Identifying Two of Tomatoes Leaf Viruses Using Support Vector Machine

Usama Mokhtar, Mona A.S. Ali, Aboul Ella Hassanien and Hesham Hefny

Abstract One of the most harmful viruses is Tomato yellow leaf curl virus (TYLCV), which is widespread over the world with tomato yellow leaf curl disease (TYLCD). It causes some symptoms to tomato leaf such as upward curling and yellowing. This paper introduces an efficient approach to detect and identify infected tomato leaves with these two viruses. The proposed approach consists of four main phases, namely pre-processing, image segmentation, feature extraction, and classification phases. Each input image is segmented and descriptor created for each segment. Some geometric measurements are employed to identify an optimal feature subset. Support vector machine (SVM) algorithm with different kernel functions is used for classification. The datasets of a total of 200 infected tomato leaf images with TSWV and TYLCV were used for both training and testing phase. N-fold cross-validation technique is used to evaluate the performance of the presented approach. Experimental results showed that the proposed classification approach obtained accuracy of 90 % in average and 92 % based on the quadratic kernel function.

Keywords Image processing · K-Mean clustering algorithm geometric features · Support vector machine (SVM)

U. Mokhtar (✉) · H. Hefny
Inst. of Stat. Studies and Res. (ISSR), Cairo University, Cairo, Egypt
e-mail: usamamokhtar@yahoo.com

H. Hefny
e-mail: hehefny@hotmail.com

M.A.S. Ali
Faculty of Computers and Information, Minia University, Minya, Egypt
e-mail: mony_4it@yahoo.com

A.E. Hassanien
Faculty of Computers and Information, Cairo University, Cairo, Egypt
e-mail: aboitcairo@gmail.com

J.K. Mandal et al. (eds.), *Information Systems Design and Intelligent Applications*,
Advances in Intelligent Systems and Computing 339,
DOI 10.1007/978-81-322-2250-7_77

1 Introduction

Agriculture is one of the most serious for national income for most countries. Tomatoes are one of the most widely cultivated food crops throughout the world due to its high nutritive value. It contains a lot of vitamins and nutrients such that vitamin C. It occupies the fourth level between word vegetables. Egypt is one of the famous countries that interested in tomatoes cultivation. It ranked fifth among leader countries in the world [1].

Although farmers do great effort in selecting good seeds of plants and creating suitable environment for plants growing, there are a lot of diseases that effect on plants. There are many factors that cause plant diseases: Plant pathogens such as (fungi, Bacteria, and Virus diseases) are the principal reasons for plant disease. Also, there are some insect that feed on the parts of plant such as (sucking insect pests), and plant nutrition such as (lack of micro elements) also, have critical effect on plant growing [2]. Viral diseases are the most common diseases of tomatoes. The amount of damage they cause varies, depending on the particular virus or combination of viruses present, the virulence of the virus strains, the susceptibility of the variety, the timing of infection, the abundance of insect vectors, and environmental conditions [3].

Now a days, Automatic detection of plant diseases attracts a lot of researcher in different domains because of there great benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves [4, 5]. Several approaches have been already introduced for image segmentation. K-means algorithm [6–8] is the most popular method for image segmentation because of its ability to cluster huge data points very quickly [9].

In [10], a new hybrid approach for image segmentation proposed for the automatic classification of leaf diseases based on high resolution multispectral and stereo images. They used Leaves of sugar beet for evaluating their approach. In [11], authors used computer image processing technique to introduce fast and accurate new method for grading of plant diseases. At the beginning, leaf region was segmented by using Otsu method [12, 13]. To detect the disease spot edges the disease spot regions were segmented by using Sobel operator. Finally, plant diseases are graded by calculating the quotient of disease spot and leaf areas.

Machine learning methods can successfully be applied as an efficient approach for disease detection. Many of these methods have been applied in agricultural researches. For example: Artificial Neural Networks (ANNs), Decision Trees, K-means, k-nearest neighbors. Support Vector Machines (SVMs) one of these approach that have been used extensively in this field. For example, in [14] authors used SVMs to identify visual symptoms of cotton diseases. In [6] authors presented an image recognition approach for wheat diseases detection using multi-class support vector machines (SVMs) after calculating features of diseased image regions of wheat disease was proposed. After calculating leaf image, image samples are trained and recognized using multi-class RBF SVM. In [15] authors employed the same classifier that trained and tested with some texture statistics features that have been computed for the useful segments.

