BOOSTING CLASSIFIERS FOR WEED SEEDS IDENTIFICATION

Pablo M. Granitto, Pablo Garralda, Pablo F. Verdes and H. Alejandro Ceccatto

Instituto de Física Rosario, CONICET and Universidad Nacional de Rosario,

Boulevard 27 de Febrero 210 Bis, 2000 Rosario, Argentina

Abstract: The identification and classification of seeds are of major technical and economical importance in the agricultural industry. To automate these activities, like in ocular inspection one should consider seed size, shape, color and texture, which can be obtained from seed images. In this work we complement and expand a previous study on the discriminating power of these characteristics for the unique identification of seeds of 57 weed species. In particular, we establish statistical bounds and confidence levels on the results reported in our preliminary study. Furthermore, we discuss the possibility of improving the naïve Bayes and artificial neural network classifiers previously developed in order to avoid the use of color features as classification parameters. Morphological and textural seed characteristics can be obtained from black and white images, which are easier to process and require a cheaper hardware than color ones. To this end we boost the classification methods by means of the AdaBoost.M1 technique, and compare the results obtained with the performance achieved when using color images. We conclude that the improvement in classification accuracy after boosting the naïve Bayes and neural classifiers does not fully compensate the discriminating power of color characteristics. However, it might be enough to make the classifier still acceptable in practical applications.

Key words: machine vision; classification; boosting; neural networks

1. INTRODUCTION

Reliable and fast identification/classification of seeds is of major technical and economical importance for the agricultural industry. Common practices based on specialized technicians are slow, have low reproducibility, and possess a degree of subjectivity hard to quantify. Machine vision seems a suitable technique to automate this task, since numerous image-processing algorithms are available for extracting classification features from seed images. Like in the standard ocular identification, automatic classification should be based on knowledge of seed size, shape, color and texture (i.e., greytone variations on the surface, see Haralick *et al.*, 1973 and Galloway, 1975).

Most previous attempts to identify seeds by machine vision have concentrated on cultivated varieties (Draper and Travis, 1984, Keefe and Draper, 1986, Sapirstein *et al.*, 1987, Chen *et al.*, 1989, Symons and Fulcher, 1988; Zayas *et al.*, 1989; Neuman *et al.*, 1989a,b, Neuman *et al.*, 1987). In these studies image analysis was essentially restricted to basic geometrical measurements to obtain different parameters (shape factor, aspect ratio, length/area, etc.). In addition, color was successfully used to separate red-, amber- and white-colored wheat. More recent studies have used color images to establish seed quality and hardseededness of some annual pasture legumes (Jansen, 1995), to characterize fungal damage, viral diseases and immature soybean seeds (Ahmad *et al.*, 1999), etc.

Besides varietal identification and cereal grain grading, early identification of weeds from the analysis of strange seeds is also of major interest in the agricultural industry. This can be done for the purpose of chemically controlling weed growth or, as occurs in many countries, it can be routinely performed as part of official requirements before a seed batch can be made commercially available (purity analysis). In particular, Argentina's law regulations require the analysis by registered laboratories of a small batch sample before a seed batch can be made commercially available. In these analyses, commercial and strange seeds present are separated, and the latter ones identified one by one. The studies in the present work are part of a development to avoid the continuous training of new technicians to perform this task, providing an automatic classifier that can be used by less skilled operators. Weed seeds are also identified by seed testing stations and seed corporations to measure the purity of the harvest, and by research stations to detect changes in seed banks in the soil. Another possible application is the identification of very strange seeds in botanical gardens, although this would require a very large database.

An early attempt to identify weed seeds (Petersen and Krutz, 1992) showed the importance of using color instead of black and white images to improve classification accuracy. More recently, Chtioui *et al.* (1996) compared the capabilities of linear discriminant analysis and artificial neural networks (ANNs) to identify weed seeds from morphological and textural parameters. However, these investigations considered only four different species, which does not provide a good characterization of inter-species seed variations. In a previous work (Granitto *et al.*, 2000), we assessed the discriminating power of different seed characteristics for the unique identification of seeds of weed species. We used a simple Bayesian approach (naïve Bayes classifier) to evaluate morphological, color and textural characteristics measured from video images, establishing their importance as classification features for weed seeds identification. The final classifier based on the optimal set of features showed a remarkable good performance. In addition, we presented classification results obtained using the same feature set as input of a committee of ANNs. These preliminary studies were conducted on a much larger basis than previous ones (Petersen and Krutz, 1992; Chtioui *et al.*, 1996), including seed images of frequent weeds found in Argentina's commercial seed production industry. In particular, to avoid introducing a bias in the selection of species considered we restricted ourselves to the 58 species listed by the Secretary of Agriculture as prohibited and primary- and secondary-tolerated weeds. From this list we finally considered 57 species for which a good number (~ 40) of young exemplars were available in the seed bank of the Seed Analysis Laboratory at the Oliveros Experimental Station of the National Institute for Agricultural Technology (INTA).

