Automatic recognition of quarantine citrus diseases

Georgina Stegmayer [a,b,*], Diego H. Milone [a], Sergio Garran [c], Lourdes Burdyn [c]

[a] Research Center for Signals, Systems and Computational Intelligence, FICH-UNL, CONICET, Ciudad Universitaria UNL, Santa Fe 3000, Argentina
[b] Centro de Investigación en Ingeniería en Sistemas de Información, CONICET, Lavaise 610, Santa Fe 3000, Argentina
[c] Instituto Nacional de Tecnología Agropecuaria (INTA), Estacion Experimental Concordia, Argentina

ARTICLE INFO

ABSTRACT

Citrus exports to foreign markets are severely limited today by fruit diseases. Some of them, like citrus canker, black spot and scab, are quarantine for the markets. For this reason, it is important to perform strict controls before fruits are exported to avoid the inclusion of citrus affected by them. Nowadays, technical decisions are based on visual diagnosis of human experts, highly dependent on the degree of individual skills. This work presents a model capable of automatic recognize the quarantine diseases. It is based on the combination of a feature selection method and a classifier that has been trained on quarantine illness symptoms. Citrus samples with citrus canker, black spot, scab and other diseases were evaluated. Experimental work was performed on 212 samples of mandarins from a *Nova* cultivar. The proposed approach achieved a classification rate of quarantine/not-quarantine samples of over 83% for all classes, even when using a small subset (14) of all the available features (90). The results obtained show that the proposed method can be suitable for helping the task of citrus visual diagnosis, in particular, quarantine diseases recognition in fruits.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Three diseases have currently quarantine restrictions for the European Union (EU) and the United States of America citrus markets: citrus canker, black spot and scab. Despite this severe limitation, many regions continue to export to these markets by following the guidelines of the so-called System Approach, individually agreed with the EU (SENASA, 2001), which includes different methods to provide the quarantine security required by the citrus trade with the EU and to certify citrus quarantine with minimal risk. A key point to the success of these programs is the effectiveness of the audit work carried out in the field by inspectors, both in the packaging area where the items are processed as well as at the boarding ports. Since tolerance to the presence of symptoms of these diseases is zero, it is essential the early detection of items with such symptoms, especially when they can reach detectable levels in the ports and markets of arrival.

At present, the diagnosis of these diseases, both in field and packing points, depends on the visual method based on the presence of symptoms. Due to the characteristics of the harvest and export operations in the case of having the dubious presence of symptoms, the diagnosis must be realized immediately. However, the visual diagnosis presents a number of disadvantages including the fact that the accuracy and reliability of the diagnostic procedure is subject to the personal capacity of those who make it (Gomez-Sanchis et al., 2012). Moreover, the decision-making process frequently involves subjective factors that provide some degree of variability to the result (Peres, Harakava, Carroll, Adaskaveg, & Timmer, 2007). There is a clear need for computational tools to identify situations in which the value of an entire production is compromised.

The literature currently available on the symptoms of diseases affecting citrus fruits is abundant (Bernal, 2000; Canteros, 2009; Dewdney & Timmer, 2011). However, symptoms of each quarantine infection are described based on a small number of very characteristic attributes (Gottwald, Graham, & Schubert, 2002; Kotze, 2000). This helps in identifying and solving the diagnoses of the most typical symptoms of each disease, but are insufficient to diagnose those that are less frequent and/or share similar attributes and variants with symptoms of other non quarantine diseases. Alternative diagnostic techniques can be used, such as the incubation of fruits under temperature and light controlled conditions, or practices of isolation of the causal agent of dubious symptoms. However, they have proved to be very slow and subject to methodological and experimental errors (Peres et al., 2007).