In [1], an automated system has been developed to classify the leaf brown spot and the leaf blast diseases of rice plant based on the morphological changes of the plants caused by the diseases. Radial distribution of the hue from the center to the boundary of the spot images has been used as features to classify the diseases by Bayes' and SVM Classifier. In [16] author try to solve the difficulty of parameter determination in the original support vector machine (SVM) by using the genetic algorithm (GA) to select the parameters of the SVM automatically and the orthogonal method is utilized to determine the best GA parameters. In [17] Automated detection using vision system and pattern recognition are implemented to detect the symptoms of nutrient diseases and also to classify the disease group. In this paper, Support Vector Machine (SVM) is evaluated as classifier with four different kernels namely linear kernel, polynomial kernel with soft margin and polynomial kernel with hard margin.

This research aimed to develop a method that detect and identify type of virus that has infected tomato leaves. In order to decide if the infected tomatoes leaf is with TSWV or TYLCV image processing techniques have been used. The proposed approach is consisted of three main stages, image clustering, feature extraction and feature classification. To evaluate these results N-fold cross-validation has been used.

The remainder of this paper is ordered as follows. Section 2 review the basic conc algorithm. Section 3 shows the proposed new neutrosophic set. Section 3 describes the different phases of the proposed identifying two of tomatoes leaf viruses. Section 4 shows the experimental results and analysis. Finally, Conclusion and future work are discussed in Sect. 5.

2 Preliminaries

2.1 K-Means Clustering Algorithm

Image Segmentation is considered as one of the most important techniques for image analysis as well as in high-level image understanding images and extracting the information from them that can be used for different tasks such that robot vision, object recognition, and medical imaging. The goal of image segmentation is to partition an image into a set of disjoint regions with uniform and homogeneous attributes such as intensity, colour, tone or texture, etc. The image segmentation approaches can be divided into four categories: thresholding, clustering, edge detection and region extraction [18]. K-means clustering algorithm is the most popular method for image segmentation. K-Means algorithm is an unsupervised clustering algorithm that can use for classifying the input data into K groups that called clusters. These clusters are non-hierarchical and they do not overlap. Every member of a cluster is closer to its cluster than any other cluster.

2.2 The Support Vector Machine

The support vector machine (SVM) is a type of classifier that is originally a binary Classification method developed by Vapnik and colleagues at Bell laboratories [19, 20]. The main advantages of SVM are: It can obtain current optimal solution under finite samples; it can obtain the global optimal solution without falling into local optimums that normal algorithms have; it transforms nonlinear problems into linear problems in a higher dimension space, and the algorithm complexity is irrelated with space dimension [21]. To explain idea of SVM we will discuss the following two cases.

The Separable Case

Fully Linearly Separable For a binary classification problem with input space X and binary class Y where $y \in \{-1, 1\}$. There may exist many separating hyper-planes that correctly classify the data. The goal of SVM is selecting between them the one that maximizes the distance between the separating hyper-planes. The goal of SVM is to search for the optimal

$$w \cdot x + b = 0 \tag{1}$$

where w is normal to the hyperplane. Since SVM search for the separating hyperplane with largest margin. This can be formulated as follows:

$$y_i(x_i \cdot w + b) - 1 \geq 0 \; \forall \; i \tag{2}$$

Not Fully Linearly Separable

In order to extend the SVM methodology to handle data that is not fully linearly separable, we relax the constraint 2 slightly to allow for misclassified points as follow:

$$y_i(x_i \cdot w + b) - 1 \geq 0 + \in_i \quad \text{where} \; \in \geq 0 \; \forall \; i \tag{3}$$

3 The proposed identifying of tomatoes leaf viruses

In general, most computer vision algorithms share a common Framework. Figure 1 depicts the layout structure of a common image processing-based disease detection algorithm.