In this work we complement and expand our previous study on automatic seed identification in two different ways:

- First, we establish statistical confidence bounds for the discriminating power of different seed characteristics reported in Granitto *et al.* (2000). For this, we perform 30 independent experiments to obtain mean values and standard deviations of the figures previously reported. In addition, we present results for paired *t*-tests between the different classification methods considered, in order to assess the significance of the obtained performance differences.
- Secondly, we explore the possibility of improving the classifiers developed in Granitto *et al.* (2000) in order to achieve similar identification capabilities without using the color variables as classification features. For this, we use the standard boosting algorithm AdaBoost.M1 (Freund and Schapire, 1997).

The first goal is important to establish how robust the results previously reported are. That is, how repetitive they are under changes in training and test sets, randomness in the learning methods, etc. Parts of this study have been recently presented in Granitto *et al.* (2002). The second one points to avoiding the use of the color variable, which would simplify the hardware requirements for a commercial system with the concomitant reduction in cost and operational complexity (black and white cameras and image acquisition boards are much cheaper than RGB ones, control of illumination conditions is far less important in processing gray-tone images than color ones, etc.).

This work is organized as follows. In Section 2 we discuss the statistical error bounds and confidence levels for the results given in our preliminary work (Granitto *et al.*, 2000). In Section 3 we introduce the boosting algorithm AdaBoost.M1, and discuss the efficiency of the boosted classifiers based only on morphological and textural seed characteristics. Finally, in Section 4 we summarize our work and draw some conclusions.

2. STATISTICAL EVALUATION OF FEATURES AND CLASSIFIERS

Image Database and Feature Extraction

We have built a database containing 3163 images of the 57 species considered (a list of these species is available on request). Details on the experimental settings used to capture these images can be found in Granitto *et al.* (2000). They consist of 768×512 pixel arrays whose entries are 24-bit records, corresponding to the 256 pixel intensity levels (8 bits) for each of the red (*R*), green (*G*) and blue (*B*) channels.

From the raw seed images we selected nearly optimal sets of 10 morphological, 7 color and 7 textural features to be used as classification parameters. For this selection we implemented standard sequential forward and backward algorithms (Jain and Zongker, 1997), using the performance of a naïve Bayes classifier as the selection criterion. The naïve Bayes classifier fits the class conditional probabilities with a product of distributions of the individual features —we used Gaussian distributions in this implementation— and, in spite of this simplification, it performs well on the problem at hand (see below). The same selection procedure applied to the whole 24 parameters retained 12 (6 morphological, 4 color and 2 textural) features, which were finally used to build the classifiers. A description of these parameters and the selection algorithm can be found in Granitto *et al.* (2000). Here, in building the boosted classifiers we will keep the original 10 morphological and 7 textural features to evaluate the possibility of disregarding color information.

Discriminating power

To compare the discriminating power of the different sets of features, in Table 1 we present the generalization capabilities of naïve Bayes classifiers built solely in terms of either the 10 morphological, 7 color or 7 textural features. Unlike the results previously reported, which were obtained in a single run, here, and in all the subsequent tables, the figures correspond to an average over 30 independent experiments as described next. In each experiment, we split the 3163 images of the 57 species considered in training and test sets, randomly choosing, for each species, 80% of the images to build the classifier and including the remaining 20% in the test set. This leaves 2530 images for training and 633 images for testing the system. Table 1 gives the average performances and standard deviations over the 30 experiments for both the training and test sets. It also shows how performance increases when the system assigns a test image to any of the *n* most probable classes. The different cases considered are indicated in the table as First Option (n=1, standard classification), First Two Options (n=2) and First Three Options (n=3). Notice that for $n > 1$ the classification is considered as correct if the test image corresponds to any of the *n* classes with the largest probabilities. This possibility is very useful in practice, since untrained operators can easily select the correct option by simple visual comparison with stored representative seed images of the *n* classes suggested by the classifier.