In recent years, highly sensitive biochemical techniques have been developed, which allow diagnosis within hours. However, they are highly expensive and require instrumental and specific training. Thus, they are not yet available today for the in situ use

* Corresponding author at: Research Center for Signals, Systems and Computational Intelligence, FICH-UNL, CONICET, Ciudad Universitaria UNL, Santa Fe 3000, Argentina.
E-mail addresses: gstegmayer@santafe-conicet.gov.ar (G. Stegmayer), d.milone@ieee.org (D.H. Milone).

required in the field and packaging areas (Baayen, 2002; Gottwald et al., 2002; Meyer, Sanders, Jacobs, & Korsten, 2006). For these reasons, efforts are needed in order to improve the current procedure of visual diagnosis for early detection of diseases (Brosnan & Sun, 2002; Sankaran, Mishra, Ehsani, & Davis, 2010), such as technological strategies using machine learning to achieve intelligent farming (Pydipati, Burks, & Lee, 2006). Current proposals study the problem of post-harvest processing of citrus (Gomez-Sanchez et al., 2012), in particular, visual detection defects through image analysis to classify the fruit depending on appearance (texture and color, Bianconi, Gonzalez, Fernandez, & Saetta, 2012) in an unsupervised way (Lopez-Garcia, Andreu-Garcia, Blasco, Aleixos, & Valiente, 2010; Sankaran et al., 2010).

This paper presents a classifier able to distinguish among the three quarantine diseases mentioned, based on a binary description of the presence or absence of disease symptoms. Such classifier could be stored in a central server that could be accessed online through a simple portable device, without special equipment nor computational processing requirements. An inspector could check the symptoms that he can see on the suspected fruits, send the data to the server and receive a response from it, indicating whether the fruit may have the quarantine disease or not.

The symptomatological descriptions available in the current literature are rather general in nature, aiming to allow the diagnosis of the most typical symptoms only and described using few attributes. Besides, some symptoms are common to symptoms of other non quarantine diseases. In this work, an accurate data set of symptoms has been created through careful observations and descriptions of different types and variations of symptoms caused by the quarantine diseases of interest. After that, a feature selection analysis on the attributes of diseases and their variants has been performed to select the most representatives ones. This allows to minimize the number of features needed to be loaded into a portable device. Several classifiers have been trained with the selected features that better represent each of the three quarantine infections of interest. Results for each classifier have been obtained using cross-validation. Then, the best classifier has been selected for further study, calculating specific performance metrics to deeply analyze its results, class by class.

This paper is organized as follows. Section 2 presents the materials used in this study. Section 3 explains in detail the proposed approach for quarantine diseases recognition, which includes a feature selection step and a classifier training. The performance measures used in this work are presented in Section 4. Section 5 shows the results obtained and their discussion. Finally, the conclusions and future work can be found in Section 6.

2. Materials

Data set used in this study includes citrus canker, black spot and scab symptoms on a group of 212 *Nova* mandarins grown along the Uruguay River citrus growing area. The database of symptoms of each quarantine disease was manually built because, while a number of the studied diseases symptoms are described in Kotze (2000), the variability observed in them is large in practice. Symptoms described in the data set were those ones that represent variants with respect to typical symptoms, for example the four typical symptoms of black spot (Kotze, 2000): freckle spot, hard spot, virulent or spreading spot and speckle blotch, and that are observed less frequently.

The data recollection period lasted one week on May 2011, where the fruits were all mature. Although infections mainly occur during the early growth of the fruits, new symptoms can be observed several months later, when the fruits reach their final color and commercial maturity (Garran & Garin, 2005). This long incubation period hampers the effectiveness of both field and packaging site monitoring inspections for the detection of the disease. The number of samples of each class (quarantine disease) was quite balanced, according to the following detail: 54 citrus canker, 43 black spot, 45 scab and 70 other (non quarantine) diseases. The whole set was randomly divided into two subsets: data set 1 (DS1) having 25% (71 samples) of the total data for feature and model selection; data set 2 (DS2) having the remaining 75% (141 samples) for training and cross-validation testing.

3. Quarantine diseases recognition

3.1. Feature selection

Feature or attribute selection is an active research area in pattern recognition, statistics, and data mining. Its main idea is to eliminate features with little or no predictive information and select only a subset of relevant features for building robust learning models. Feature selection can significantly improve the performance of learning models by removing most irrelevant and redundant features from the data, thus achieving better generalization to test points. Besides, it can help to improve model interpretability and comprehension (Guyon & Elisseeff, 2003).

