Automatic classification of seed species by neural networks and optic RGB images

Anonymous ECCV submission

Paper ID 3658

Abstract. In this paper we propose a neural network approach aiming at automatically detecting native seeds species from the paramo ecosystem based on optic RGB images. The network architectures which are explored for this purpose consist in shallow feed-forward networks with one hidden layer holding up to 15 neurons, and deep convolutional neural networks (CNNs). Focus is placed on four species which are commonly found under natural conditions in the paramo ecosystem, namely *Espeletia congestiflora, Bucquetia glutinosa, Calamagrostis effusa* and *Puya santosii*. Images of the individual seeds were taken on both a white and a black-soil background, the latter simulating the natural conditions where the seeds can be found. We show that relevant knowledge for classification of the seeds' species can be extracted only from their optical information. Under a double cross-validation scheme, a 14-neuron shallow network achieves an 88% test accuracy, while a CNN achieves 94%. On the other hand, after augmenting the available image-data, a CNN is built with a 100% accuracy on validation and a small control-test set. Overall, this neural network approach suggests a promising methodology for seed species prediction based on optical RGB images.

Keywords: Neural networks, Robust learning, Image augmentation, Deep learning application, Seed phenotyping

1 Introduction

An application of neural networks is explored in this study for predicting the species of native seeds from the paramo of *La Rusia*, at The Andes mountains in Boyaca, Colombia. Paramos are neotropical mountain environments well known for their high biodiversity, holding approximately 5000 plant species [**?**]. As study cases, we selected four characteristic species: the tussock grass *Calamagrostis effusa*, a common species forming most paramo grasslands in Colombia, *Espeletia congestiflora*, a giant stemrosette typical of the paramo La Rusia, *Puya santosii*, a giant basal rosette that is found mostly in paramo grasslands and *Bucquetia glutinosa*, representantive of tall shrublands in the area.

Here the problem of seed species prediction is conceived as a multi-class classification problem, with the purpose of developing models for fast and non-destructive real-time detection and species' prediction of native seeds.

Revising the existing literature, different studies have explored the use of multi-spectral and hyper-spectral images (containing information on ultraviolet

and near infrared wavelengths) for seed variety prediction [7,8,9], or optical RGB images for seedling species prediction [1]. Regarding the studies on variety, it has been shown that a combined spectral, morphological, and texture approach, extracting features from visible and near-infrared hyperspectral (VIS-NIR) images can achieve an accuracy of up to 98.2% by a Support Vector Machine (SVM) model [7], or based only on VIS-NIR hyperspectral images and texture features (contrast, homogeneity, energy and correlation) [8], six varieties of maize seeds were predicted by means of principal components and an SVM model (PCA-SVM) with a 98.89% accuracy. One last study based only on hyperspectral images of three varieties of grape seeds achieves a prediction accuracy of 88.7% also following a PCA-SVM technique.

Here it is conjectured here that the availability of hyper-spectral images allows using traditional techniques such as PCA-SVM, but studies based only on optic (RGB) images could require more general statistical techniques for arriving at possible satisfactory results. In this sense, somehow closer to this study, a study on the prediction of seedling species was developed on 5,000 RGB images with 960 unique plants belonging to 12 species with Convolutional Neural Networks (CNNs) algorithms, achieving a test accuracy of 99.48% [1].

In particular, this study on the prediction of seed species is a pioneer study as there is no other study in literature up to the authors knowledge dealing with such a problem, even more if we focus on paramo samples of seeds recollected from The Andes mountains in Boyaca, Colombia. Hence, a direct comparison with previous results cannot be developed, but the statistical experiments performed here will compare the performance of simple (linear) to more complex (non-linear) classifiers. Namely, we explore one hidden-layer feed forward neural networks, holding from 1 to 15 neurons, and CNNs [2].

The methodology followed in this study is built on images of seeds that have been manually selected from soil samples recollected from the paramo of La Rusia, with both clean-white and black-soil backgrounds, the latter resembling the natural conditions where seeds were found. In this way, seeds should be visible on the surface of the earth for the recognition procedure to be useful in practice. Therefore, it should be noted that other technologies perhaps based on hyper-spectral imaging could be explored in a future line of research for extending this methodology for image species prediction on a thicker layer of the surface of the earth.

This study is organized as follows. Section 2 presents the procedures for the recollection of the seed samples, and how images were captured. Then in Section 3 we explain the data pre-processing and neural network approach to image analysis, building single hidden-layer, feed-forward neural nets and CNNs. Finally, in Section 4, we present our results, and in Section 5, some final comments are given for future research.

2 Biological experiments

The seeds used in this study were obtained from fructiferous plant samples collected at the paramo of La Rusia (5.9415°N, 73.0517°W). These samples were taken to the Environmental Engineering laboratory at Los Andes University, and when a sample contained at least 50 seeds in good state, then the species was considered adequate for further analysis. Seeds were manually selected and set aside, being classified according to their species, and photographed on a white background as well as on a black-soil background by an optic camera, model NIKON D-3000, mounted on a stereoscope Zeiss Stemi 508, capturing images of the seeds with a 35x magnification.