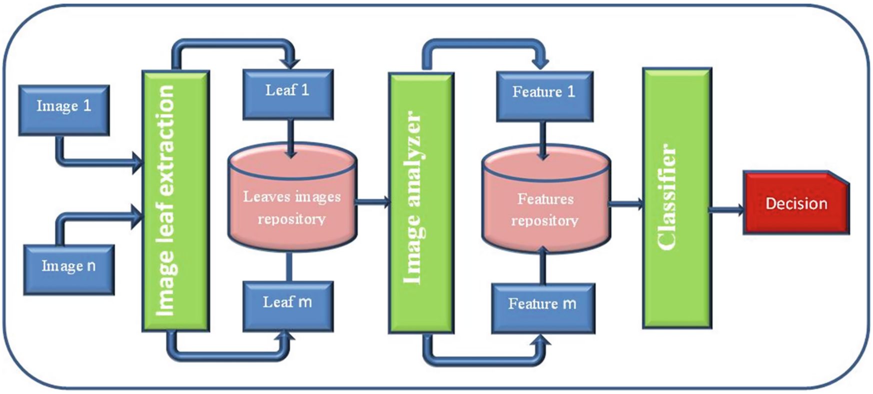

Fig. 1 Over all system structure

The algorithm begins with acquiring and collecting digital images from suitable environment. To prepare the acquired images to next step, image-processing techniques are applied such that image transformation, image resizing, image filtering etc. the next step is to segment image using a suitable image segmentation technique such that Clustering, edge detection, region growing to extract the infected part of leaf image. Then we extract useful features that are necessary for further analysis using suitable feature extraction techniques. These features may be color, shape texture features. After that, several analytical discriminating techniques are used to classify the images according to the specific problem. Figure 2 shows the overall architecture of the proposed identifying two of tomatoes leaf viruses using Support Vector Machine.

3.1 Image Acquisition

The first stage of this approach is the image acquisition stage. This phase plays an important role in any image classification system. We must select these images carefully to achieve the intended task in this approach. Science, the aim of this article is to distinguish between two types of virus, TSWV and TYLCV that infect tomatoes leaves, for this work we collected different specimens of infected tomatoes leaves for each type of virus from the enternet [22]. Since, some of these images contain more than one tomato leaves, therefore we addressed this issue in the preprocessing stage, for more details see the next section at the end of this phase we have 200 image of infected tomato leaves 100 for each virus type.

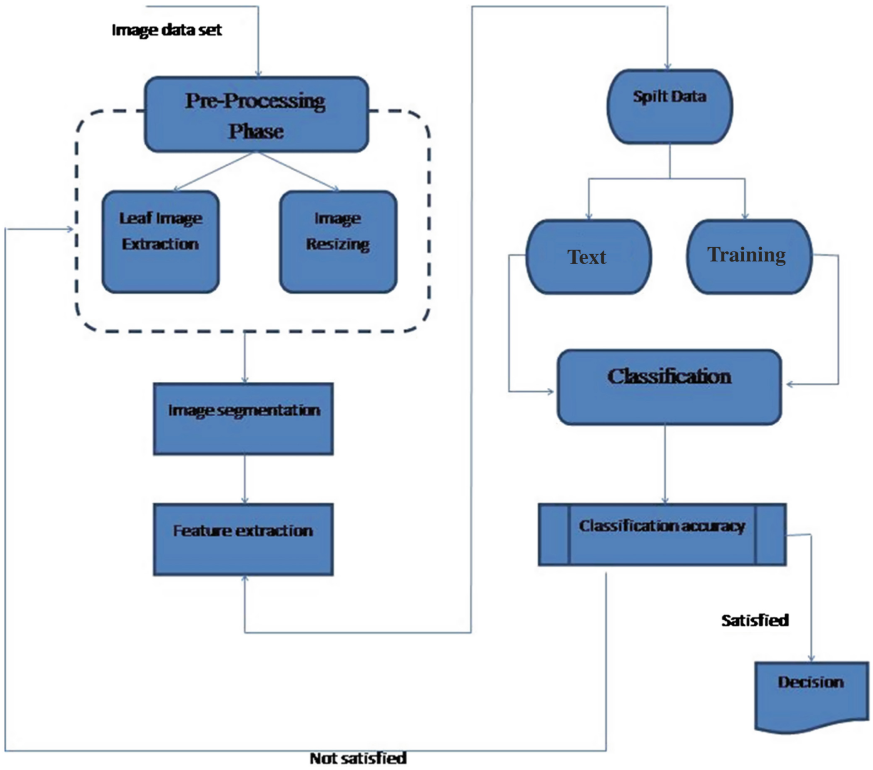

Fig. 2 The layout structure of the proposed system

3.2 *Pre-processing*

To increase the efficiency of the classification and prediction process, a pre-processing phase should be considered to enhance the quality of the input tomato leaves images and to make the feature extraction phase more reliable. it can be decomposed into the follow two steps:

Extracting image of tomato leaves from the acquired images
As we mention before, most of acquired images have more than one leaf image and may be other things, since the aim of this research is to identify type of virus that affect tomato leaf, we need to deal with single tomato leaf at time. In order to achieve that, these image have been manually cropped to extract a single leaf that saved in jpeg format.