The problem of feature subset selection is that of finding a subset of the original features of a dataset, such that an algorithm that is run on data containing only these features be able to generate a classifier with the highest possible accuracy. Given an algorithm I that will be used for classification and a dataset $\mathcal{D}$ with features $F_1, F_2, \ldots, F_n$, from a distribution over the labeled instance space, an optimal feature subset, $\mathcal{F}_{opt}$, is a subset of the features such that the accuracy of the classifier $C = I(\mathcal{D})$ is maximal (Kohavi & John, 1997).

Techniques for feature selection can be divided in two approaches: feature ranking, where features are ranked by some criteria and then features above a defined threshold are selected; and subset selection, where one searches a space of feature subsets for the optimal subset. Such approach works by using a function (for example, classifier accuracy) that takes a subset and generates an evaluation value for that subset.

A search is performed in the subsets space until the best solution can be found. For example, best-first search is a commonly used search algorithm which explores a state space by expanding the node with the best score first (Russell & Norvig, 2009). An evaluation function is used to assign a score to each candidate node. The algorithm maintains two lists, one containing a list of candidates yet to explore, and one containing a list of visited nodes. This algorithm always chooses the best of all unvisited nodes, rather than being restricted to only a small subset, such as immediate neighbors. Other search strategies, such as depth-first and breadth-first, have this restriction. In this work we have used the best-first search method for feature selection, which searches the attribute subset space by finding low-dimensional projections of the data that score highly. The features that have the largest projections in the lower dimensional space are then selected (Witten, Frank, & Hall, 2011).

3.2. Classification

A classifier is a mapping from the space of feature values to the set of class values. Each technique uses a learning algorithm to identify a model that best fits the relationship between the attribute set and the class labels in the training data. The models should fit the training data well and also correctly predict the class label of points not seen during the training process (Witten et al., 2011). After the feature selection step explained above, using only the

selected attributes, we have trained three different classifiers (decision trees, neural networks and naive Bayes).

3.2.1. Classification and regression tree (CART)

CART is a classification method that uses data to construct a so-called decision-tree, which is then used to classify new data. The goal of CART is to create a model that predicts the value of a target variable based on several input variables. CART can handle numerical as well as categorical variables.

Decision trees are formed by a set of rules, based on variables in the training set, selected to get the best split to differentiate observations based on the independent variables (the classes). Once a rule is selected and splits a node into two, the same process is applied to each child node of the resulting tree, recursively, until no further changes can be made. That is to say, until a node has the same value of the target variable, or when splitting no longer adds value to the predictions. Each branch of the tree ends in a terminal node. Each observation falls into one and exactly one terminal node, and each terminal node is uniquely defined by a set of rules (Breiman, Friedman, Olshen, & Stone, 1984).

In summary,

1. Take all of the data in the training set.
2. Consider all possible values of all variables.
3. Select the variable/value that produces the greatest separation in the target ($x = t_i$ is called a split).
4. If $x < t_i$ then send the data to the left part of the tree; otherwise, send data point to the right branch of the tree.
5. Repeat same process from 3 on these two nodes of the tree and the data in each node, until no further changes can be made.

3.2.2. Naive Bayes (NB)

The naive Bayes model for joint distributions has been studied extensively in the pattern recognition literature (Duda & Hart, 1973). A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. The naive Bayes model assumes the conditional independence of all effect variables, given a single cause variable. In this model, the class variable (which is to be predicted) is the root and the attribute variables are the leaves. The model is naive because it assumes that the attributes are conditionally independent of each other, given the class. Once the model has been trained, it can be used to classify new examples for which the class variable is unobserved. A deterministic prediction can be obtained by choosing the most likely class (Russell & Norvig, 2009).