Table 1: Naïve Bayes classifier performances as percentage of correct seed identifications using only one particular set of features at a time. Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.

Features	First Option		First Two Options		First Three Options	
	Training	Test	Training	Test	Training	Test
Morphology	86.7 ± 0.3	84.5 ± 1.1	95.8 ± 0.5	94.4 ± 0.9	98.2 ± 0.1	97.3 ± 0.5
Color	60.2 ± 0.7	57.2 ± 1.7	73.6 ± 1.5	70.3 ± 2.1	81.6 ± 0.4	78.6 ± 1.6
Texture	55.3 ± 0.5	53.1 ± 1.4	69.3 ± 1.3	66.7 ± 2.2	76.9 ± 0.5	74.2 ± 1.5

A quick look at Table 1 confirms the conclusions anticipated in Granitto *et al.* (2000) on the basis of a single realization: The largest discriminating power corresponds to morphological features, and color and texture characteristics are not very good as classification parameters. Furthermore, the best generalization results for a combination of two type of characteristics corresponds to morphology and color (see Table 2). Notice, however, that using morphology and texture would require to consider only black and white images, which, as stated in the Introduction, constitutes an important simplification and a reduction in hardware cost. Finally, the performances of naïve Bayes classifiers built in terms of the optimal set of 12 features are given in Table 3.

Table 2: Naïve Bayes classifier performances as percentage of correct seed identifications using different combination of two sets of features. Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.

Features	First Option		First Two Options		First Three Options	
	Training	Test	Training	Test	Training	Test
Morphology + Color	96.6 ± 0.2	94.8 ± 0.8	99.1 ± 0.1	98.2 ± 0.6	99.4 ± 0.1	98.9 ± 0.4
Morphology + Texture	91.7 ± 0.3	89.1 ± 0.9	97.6 ± 0.3	95.9 ± 0.7	98.7 ± 0.1	97.5 ± 0.5
Color + Texture	83.2 ± 0.3	80.4 ± 1.3	91.0 ± 0.7	88.6 ± 1.4	94.6 ± 0.3	92.5 ± 0.1

Table 3: Performances of different classifiers as percentage of correct seed identifications using the optimal set of 12 features. Mean values and standard deviations are estimated from 30 independent experiments, as described in the main text.

Classifier		First Option		First Two Options		First Three Options	
		Training	Test	Training	Test	Training	Test
Naïve Bayes		97.5 ± 0.2	95.8 ± 0.9	99.4 ± 0.1	98.7 ± 0.4	99.7 ± 0.1	99.2 ± 0.3
Single ANN		100	95.6 ± 0.8	100	98.6 ± 0.4	100	99.4 ± 0.3
Committee	MR	100	96.6 ± 0.7	100	98.3 ± 0.5	100	98.5 ± 0.4
	AP	100	96.7 ± 0.7	100	99.0 ± 0.5	100	99.5 ± 0.3

It is important to stress that differences in performance scores in Tables 1 and 2 are statistically significant in all cases, with at least 29 out of the 30 experiments performed behaving consistently with these differences. For instance, color features have larger discriminating power than textural ones in 29 out of the 30 experiments [remember that ideally 22/30 would correspond to a 99% confidence level, although there are risks in applying the paired-difference *t*-test to different random train-test splits of the data (Dietterich, 1996)].

Neural network classifier

Like in Granitto *et al.* (2000), to compare with the naïve Bayes classifier we have also developed a classifier based on ANNs (Bishop, 1995). For this we used the same feature set selected in the Bayesian approach and trained feedforward networks with 12 input, h hidden, and 57 output units. The numbers of input and output units correspond to the number of parameters used and seed species to be identified respectively. The number of hidden units was varied from h=20 to h=80, monitoring the performance on cross-validation samples set aside from the training data; the results presented below correspond to h=40 units, which lead to the smallest classification error on these samples. We employed output units with softmax (normalized exponential) activation functions to allow the interpretation of outputs as class probabilities. Furthermore, a cross-entropy error measure was used, which is the standard choice for classification problems. We trained the ANNs with the usual backpropagation rule until convergence, since only negligible overfitting problems were observed. This avoided the use of part of the training set for validation purposes (except for the initial selection of the optimal number of hidden units).