Image Resizing
In this work, all images must be with the same size and equal dimensions. So, the gray image should be resized to equal dimensions.

3.3 Image segmentation phase

In this paper, in order to extract regions of interest, we have selected the most frequently used K-Mean clustering algorithm. Since it is challenging to visually distinguish between colors of leaf images, we have been used The L * a * b color space that also known as CIELAB or CIE L * a * b hence, it enables us to quantify these visual differences. The L * a * b color space is derived from the CIE XYZ tristimulus values. The L * a * b space consists of a luminosity L* or brightness layer, chromaticity layer a* indicating where color falls along the red-green axis, and chromaticity layer b* indicating where the color falls along the blue-yellow axis [23].

3.4 Feature Extraction Phase

After the leaf image has been segmented and isolated the infected parts of leaf, the next step is to extract some useful features. The purpose of feature extraction is to reduce the original dataset by measuring certain features or properties of each image such as texture, color and shape. In the proposed approach, we measure two types of features, geometric features and histogram feature to be later use for classification.

Geometric Features

At this stage, we need to calculate some geometric features (refer to [24]).

- Image Length: The total length of infected part of leaf is calculated.
- Image Area: Compute area of infected part of leaf.
- Image Area Estimate: Compute area of infected part of leaf with respect to edges and corners.
- Image Perimeter: Compute perimeter of infected part of leaf.
- Euler number: In this stage, we compute the Euler number, or Euler Poincare characteristic, of a infected part of image that correspond to the number of connected components in the image.

Histogram Feature

In this step we calculate histogram to image in HSV color space. So we first convert image from RGB color V color space. Then image must be quantized in 8 * 2 * 2 equal bins.

3.5 Classification Phase

The final step of this work is a classification phase, in this approach, SVM technique has been applied for classification of tomato leaf image to any of the following states, infected with TSWV or with TYLCV. The inputs of this stage are training dataset feature vectors and their corresponding classes, whereas the outputs are the decision that determines type of virus that infected tomato leaf. To achieve good results, SVM was trained and tested using different kernel functions that are: Linear kernel, radial basis function (RBF) kernel, QP kernel, Multi-Layer Perceptron (MLP) kernel and Polynomial kernel [25, 26].

For the process of evaluating the results, we used N-fold cross-validation technique [27] with N = 10, firstly, dataset is divided into equally (or nearly equally) N-subsets. Then the cross-validation process is performed N times with each sub-set being once the test dataset and all the others being the training dataset. This process is repeated N times. Hence we take the average of performance of N runs. The algorithm performance measure can be calculated as the average of the performances of 50 runs.

4 Experimental Results

As we mention before the aim of this work is to detect and identify type of virus that infected tomato leaves. here we focus on two types of virus TSWV and TYLCV. Tomato spotted wilt is caused by a virus that is usually spread by thrips. Tomato plants that infected with spotted wilt become stunted and often die. Initially, leaves in the terminal part of the plant stop growing, become distorted, and turn pale green. In young leaves, veins thicken and turn purple, causing the leaves to appear bronze. Necrotic spots, or ring spots, are frequently present on infected leaves and stems often have purplish-brown streaks [28]. Figure 3 illustrates an example of tomato plant image infected with TSWV.

Tomato yellow leaf curl disease (TYLCV) caused by the whitey-transmitted begomovirus (genus Begomovirus, family Geminiviridae) is one of the most damaging diseases of tomato. [29] If virus infected tomato plants at a young stage, tomato plants can be severely stunted and will not produce fruit. Foliage shows an upright or erect growth habit, leaves curl upwards and may be crumpled. Interveinal chlorosis is also observed in the leaves. Figure 4 illustrates an example of tomato plant image infected with TYLCV.

After we collected some prototypes of infected tomato leaves for each type of virus, we cropped them to extract single tomato leaf for each image. K-mean clustering algorithm has been employed to extract regions of interest. Figure 5 demonstrates sample resultant images before and after the segmentation.

Fig. 3 Leaf lesions due to Tomato spotted wilt virus on tomato

Fig. 4 Yellow, distorted tomato foliage caused by Tomato yellow leaf curl virus

After preprocessing phase is done, we calculated some geometric features and histogram as we mention before that constrict feature vectors. The last phase is the classification and prediction of new objects, it is dependent on the support vector machine (SVM). SVM was trained and tested using different kernel functions that are: (Linear kernel, (RBF) kernel, QP kernel, (MLP) kernel and Polynomial kernel). Figure 6 shows the overall accuracy of the classification based on different SVM kernel functions. The Quadratic (QP) kernel function shows the better accuracy classification accuracy.