The probability model for a classifier can be stated as a conditional model $p(C|F_1,\ldots,F_n)$ over a dependent class variable C with a small number of outcomes or classes, conditional on several feature variables F_1 through F_n. Using Bayes' theorem, we can write

$$p(C|F_1, F_2, \ldots, F_n) = \frac{p(C)p(F_1, \ldots, F_n|C)}{p(F_1, \ldots, F_n)}. \tag{1}$$

In practice we are only interested in the numerator of that fraction, since the denominator does not depend on C and the values of the features F_i are given, so that the denominator is effectively constant. The numerator is equivalent to the joint probability model $p(C, F_1,\ldots,F_n)$. Under the conditional independence assumption, we assume that each feature F_i is conditionally independent of every other feature F_j for $j \neq i$. This means that $p(F_i|C,F_j) = p(F_i|C)$ for $j \neq i$ and so the joint model can be expressed as

$$p(C, F_1, F_2, \ldots, F_n) \propto p(C)\prod_{i=1}^{n} p(F_i|C). \tag{2}$$

The training of a naive Bayes model is computed by simple frequencies (maximum likelihood estimate). The class distribution is estimated by $p(C) = \#(C)/|\mathcal{D}|$, where $\#(C)$ is the number of times the class C shows up in the training data $\mathcal{D}$, with the denominator being the total number of training instances (each instance has a unique class).

The naive Bayes classifier combines the naive Bayes probability model with a decision rule, such as selecting the hypothesis that is most probable; this is known as the maximum a posteriori rule. The corresponding classifier is a function that can be defined as follows:

$$\hat{c}(f_1, f_2, \ldots, f_n) = \arg\max_c p(C = c)\prod_{i=1}^{n} p(F_i = f_i|C = c). \tag{3}$$

This means that for each possible class label, the conditional probability of each feature has to be multiplied together, given the class label. The label for which the largest product is obtained is the label returned by the classifier.

3.2.3. Multilayer perceptron (MLP)

MLP is a type of artificial neural network model, which can be loosely defined as a large set of interconnected units (neurons) that are executed in parallel to perform a common global task. The units undergo a learning or training process in response to input signals, adjusting the internal parameters of the neural model (weights between neurons).

The MLP model, which is the most widely used for classification problems (Zheng et al., 2011), has distinct layers such as input, hidden and output, with no connections among neurons belonging to the same layer. The number of layers and neurons in each layer is chosen a priori, as well as the type of activation functions for the neurons. Each neuron j in each layer computes a weighted sum of its inputs and then applies an activation function to produce an output, as follows:

$$y_j = \phi_j\left(\sum_{i=1}^{n} w_{ji}x_i\right), \tag{4}$$

where y_j is the neuron output, n is the number of inputs, w_{ji} is the synaptic weight connecting the input signal x_i to the neuron j and ϕ_j is the activation function, which will provide a nonlinear mapping between input and target signals (Haykin, 2008).

In the MLP model, learning is supervised and the basic learning algorithm used is backpropagation, which uses gradient descend to minimize a cost function. The cost function is generally defined as the mean square error

$$E = \frac{1}{2}\sum_{k=1}^{p} (t_k - y_k)^2 \tag{5}$$

between the desired or target output (t_k) and each actual network output (y_k), for p output neurons.

During learning, the error propagates backwards through the network and the model parameters are changed according to the so-called delta rule (Haykin, 2008):

$$\delta w_{ji} = -\eta \frac{\partial E}{\partial w_{ji}}. \tag{6}$$

4. Performance measures

The objective in supervised learning is to approximate an unknown function, using a set of data, searching for the model that better predicts the outputs of the unknown function. To fit the parameters of the different models, their performance is compared on a dataset not used during training, to evaluate the generalization capability of each model. This is named cross-validation and it is an effective method for estimating the prediction error of a

classifier to an independent data set. To reduce variability, multiple folds of cross-validation can be performed using different partitions, and the validation results are averaged over the folds.

In *k*-fold cross-validation, the original data set is randomly partitioned into *k* subsets: $k - 1$ subsets are used as training data and the remaining single subset is used for testing the model. The cross-validation process is then repeated *k* times (the folds), with each of the *k* subsets used exactly once. The *k* results from the folds can then be averaged to produce a single estimation (Haykin, 2008).

In classification problems, the primary source of performance measurements in classification is the overall accuracy of a classifier estimated through the classification rate or accuracy (*A*), which is the proportion of correctly classified examples, calculated as

$$A = \frac{tp + tn}{M}, \tag{7}$$

where *M* is the total number of samples, *tp* (true positives) is number of correct predictions of a class sample; and *tn* (true negatives) is the number of correct predictions of a no-class sample.