The performance of a single (generic) ANN and the results obtained by structuring 10 networks in a committee are shown in Table 3. In the case of the ANN committee we considered two options: i) each network votes for the class with the largest probability according to its own outputs, and the image is finally assigned to the class with the majority of votes (Majority Rule, MR), and ii) the class probabilities output by the 10 networks are added and the image is assigned to the class with the largest sum value (Added Probabilities, AP). Again in this case, all the results quoted correspond to an average over 30 independent realizations of the whole procedure.

A complete comparative description of the different methods' performances is given in Table 4, where the entry (*i*,*j*) gives the number of experiments in which the method in row *i* produced better results (at first option) than the method in column *j*. Using standard paired *t*-tests with the proviso mentioned above concerning random splits of the data, these figures show that the two ANN committee implementations are better than the naïve Bayes and single ANN classifier with more than a 99% confidence level. Moreover, the strategy of adding probabilities in the committee is better than the majority-rule vote with a confidence level also above 99%.

Several comments are in order at this point. First, we stress the excellent performance of the naïve Bayes classifier, which might be related to an effective near independence of the selected classification parameters. Secondly, since in the ANN approach the performance of a single network is already very good, there is little room left for improvement by adding several predictors in a committee. Notice that when the system is allowed to suggest three options for class membership, from the 633 images in the test set only 5 images are misclassified by the naïve Bayes classifier and 4 images by the single ANN (for both methods the performance reaches 100% with five options). Of course, for a much larger number of species the classification problem would be more demanding and the ANN committee might have an advantage over other simpler methods. Furthermore, feature selection should be performed using this classifier as selection criterion to obtain an optimal set for the ANN approach. In passing we mention that there are more sophisticated feature selection methods than the one implemented in this work (Jain and Zongker, 1997, Chtioui *et al.*, 1998). Finally, we stress the fact that different realizations of training and test sets do not substantially change performances (Tables 1 to 3), as indicated by the low standard deviations observed in the 30 independent runs. Consequently, all the above findings support the conclusions already drawn in Granitto *et al.* (2000).

Table 4: Number of experiments in which the classification methods indicated on the left column perform better than the methods indicated on the top row, for the 30 independent experiments performed.

Method		Naive Bayes	Single ANN	Committee MR	Committee AP
Naïve Bayes			16	6	7
Single ANN		14		0	0
Committee	MR	24	30		5
Committee	AP	23	30	25	

3. BOOSTING CLASSIFIERS: THE ADABOOST.M1 ALGORITHM

Boosting

Boosting is a general method to improve the performance of any learning algorithm that consistently generates classifiers with misclassification errors smaller than 50% on a given problem. The first effective boosting algorithm was developed by Schapire (1990); more recently, Freund and Schapire (1997) introduced AdaBoost, a new algorithm that has undergone intense theoretical study and empirical testing in the last years. In particular, in this work we will implement the simplest extension of AdaBoost to multiclass problems, the so called AdaBoost.M1 algorithm, whose pseudocode is given next.

AdaBoost.M1 Algorithm

Input: Data set $D=\{(x_i, y_i), i=1,m\}$, where $x_i \in X$ and $y_i \in Y=\{1, ..., k\}$
Base learning algorithm *WeakLearn*
Integer T

Initialize: $w_i(1) = 1/m \quad \forall\ i=1,m$

For $t = 1, 2, ... , T$:

- generate data set D_t by re-sampling m examples from D with probability $w_i(t)$
- train *WeakLearn* on D_t to obtain the base classifier $h_t : X \rightarrow Y$
- compute the error of h_t: $\varepsilon_t = \Pr_{i\sim D_t} [h_t(x_i) \neq y_i]$
  If $\varepsilon_t > ½$, then $T = t - 1$ (abort loop)
- assign $\beta_t = \varepsilon_t /(1-\varepsilon_t)$
- update distribution $w(t)$:

$$w_i(t+1) = w_i(t) \times \begin{cases} {}_t & \text{if } h_t(x_i) \neq y_i \\ 1 & \text{otherwise} \end{cases}$$