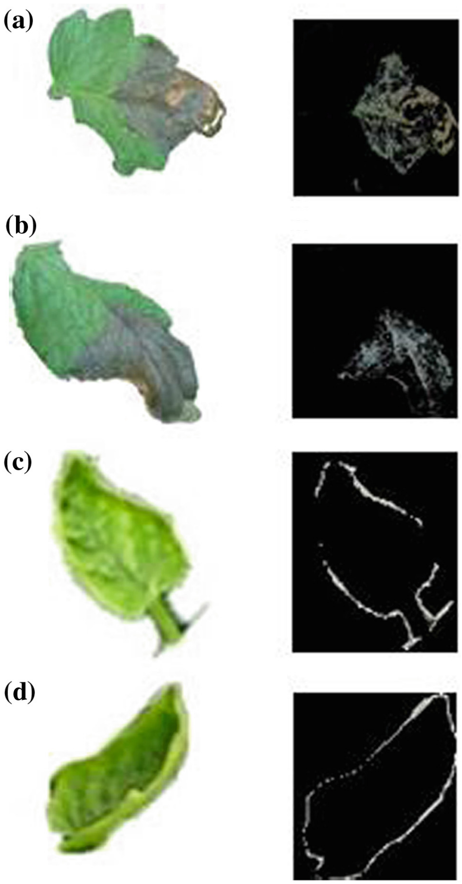

Fig. 5 9-a, 9-b are samples of Leaf infected with TSWV before and after clustering, 9-c, 9-d are samples of Leaf infected with TYLCV before and after clustering

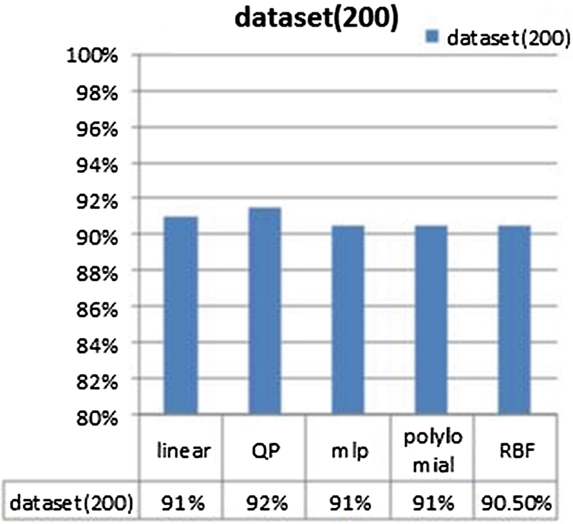

Fig. 6 Results for different SVM kernel functions

5 Conclusions and Future Works

Support vector machine (SVM) algorithm with different kernel functions is used for classification and identifying two of tomatoes leaf viruses. The datasets of total 200 infected tomato leaf images with TSWV and TYLCV were used for both training and testing phase. N-fold cross-validation technique is used to evaluate the performance of the presented approach. Experimental results showed that the proposed classification approach has obtained with accuracy of 90 % in general and the highest classification accuracy of 91.5 % has been achieved using Quardratic kernel function. Our future works will focus on detection and identification other types of virus that infect tomato plants.