Another usual performance measures commonly used in pattern recognition are precision and recall (Olson & Delen, 2008), defined as:

$$P = \frac{tp}{tp + fp}, \tag{8}$$

$$R = \frac{tp}{tp + fn}, \tag{9}$$

where *fp* (false positives) is the number incorrect predictions of a class example; *fn* (false negatives) corresponds to the number of incorrect prediction of a no-class example. Precision for a class can be defined as the ratio between samples correctly classified and the total number of samples assigned to a class. That is to say, it is a measure of the fraction of classified samples that are relevant. Recall for a class is the ratio between samples correctly detected over the total number of samples that actually belong to a class.

Precision can be seen as a measure of exactness (fidelity), recall is a measure of completeness, and the *F*-score is a measure that combines precision and recall through their harmonic mean as follows:

$$F = 2 \cdot \frac{P \cdot R}{P + R}. \tag{10}$$

Relative operating characteristic (ROC) curves can be used to visualize the achieved trade-offs between correctly classifying positive and negative cases. A ROC curve is a graphical plot of the true positive rate also known as sensitivity, versus the false positive rate or one minus the specificity, for a classifier system as its discrimination threshold is varied. Each point on the ROC curve represents a classifier with a particular trade-off between sensitivity and specificity. Comparing the performance of multiple classification schemes with statistical tools requires the information represented by the ROC curve to be collapsed into a single response variable (Pietersma, Lacroix, Lefebvre, & Wade, 2003). To this end, the area under the entire ROC curve (AUC) was proposed as a suitable performance index (Bradley, 1997) since it is a value between 0 and 1 and makes easier the comparison of classifiers among them. When AUC is close to 1, it means that most of the positive class samples have been assigned a score higher than any no-class sample, meaning that there is a threshold that perfectly separates the classes.

5. Results and discussion

This section reports the results obtained on the data set described in Section 2. First of all, the feature selection step has been performed. Classification rates for three classifiers are reported before and after the feature selection step. After that, using only the selected features, the three classifiers have been tuned and compared in order to select the most adequate for the recognition of the three quarantine diseases of interest. The results on global, as well as per class classification rates are reported and discussed.

5.1. Feature and model selection

The feature selection, as well as the classification models tuning, have been performed through cross-validation on the smaller subset only, as suggested in the pattern recognition literature (Demsar, 2006). The data set 2 was used for model testing and discussion of results.

First of all, all of the features (90) of the quarantine diseases of interest in this study were used to train and test three classifiers. Table 1 shows the global classification rate obtained with each of the studied classifiers (in columns) after a 5-fold cross-validation procedure performed on each subset (in rows).

It can be seen that for each partition of the original dataset, the global classification rate is higher than 60% for both NB and MLP, being significantly lower in the case of CART for the data set 1. In the case of the data set 2, accuracies higher than 70% are achieved. Note that these rates are obtained if all the 90 features (symptoms) are used for the classification task. This implies that, at the moment of diagnosis, for example at a packaging site, an inspector having a mobile device would have to fulfill 90 boxes in a form in order to send all of the required information to a remote server and get a classifier response. However, if a feature selection procedure is performed over the original 90 features, only the most important features would be used. This would significatively speed the in situ diagnosis task, by reducing the amount of information to be provided to the classifier. This approach, however, implies that less information is given to the classifier, which could reduce its performance. We will show, however, that the performance of the classifier can be maintained, and even, increased, if only the relevant features to discriminate between classes is provided, and noisy and redundant features are not considered for the classification.

For the feature selection step, a 5-fold cross-validation procedure has been performed to decide which features to use. In this step, DS1 was used for feature and model structure selection, and DS2 for estimation of the final error (Arlot & Celisse, 2010). Once the feature selection step was completed by performing a classical best-first search, 14 of a total of 90 features were selected (see Table 2). For CART, the minimal number of observations at the terminal nodes was two and minimal cost-complexity pruning has been used (Breiman et al., 1984). For the MLP model, the best results have been obtained after 100 training epochs with a topology of five hidden neurons and using a 10% of the training set for monitoring the generalization peak.

Table 1
Accuracy (*A* %) for 5-fold cross-validation without feature selection (90 features) in both datasets.