- normalize distribution $w_i(t+1)$

$$Z_t=\sum_{i=1,m} w_i(t+1) \qquad w(t+1) = w(t+1)/Z_t$$

Output: Final hypothesis

$$h_{Boost}(x) = \underset{y\in Y}{\text{argmax}} \sum_{t:\, h_t(x)=y} \log(1/\beta_t)$$

AdaBoost.M1 takes as inputs a dataset $D=\{(x_i, y_i)\ ,\ i=1,m\}$ with m examples, where $x_i \in X$ is an attribute vector and $y_i \in Y$ the corresponding class label (we will consider k classes), and a weak classification method (*WeakLearn*). It calls repeatedly *WeakLearn*, applying, in each iteration t, this algorithm on a training dataset D_t obtained by re-sampling from D with probability $w_i(t)$. That is, *WeakLearn* finds a new classifier $h_t : X \rightarrow Y$ seeking to minimize the training error $\varepsilon_t = \Pr_{i\sim Dt}[h_t(x_i) \neq y_i]$ (notice the error is measured with respect to the distribution of examples in D_t). This process is iterated T times, and the hypotheses $h_1, \ldots, h_T$ obtained are combined in a final classifier h_{Boost}. The re-weigthing of examples in D (starting from $w_1(t) = 1/m$) and the way of combining the succesive hypotheses h_t are indicated in the algorithm's pseudocode above. Other boosting schemes change these particular rules. The idea behind boosting techniques is that "easy" examples that are correctly classified by most previous hypotheses get a small weight, while "hard" examples usually wrongly classified get larger weights. Thus, boosting concentrates the efforts of *WeakLearn* in those examples that are difficult to learn by this base algorithm. The final classifier is a weighted voting of the weak hypotheses obtained during the T iterations. The most important fundamental property of this technique is the fact that if *WeakLearn* has consistently errors $\varepsilon_t < ½$, then the misclassification error of h_{Boost} on D drops to zero exponentially fast. Of course, this does not mean that the test error will be small. However, if T is not too large, theoretical and empirical investigations indicate that h_{Boost} may have very good generalization capabilities.

The main drawbacks of boosting are: 1) The base learner must produce hypotheses with misclassification errors $\varepsilon_t < ½$. For random guessing among k classes the expected error is $1 - 1/k$, so that for $k > 2$ the requirement on ε_t can be difficult to achieve. Fortunately, for the problem under consideration in this work the weak classifiers used (Naïve Bayes and ANN) have errors smaller than ½ (see Fig. 2 below). 2) For very noisy datasets, containing many misclassified objects, the algorithm places too much attention on these wrong examples and the generalization performance deteriorates. In these cases regularization methods become necessary.

Alternatively, one may use a simpler committee method like *bagg*ing (for *b*oostrap *agg*regation, see Breiman, 1996). Bagging simply trains T base classifiers on boostrap re-samples of D, and outputs as final hypothesis

$$h_{Bagging}(x) = \underset{y\in Y}{\text{argmax}} \sum_{t=1,T} \Pr[h_t(x)=y]\ .$$

That is, it assigns instance x to the class with the largest added probability according to the T base classifiers. In the following we will use the performance of this committee method as an additional basis of comparison for the performance of the more sophisticated boosting algorithm.

As stated in the Introduction, we have boosted the base classifiers described in the previous section (naïve Bayes and ANN) without using the color features. For this we considered only the intensity $I=(R+G+B)/3$ instead of the three color channels. In the case of the naïve Bayes classifier, we considered two standard ways of representing the class distribution of the individual features: i) by fitting a normal distribution, like in the calculations in Section 2, and ii) by using a discrete histogram and optimizing the number of bins. This last alternative gives more flexibility to the base classifier, allowing it to learn perfectly all the training examples after some rounds of the boosting algorithm. Although this seems to be preferable from a methodological point of view (no examples are "too hard" for the base classifier), we will see that overfitting problems deteriorate the final results. In Fig. 1 we give the misclassification error, averaged over the 30 experiments, for the training and test sets, as a function of the number of ensemble members (equivalent to boosting rounds T). The training error plot shows, as salient feature, the flexibility of ANN and discrete naïve Bayes base learners, whose errors reach zero after some rounds of the algorithm. From the test error plot we see that boosting performs always slightly better than bagging for all learning algorithms, and that ANNs produce the best generalization results for this problem. A summary of the most relevant results is given in Table 5. In Fig. 2 we plot the error ε_t on the re-weighted training set D_t during the boosting process. We see that, for all base learners, ε_t stays below 0.5 as required, being particularly small for the oversized ANNs used that are able to fit even the hardest examples.

