus genetic programming and minimax probability machine regression, Handbook of Probabilistic Models, Elsevier, 2020, pp. 37–87.
[24] K. Mochida, S. Koda, K. Inoue, T. Hirayama, S. Tanaka, R. Nishii, F. Melgani, Computer vision-based phenotyping for improvement of plant productivity: a machine learning perspective, GigaScience 8 (2019) giy153.
[25] S. Nagai, H. Morimoto, T.M. Saitoh, A simpler way to predict flowering and full bloom dates of cherry blossoms by self-organizing maps, Ecol. Inf. 56 (2020), 101040.
[26] P. Heidari, A. Rohani, M. Rezaei, Soft computing-based approach on prediction promising pistachio seedling base on leaf characteristics, Sci. Hortic. 274 (2020), 109647.
[27] A. Dhomne, R. Kumar, V. Bhan, Gender recognition through face using deep learning, Procedia Comput. Sci. 132 (2018) 2–10.
[28] H. Wang, H. Su, P. Chen, R. Hou, Z. Zhang, W. Xie, Learning deep features for giant panda gender classification using face images. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019.
[29] S. Hosseini, Genotype by temperature interaction effects on sex determination in zebrafish (*Danio rerio*), Niedersächsische Staats-und Universitätsbibliothek Göttingen (2019).
[30] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley New York, 1973.
[31] J. Wang, P. Neskovic, L.N. Cooper, Neighborhood size selection in the k-nearest-neighbor rule using statistical confidence, Pattern.Recogn. 39 (2006) 417–423.
[32] A.K. Agrawala, Machine Recognition of Patterns, IEEE Press, New York, 1977.
[33] E. Fix, J.L. Hodges, Discriminatory analysis. Nonparametric discrimination: consistency properties, Int. Stat. Rev./Revue Internationale de Statistique 57 (1989) 238–247.
[34] A.K. Ghosh, P. Chaudhuri, C. Murthy, On visualization and aggregation of nearest neighbor classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1592–1602.
[35] T. Kohonen, The self-organizing map, Proc. IEEE 78 (1990) 1464–1480.
[36] J.S. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man Cybernet. 23 (1993) 665–685.
[37] Y. Wu, H. Wang, B. Zhang, K.L. Du, Using radial basis function networks for function approximation and classification, ISRN Appl. Math. 2012 (2012), 324194.
[38] F. Kamiab, A. Ebadi, B. Panahi, A. Tajabadi, RAPD Analysis for Sex Determination in *Pistacia vera* L, J. Nuts 5 (2014) 51–55.
[39] D.M. Farid, M.A. Al-Mamun, B. Manderick, A. Nowe, An adaptive rule-based classifier for mining big biological data, Expert Syst. Appl. 64 (2016) 305–316.
[40] V. Vapnik, The Nature of Statistical Learning Theory, Springer Science & Business Media, 2013.
[41] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Mach. Learn. 46 (2002) 389–422.
[42] Q. Shen, W.-M. Shi, W. Kong, B.-X. Ye, A combination of modified particle swarm optimization algorithm and support vector machine for gene selection and tumor classification, Talanta 71 (2007) 1679–1683.
[43] A. Rohani, M. Mamarabadi, Free alignment classification of dikarya fungi using some machine learning methods, Neural Comput. Appl. 31 (2019) 6995–7016.
[44] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006) 861–874.
[45] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1950) 32–35.
[46] R. Kumar, M.K. Goyal, P. Ahmed, A. Kumar, Unconstrained handwritten numeral recognition using majority voting classifier, in: 2012 2nd IEEE International Conference on Parallel Distributed and Grid Computing (PDGC), IEEE, 2012, pp. 284–289.
[47] N. Jafari, A. Chodorowski, Histology-based oral lesion classification, in: 2012 20th Iranian Conference on Electrical Engineering (ICEE), IEEE, 2012, pp. 1612–1617.
[48] S. Delgado, F. Morán, A. Mora, J.J. Merelo, C. Briones, A novel representation of genomic sequences for taxonomic clustering and visualization by means of self-organizing maps, Bioinformatics 31 (2015) 736–744.
[49] Y. Nam, S.-H. Koh, S.-J. Jeon, H.-J. Youn, Y.-S. Park, W.I. Choi, Hazard rating of coastal pine forests for a black pine bast scale using self-organizing map (SOM) and random forest approaches, Ecol. Inf. 29 (2015) 206–213.
[50] A.J. George, G. Gopakumar, M. Pradhan, K.A. Nazeer, M.J. Palakal, A self organizing map-harmony search hybrid algorithm for clustering biological data, in: 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES), IEEE, 2015, pp. 1–5.
[51] S. Kafkas, M. Khodaeiaminjan, M. Güney, E. Kafkas, Identification of sex-linked SNP markers using RAD sequencing suggests ZW/ZZ sex determination in Pistacia vera L, BMC Genomics 16 (2015) 98.
[52] M. Khodaeiaminjan, E. Kafkas, M. Güney, S. Kafkas, Development and linkage mapping of novel sex-linked markers for marker-assisted cultivar breeding in pistachio (*Pistacia vera* L.), Mol. Breed. 37 (2017) 98.
[53] N.R. da Silva, J.B. Florindo, M.C. Gómez, D.R. Rossatto, R.M. Kolb, O.M. Bruno, Plant identification based on leaf midrib cross-section images using fractal descriptors, PLoS ONE 10 (2015).
[54] H. Lu, W. Jiang, M. Ghiassi, S. Lee, M. Nitin, Classification of Camellia (Theaceae) species using leaf architecture variations and pattern recognition techniques, PLoS ONE 7 (2012).
[55] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural networks based recognition of plant diseases by leaf image classification, Comput. Intell. Neurosci. 2016 (2016).
[56] G. Dhingra, V. Kumar, H.D. Joshi, A novel computer vision based neutrosophic approach for leaf disease identification and classification, Measurement 135 (2019) 782–794.
[57] E. Omrani, B. Khoshnevisan, S. Shamshirband, H. Saboohi, N.B. Anuar, M.H.N. M. Nasir, Potential of radial basis function-based support vector regression for apple disease detection, Measurement 55 (2014) 512–519.
[58] S. Fuentes, E. Hernández-Montes, J.M. Escalona, J. Bota, C.G. Viejo, C. Poblete-Echeverría, E. Tongson, H. Medrano, Automated grapevine cultivar classification based on machine learning using leaf morpho-colorimetry, fractal dimension and near-infrared spectroscopy parameters, Comput. Electron. Agric. 151 (2018) 311–318.
[59] L. Li, Y. Wu, M. Ye, Experimental comparisons of multi-class classifiers, Informatica 39 (2015) 71.
[60] S. Ghosh-Dastidar, H. Adeli, N. Dadmehr, Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection, IEEE Trans. Biomed. Eng. 55 (2008) 512–518.