Data set	CART	NB	MLP
DS1 (25% data)	54.93	76.06	60.56
DS2 (75% data)	68.43	79.43	73.76

Table 2
Selected features. All features are binary.

Feature		Description
Shape	Circular spot	Shape of the symptom in the infected area
Topography of the surface	Depressed or Prominent	Topography of the symptom on the surface of the infected area
Deepness	Shallow	Deepness reached by the necrotic tissue within the infected area
Transition zone	Constellation or Aureole	Kind of transition zone between healthy and necrotic tissue
Color of the transition zone	Oily-water-soaked	Color-aspect of the transition zone between healthy–ill tissue
Central color	White	Predominant color of the symptom central zone
Ruggedness of the central surface	Eruptive or Edge perimeter	Type of the surface of the central zone of the symptom
Pattern of the central zone	Flat and smooth	External aspect of the central zone
Central texture	Corky & granular or scabby	Texture of the tissue in the central zone of the symptom
Presence of fruiting bodies	Pycnidia present	Presence of pycnidia on the central zone of the symptom

Table 3
Accuracy (A %) for 5-fold cross validation on DS2 after the feature selection step performed on DS1 (14 features selected).

Class	CART	NB	MLP
Citrus canker	83.80	91.90	94.60
Black spot	70.00	83.30	80.00
Scab	72.40	82.80	86.20
Global	70.92	78.72	83.69

5.2. Classifiers training and testing on selected features

After the feature and model selection steps, the classifiers were trained using the WEKA software[1] by performing 5-fold cross-validation on DS2 (Bouckaert et al., 2010). With this approach, the data set was divided into five mutually exclusive folds with approximately the same class distribution as the original data set. Each fold was used once to test the performance of the classifier that was generated from the combined data of the remaining folds, leading to five independent performance estimates (Pietersma et al., 2003).

Table 3 reports the classifiers performance obtained on DS2 using the 14 main features previously selected. It can be seen that, while NB has maintained its classification rate, both CART and MLP have even increased their performance. In fact, the MLP model has reached a very high classification rate (almost 84%) on the 5-fold test partitions of the larger data subset. This result can be explained with the fact that, once only the most informative features haven been considered for classification, removing noisy and redundant information, the generalization capability of the models has been improved. Moreover, by using the best features for class discrimination, we are performing a better training of the models because there is a better relationship between data size and number of parameters to estimate.

Once the better classifier was obtained, per class performance has been analyzed. Table 4 shows the detail of the performance measures used in this study (columns) for each of the classes (quarantine diseases) included (rows). It can be noticed that the best classifier found (the MLP model) has very high values, almost all of them are near the optimum (1.0). In particular, for the ROC area index (AUC), the values obtained assure a very high performance for the automatic recognition of any of the quarantine diseases under study, since the obtained values are, in all cases, higher than 0.90.

Table 5 shows the performance of the MLP classifier for the recognition of each of the diseases, through a confusion matrix. Looking at the detail of the confusion matrix, it can be seen that the higher mis-classification occurs with the *other* class, not the classes of interest. That is to say, at the presence of a sample of a quaran-

---

Table 4
Detailed performance by class of the MLP (best) model.

Class	P	R	F	AUC
Citrus canker	0.92	0.95	0.93	0.99
Black spot	0.80	0.80	0.80	0.94
Scab	0.93	0.86	0.89	0.95

Table 5
Confusion matrix for the best classification model (MLP) – multiclass problem.

Classified as →	a	b	c	d
a = citrus canker	35	0	0	2
b = black spot	0	24	0	6
c = scab	0	0	25	4
d = other	3	6	2	34

tine disease the classification will be correct. However, at the presence of a sample of a non-quarantine disease, there are 11 samples that would be indicated as quarantine, and 13 quarantine samples that would be indicated as other disease.

It is important to highlight the fact that the model has no confusion among the three classes of interest. That is to say, for the quarantine diseases of interest in this study, there are no misclassified elements: each sample of each quarantine disease has been correctly identified.

For the classification problem stated as binary, that is to say, if only two classes are considered: *quarantine disease* and *not quarantine disease*, the corresponding confusion matrix is shown in Table 6. The classification rate corresponding to the binary problem and the MLP model is of 83%, which maintains the result obtained with this classifier in the multiclass problem.