References

1. Peralta, I.E., Spooner, M.D., Razdan, M.K. Mattoo, A.K.: History, origin and early cultivation of tomato (Solanaceae). Genet Improv Solanaceous Crops Tomato **2** (2007)
2. Agrios, G.N.: Plant Pathology, 4th edn. Academic Press (1997)
3. Sikora, E.J.: Virus diseases of tomato. ANR-0836 (2011)
4. Rumpf T., Mahlein, A.K., Steiner, U., Oerke, E.C., Dehne, H.W., Plümer L.: Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Comput. Electron. Agric. **47**(1), 91–99 (2010)
5. Hillnhuetter, C., Mahlein, A.K.: Early detection and localisation of sugar beet diseases: new approaches. Gesunde Pflanzen **60**(4), 143–149 (2008)
6. Lin, W.T., Lin, C.H., Wu, T.H., Chan, Y.K.: Image segmentation using the k-means algorithm for texture features. World Acad. Sci. Eng. Technol. **65** (2010)
7. Goclawski, J., Sekulska-Nalewajko, J., Gajewska, E., Wielanek, M.: An automatic segmentation method for scanned images of wheat root systems with dark discolourations. Int. J. Appl. Math. Comput. Sci. **19**(4), 679–689 (2009)
8. ISA, N.A.M.: Automated edge detection technique for Pap smear images using moving K-means clustering and modified seed based region growing algorithm. Int. J. Comput. Internet Manage. **13**(3), 45–59 (2005)
9. Gonzales, R.C., Richard, E.W.: Digital image processing, 2nd edn. (2002)
10. Valliammal, N., Geethalakshmi, S.N.: Plant leaf segmentation using non linear K means clustering. Int. J. Comput. Sci. Issues (IJCSI) **9**(3), 212–217 (2012)
11. Weizheng, S., Yachun, W., Zhanliang, C., Hongda, W.: Grading method of leaf spot disease based on image processing. In: Proceedings of IEEE International Conference on Computer Science and Software Engineering, vol. 6, pp. 491–494 (2008)
12. Sezgin, M.: Survey over image thresholding techniques and quantitative performance evaluation. J. Electron. Imaging **13**(1), 146–168 (2004)
13. Otsu, N.: A threshold selection method from gray level histograms. IEEE Trans. Syst. Man Cybern **9**(1), 62–66 (1979)
14. Camargo, A., Smith, J.S.: An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosyst. Eng. **102**(1), 9–21 (2009)
15. Arivazhagan, S., Shebiah, R.N., Ananthi, S., Varthini, S.V.: Detection of un-healthy region of plant leaves and classification of plant leaf diseases using texture features. Agric. Eng. Int. CIGR J. **15**(1), 211–217 (2013)
16. Tian, J., Hu, Q., Ma, X.X., and Han, M.: An improved kpca/ga-svm classification model for plant leaf disease recognition. J. Comput. Inf. Syst. **8**(18), 7737–7745 (2012)

17. Asraf, H.M., Nooritawati, M.T., Rizam, M.S.B.: A comparative study in kernel-based support vector machine of oil palm leaves nutrient disease. Procedia Eng. **41**, 1353–1359 (2012)
18. Fu, K.S., Mui, J.K.: A survey on image segmentation. Pattern Recogn. **13**(1), 3–16 (1981)
19. Vapnik, V.: The nature of statistical learning theory. Springer (2000)
20. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)
21. Zhang, W., Jin, X.: Image recognition of wheat disease based on RBF support vector machine. In: Proceedings of the International Conference on Advanced Computer Science and Electronics Information (ICACSEI 2013). Atlantis Press (2013)
22. http://www.ipmimages.org
23. Subbaiah, V., Aparna, G.S., Gopal, D.V.R.S.: Computer aided molecular modeling approach of H & E (Haemotoxylin & Eosin) images of colon cancer. Int. J. Comput. Appl. **44**(9), 5–8 (2012)
24. Legland, D., Kiêu, K., Devaux, M.F.: Computation of minkowski measures on 2d and 3d binary images. Image Anal. Stereology **26**(2), 83–92 (2011)
25. Vanschoenwinkel, B., Manderick, B.: Appropriate kernel functions for support vector machine learning with sequences of symbolic data. Deterministic Stat. Methods Machine Learn. 256–280 (2005)
26. Boolchandani, D., Sahula, V.: Exploring efficient kernel functions for support vector machine based feasibility models for analog circuits. Int. J. Des. Anal. Tools Circuits Syst. **1**(1) (2011)
27. Prekopcsák, Z., Henk, T, Gáspár-Papanek, C.: Cross-validation: the illusion of reliable performance estimation. In: RCOMM RapidMiner Community Meeting and Converence (2010)
28. Sikora, E.J., Gazaway, W.S.: Wilt Diseases of Tomatoes. Published by the Alabama Cooperative extension system. Reviewed for web June 2009, Anr-0797
29. Rojas, M.R., Kon, T.: First report of tomato yellow leaf curl virus associated with tomato yellow leaf curl disease in California. Am. Phytopathol. Soc. **91**(8), 1056 (2007)
30. Al Bashish, D., Braik, M., Sulieman B.A.: Detection and classification of leaf diseases using K-means-based segmentation and neural-networks-based classification. Inf. Technol. J. **10**(2), 267–275 (2011)
31. Phadikar, S., Sil, J., Das, A.K.: Classification of rice leaf diseases based on morphological changes. Int. J. Inf. Electron. Eng. **2**, 460–463 (2012)