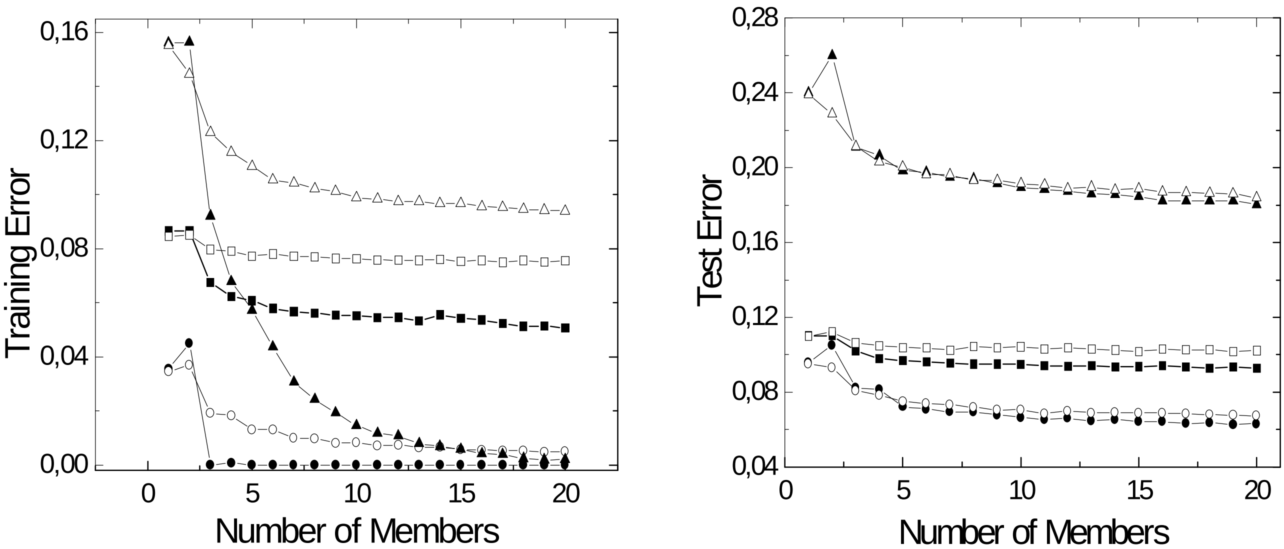

Fig. 1: Training and test errors as a function of the number of ensemble members. Full and open symbols indicate the results of boosting and bagging the base learners. Squares and triangles correspond to naïve Bayes with Gaussian distributions and discrete histograms respectively; circles are the results for ANN.

The means and standard deviations of the results obtained after 30 runs of AdaBoost.M1 are given in Table 5. For the sake of comparison, in this table we include also the results obtained in Section 2 without boosting the classifiers, both using the optimal set of 12 features (first column) and only the 10 morphological and 7 textural parameters (second column). These last results correspond to the average performances of the first classifiers obtained while running AdaBoost.M1 for each base learning method (T=1). The third column in Table 5 gives the results of simply bagging 20 classifiers, and the last column contains the results of AdaBoost.M1 with T=20. A larger number of iterations of the algorithm does not lead to a sensible improvement in the final results. A few comments are in order at this point: i) The comparison of the first and last columns clearly shows that the performance of a single classifier improves via boosting, but this is not enough to regain the accuracy obtained with the inclusion of color features. ii) Although the AdaBoost.M1 algorithm produces the best black-and-white images classifiers, the improvement over the standard classifiers and the simpler bagging approach is not substantial. This certainly points to the fact that, for this problem, the naïve Bayes and ANN approaches are already producing very good classifiers, without leaving much room for improvement. iii) The histogram implementation of the naïve Bayes classifier shows the largest improvement due to boosting. Unfortunately, this base learner has a poor performance due to overfitting and the boosted classifier does not reach a competitive performance.