If the confusion matrix is analyzed in detail, it can be seen that, as stated before, the quarantine diseases samples of interest are correctly recognized in most cases: in 92 out of 96 examples of quarantine disease class are correctly classified. The other class has less examples and the confusion if higher. This fact should be tackled in future works, increasing the number of non quarantine diseases samples or through a better classifier design. Furthermore, existing proposals for visual detection of citrus defects through images and machine vision (Wen & Tao, 1999), that is to say, depending on the texture and colour of the fruit (Lopez-Garcia

Table 6
Confusion matrix for the best classification model (MLP) – two class problem.

Classified as →	a	b
a = quarantine disease	92	4
b = not quarantine disease	20	25

et al., 2010; Sankaran et al., 2010), could be used as a preliminary step to our classifier, obtaining the features automatically through image analysis.

6. Conclusions and future work

It has been explained how citrus exports to foreign markets are limited today, mainly, by fruit diseases. Some of them are quarantine for the markets and have zero-tolerance at the destination market. For this reason, it is important to perform good controls before fruits are exported. Nowadays, technical decisions are highly dependent on the degree of individual skills on human experts, with previous experience in visual diagnosis. This work has presented a model capable of recognizing three quarantine diseases (citrus canker, black spot and scab) in an automatic way and achieving high classification rates, by using barely a bit more than a dozen of characteristics or symptoms that can be seen in a fruit.

The proposed approach is based on the combination of a feature selection method and a classifier that has been trained on the illness symptoms. Experimental work was performed on 212 *Nova* mandarins. The proposed approach achieved a classification ratio of quarantine/not-quarantine samples of 83% for all classes, even when using a small subset (14) of all the available features (90). When the problem was stated as multiclass, also high classification rates of 84% was achieved and AUC values higher than 95%, very close to the optimum. Since only the most informative features have been considered, removing noisy and redundant information, the generalization capability of the models has been improved, with a direct impact on the model performance.

The high classification rates that have been obtained on the task of automatic recognition of quarantine citrus diseases show the usefulness of the proposed approach. Another advantages of the proposed method is the significant reduction of the number of features that have to be used in order to obtain a high classifier response, which could be very helpful for visual inspection on the field, if, for example, the classifier was implemented in a mobile device used at field and packaging site monitoring inspections for the detection of the diseases.

The results obtained show that the proposed method can be suitable for helping the task of citrus visual diagnosis, in particular, quarantine diseases recognition in fruits in the field. All of the quarantine diseases samples are correctly recognized into each corresponding class. As future work, it would be very interesting to design a hierarchical classifier, for the *not quarantine* class only, capable of better discerning with which quarantine disease the sample is being confused.

Acknowledgements

References

SENASA (2001). Programa de Certificación de Fruta Fresca Cítrica de la Región NEA para exportacion a la Unión Europea. Resolución 78/2001, SENASA.

Gomez-Sanchez, J., Martfn-Guerrero, J. D., Soria-Olivas, E., Martfnez-Sober, M., Magdalena-Benedito, R., & Blasco, J. (2012). Detecting rottenness caused by *Penicillium* genus fungi in citrus fruits using machine learning techniques. *Expert Systems with Applications, 39*(1), 780–785.

Peres, N. A., Harakava, R., Carroll, G., Adaskaveg, J., & Timmer, L. (2007). Comparison of molecular procedures for detection and identification of *Guignardia citricarpa* and *G. mangiferae. Plant Disease, 91*(5), 525–531.

Bernal, R. (2000). Some aspects of the epidemiology of citrus scab (*Elsinoe* spp.) in *Satsuma mandarin* cv OwariG and Valencia orange *Citrus sinensis* (L.) Osbeck. In *Proceedings ninth international society of citriculture congress* (Vol. 1, pp. 990–993). Orlando, Florida: EEUU.

Canteros, B. (2009). Guía para la Identificación y Manejo de las Enfermedades Fúngicas y Bacterianas en Citrus, CFI-INTA.

Dewdney, M., Timmer, L. (2011). Alternaria brown spot. UF IFAS Extension, Publication PP152.