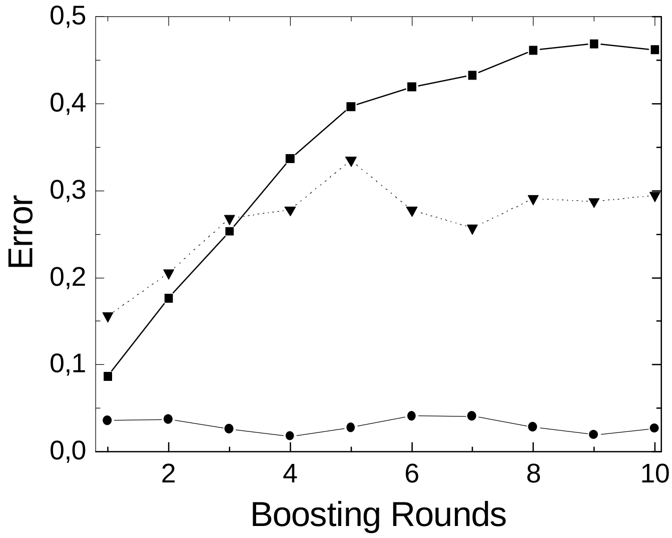

Fig. 2: Error ε_t evolution as a function of boosting rounds t, for different single base learners: Naïve Bayes with Gaussian distribution (squares) and discrete histograms (triangles) and ANNs (full circles).

Table 5: Average results and corresponding standard deviations of 30 boosting experiments with T=20, for the base classifiers indicated and using black and white images. Also given for comparison, results corresponding to the standard (single) classifiers and 20 bagged base learners and performances obtained in the previous section using color images.

Method		Color Images	Black and White Images		
			Standard	Bagging	Boosting
Naïve Bayes	Gaussian	95.8±0.9	89.0±1.0	89.8±1.1	90.7±1.1
	Histogram	-	76.0±2.3	81.6±1.2	82.0±2.0
ANN	Single	95.6±0.8	90.4±1.0	93.1±1.0	93.7±0.8
	Committee	96.6±0.7	-	-	-

As mentioned previously in connection with Table 3, instead of simply trying to identify the seed one can let the system suggest several probable options to the operator, so that he/she can make the final decision. In this practical situation, boosted ANNs predict the correct species within three options with (98.8±0.4)% of accuracy. This performance might be acceptable for a commercial system, which opens up the possibility of adapting the existing software to work only with gray-tone images.

4. SUMMARY AND CONCLUSIONS

We performed a statistical analysis of the discriminating power of different characteristics of weed seeds measured from color images, using the performance of a naïve Bayes classifier as evaluation criterion. In agreement with the preliminary conclusions in Granitto *et al.* (2000), we established that size and shape are the principal characteristics for seed identification (see Table 1), although color and texture must be also considered to obtain a good performance. These last properties have approximately the same discriminating power when considered independently of morphology.

The implementation of a naïve Bayes classifier on the basis of the selected parameters produced excellent results, as shown in Table 3. This is in part due to the careful parameter selection, which lead to a set that approximately fulfills the independence of classification features assumed by the naïve Bayes approach. Table 3 also gives the performances of classifiers based on ANNs for comparison. In particular, the best ANN committee correctly identified 99.5% of the test images within the three most probable classes for each case. Furthermore, we determined with a 99% confidence level that this ANN committee performs better than the simple Bayesian approach (Table 4).

In addition to these statistical studies, which support the conclusions in Granitto *et al.* (2000), we have also investigated the possibility of avoiding color features in the identification problem. For this, we improved the Bayesian and ANN methodologies by boosting them via the AdaBoost.M1 algorithm. The purposes here were simplifying operating conditions (illumination control) and reducing the cost of a potential commercial system based on the current development. The boosted classifiers developed on gray-tone features (morphology and texture) were not able to achieve the same performance reached by using color characteristics of the seeds. Notwithstanding this, the best result reported in Table 5 of approximately 94% of accuracy at first option, and the 98.8% within three options quoted at the end of the previous section, might well be acceptable in commercial applications.

For the number of species considered in this study, the preprocessing of images and the careful selection of measured features reduced considerably the complexity of the classification problem. However, one might expect this problem to become more demanding for databases containing several hundreds of species, as required in some applications. In such a case, the improvement of base classifiers via boosting might be more important. Work in this direction requires the lengthy acquisition of an extended database, which is currently in progress.

ACKNOWLEDGEMENTS

We acknowledge the constant assistance of Eng. Roque Craviotto and technicians of the Seed Analysis Laboratory at EEA Oliveros of INTA. This project was partially financed through grant PICT 11-03834 from ANPCyT.

REFERENCES

Ahmad, I.S., Reid, J.F., Paulsen, M.R., Sinclair, J.B., 1999. Color classifier for symptomatic soybean seeds using image processing. Plant Disease 83, 320-327.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Clarendon Press, Oxford.

Breiman, L., 1996. Bagging predictors. Machine Learning 26, 123-140.

Chen, C., Chiang, Y.P., Pomeranz, Y., 1989. Image analysis and characterization of cereal grains with a laser range finder and camera contour extractor. Cereal Chem. 66(6), 466-470.

Chtioui, Y., Bertrand, D., Dattée, Y., Devaux, M.F., 1996. Identification of seeds by color imaging: Comparison of discriminant analysis and artificial neural networks. J. Sci. Food Agric. 71, 433-441.

Chtioui, Y., Bertrand, D., Barba, D., 1998. Feature selection by a genetic algorithm. Application to seed discrimination by artificial vision. J. Sci. Food Agric. 76, 77-86.

Dietterich, T.G., 1996. Proper statistical tests for comparing supervised classification learning algorithms. Technical Report. Department of Computer Science, Oregon State University, Corvallis, OR.

Draper, S.R., Travis, A.J., 1984. Preliminary observations with a computer based system for analysis of the shape of seeds and vegetative structures. Journal of National Institute of Agricultural Botany 16, 387-395.

Freund Y., Schapire, R.E., 1997. A decision theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119-139.

Galloway, M.M., 1975. Textural analysis using gray level run length. Computer Graphics and Image Processing 4, 172-179.

Granitto, P.M., Verdes, P.F., Navone, H.D., Ceccatto, H.A., 2000. Automatic Identification of Weed Seeds by Color Image Processing, VI Argentine Congress on Computer Science (Ushuaia, Argentina).

Granitto, P.M., Navone, H.D., Verdes, P.F., Ceccatto, H.A., 2002. Weed Seeds Identification by Machine Vision, Computers and Electronics in Agriculture 33, 91.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics 3(6), 610-621.

Jain, A., Zongker, D., 1997. Feature selection: Evaluation, application, and small sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(2), 153-158.

Jansen, P.I., 1995. Seed production quality in *Trifolium balansae* and *T. resupinatum*: The role of seed color. Seed Sci. Technol. 23, 353-364.

Keefe, P.D., Draper, S.R., 1986. The measurement of new characters for cultivar identification in wheat using machine vision. Seed Sci. Technol. 14, 715-724.

Neuman, M.R., Sapirstein, H.D., Shwedyk, E., Bushuk, W., 1987. Discrimination of wheat class and variety by digital image analysis of whole grain samples. J. Cereal Sci. 6, 125-132.

Neuman, M.R., Sapirstein, H.D., Shwedyk, E., Bushuk, W., 1989[a]. Wheat grain color analysis by digital image processing I. Methodology. J. Cereal Sci. 10, 175-182.

Neuman, M.R., Sapirstein, H.D., Shwedyk, E., Bushuk, W., 1989b. Wheat grain color analysis by digital image processing II. Wheat class discrimination. J. Cereal Sci. 10, 183-188.

Petersen, P.E.H., Krutz, G.W., 1992. Automatic identification of weed seeds by color machine vision. Seed Sci. Technol. 20, 193-208.

Sapirstein, H.D., Neuman, M., Wright, E.H., Shwedyk, E., Bushuk, W., 1987. An instrumental system for cereal grain classification using digital image analysis. J. Cereal Sci. 6, 3-14.

Schapire, R.E., 1990. The strength of weak learnability. Machine Learning 5(2), 197-227.

Symons, S.J., Fulcher, R.G., 1988. Determination of wheat kernel morphological variation by digital image analysis: I. Variation in Eastern Canadian Milling Quality Wheats. J. Cereal Sci. 8, 211-218.

Zayas, I., Pomeranz, Y., Lai, F.S., 1989. Discrimination of wheat and nonwheat components in grain samples by image analysis. Cereal Chem. 66(6), 233-237.