Gottwald, T.R., Graham, J.H., Schubert, T.S. (2002) Citrus canker: The pathogen and its impact. *Plant Health Progress* http://dx.doi.org/10.1094/PHP-2002-0812-01-RV.

Kotze, J. (2000). Black spot. In W. L. Timmer, S. M. Garnsey, & J. H. Graham (Eds.), *Compendium of citrus diseases. 2nd ed.*. APS Press.

Baayen, R. P., Bonants, P. J. M., Verkley, G., Carroll, G. C., van der Aa, H. A., Jr., C. Glienke de Blanco, et al. (2002). Nonpathogenic isolates of the citrus black spot fungus Guignardia citricarpa identified as a cosmopolitan endophyte of woody plants G. *mangiferae (Phyllosticta capitalensis). Phytopathology, 92*(5), 464–477.

Bouckaert R Frank, E., Hall, M. A., Holmes, G., Pfahringer, B., et al. (2010). Weka-experiences with a java open-source project. *Journal of Machine Learning Research, 11*(1), 2533–2541.

Meyer, L., Sanders, G., Jacobs, R., & Korsten, L. (2006). A one-day sensitive method to detect and distinguish between the citrus black spot pathogen *Guignardia citricarpa* and the endophyte *Guignardia mangiferae. Plant Disease, 90*(1), 97–101.

Brosnan, T., & Sun, D.-W. (2002). Inspection and grading of agricultural and food products by computer vision systems – A review. *Computers and Electronics in Agriculture, 36*(2), 193–213.

Sankaran, S., Mishra, A., Ehsani, R., & Davis, C. (2010). A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture, 72*(1), 1–13.

Pydipati, T., Burks, T., & Lee, W. (2006). Identification of citrus disease using color texture features and discriminant analysis. *Computers and Electronics in Agriculture, 52*(1), 49–59.

Bianconi, F., Gonzalez, E., Fernandez, A., & Saetta, S. A. (2012). Automatic classification of granite tiles through colour and texture features. *Expert Systems with Applications, 39*(12), 11212–11218.

Lopez-Garcia, F., Andreu-Garcia, G., Blasco, J., Aleixos, N., & Valiente, J.-M. (2010). Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach. *Computers and Electronics in Agriculture, 71*(2), 189–197.

Garran, S., Garin, R. (2005). Evolucion pre- y postcosecha de los sintomas de mancha negra (guignardia citricarpa kiely) en naranja valencia late y mandarina nova en la region citrcola de concordia en entre rios. In *Programa y Resumenes V Congreso Argentino de Citricultura* (p. 84).

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research. Special Issue on Variable and Feature Selection, 3*(3), 1157–1182.

Kohavi, R., & John, G. (1997). Wrappers for feature subset selection. *Artificial Intelligence, 97*(2), 273–324.

Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). New York: Prentice Hall.

Witten, I. H., Frank, E., & Hall, M. (2011). *Data mining: Practical machine learning tools and techniques* (3rd ed.). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc..

Breiman, L., Friedman, J., Olshen, R., Stone, C. (1984). *Classification and regression trees.* Monterey, CA: Wadsworth Brooks/Cole Advanced Books Software.

Duda, R., & Hart, P. (1973). *Pattern classification and scene analysis.* New York: Wiley.

Zheng, H., Fang, S., Lou, H., Chen, Y., Jiang, L., & Lu, H. (2011). Neural network prediction of ascorbic acid degradation in green asparagus during thermal treatments. *Expert Systems with Applications, 38*(5), 5591–5602.

Haykin, S. (2008). *Neural Networks and learning machines* (3rd ed.). New York: Prentice Hall.

Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques.* Springer.

Pietersma, D., Lacroix, R., Lefebvre, D., & Wade, K. M. (2003). Performance analysis for machine-learning experiments using small data sets. *Computers and Electronics in Agriculture, 38*(1), 1–17.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145–1159.

Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research, 7*(1), 1–30.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*(1), 40–79.

Wen, Z., & Tao, Y. (1999). Building a rule-based machine-vision system for defect inspection on apple sorting and packing lines. *Expert Systems with Applications, 16*(3), 307–313.