Taking the four categories for the different seeds species considered in this study, neural network models were trained and validated under different configurations for their hidden architecture. In the following section we explore these configurations and explain the methodology for identifying the best network model for native paramo seed species prediction.

3 Methodology

In this section we review the main concepts and procedures for building the logistic regression implementation of a neural net (one hidden unit), and the more general models consisting in a single layer with multiple hidden units/neurons, and a deep CNN architecture. The code was all developed on the open-source programming language Python 3.0.

For the statistical experiments we have 30 images for each of the 4 species of seeds, where each seed is photographed from different angles both on a clean-white and a soil-black background (as illustrated in Figure 1). Hence, in total we have 124, 146, 150 and 153 images for the four species *Espeletia congestiflora, Bucquetia glutinosa, Calamagrostis effusa* and *Puya santosii*, respectively, being the four classes well balanced and represented.

The statistical experiments for the development of the single hidden-layer and the CNNs follows a double cross validation scheme, i.e., repeating multiple times the training, validation and test of the same network architectures with different partitions of the data. This procedure allows addressing the difficulty of having a small amount of data for training complex models while controlling the risk of over-fitting. Besides, reducing the bias of our estimation of the loss in prediction, with respect to running only one time the validation and test procedure. Different partitions of training, validation and test sets are also explored.

In this way, the development of the network models follow three scenarios for the double cross validation scheme. The first scenario randomly takes 70% of the images for training and 15-15% for validation-test. The second one takes 80% for training and 10-10% for validation-test, and a third one tries a more demanding training design consisting in taking 50% for training, 15% for validation and 35% for test. All random partitions are repeated 10 times, and are developed in a stratified fashion maintaining the balance of the classes.

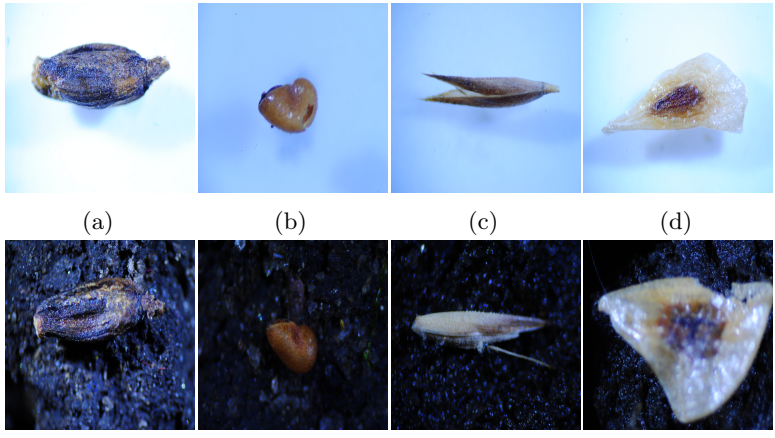


Fig. 1: An example of species (a,e) Espeletia congestiflora, (b,f) Bucquetia glutinosa, (c,g) Calamagrostis effusa and (d,h) Puya santosii on clean-white and soil-black backgrounds.

Lastly, the performance of CNNs is further explored, under one single run for train, validation and test, after augmenting the image data by rotation of the seed images.

3.1 Single hidden-layer neural networks

For the feed-forward and error retro-propagation architecture of the single-hidden layer network, the activation function of each neuron in the hidden layer is defined as a sigmoid function σ, such that the feed forward linear combination is computed by

$$A = \sigma(W \cdot X + b) = \frac{1}{1 + exp^{-(W \cdot X + b)}} \tag{1}$$

where W is a $n_n \times n_x$ matrix of weights associated to the n_n neurons and the n_x features or rows of the matrix X. Hence, X has dimensions $n_x \times m$, containing the m images, each one represented as a (*flattended*) column-vector with all the pixel values for its RGB decomposition, and b is the bias term.

The loss function to be minimized corresponds here with the categorical cross-entropy, defined for binary encoded labels y_i and elements a_i of matrix A, as in

$$J(w, b) = -\frac{1}{m} \sum_{i=1}^{m} y_i \log(a_i) + (1 - y_i) \log(1 - a_i). \tag{2}$$

In our experiments, we try from 1 to 15 neurons, thus being the network with only one neuron the equivalent to the neural network implementation of logistic regression. Parameters are randomly initialized once, under a Normal

distribution with zero-mean and standard deviation of 0.05, and early-stopping is used under a double cross validation scheme in order to identify the best epoch according to the error in validation, arriving at the minimum error by means of a gradient descent search under a fixed error rate of 0.01. Finally, the best model is selected according to the number of neurons and epochs that minimize the loss and exhibits the maximum accuracy on validation, according to

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}, \tag{3}$$

where TP is the number of true positives, TN the true negatives, FN the false negatives and FP the false positives.

3.2 Convolutional Neural Nets

CNNs allow examining the information contained in the images according to the possible correlations among neighboring pixels, through convolutional filters which extract attributes on low-level characteristics such as borders or contours [2]. Such a deep-convolutional architecture allows automatically tuning the filters' parameters, leading to a possibly greater performance for classification.

Here the images are not transformed to column-vectors, as they are passed to the network under their three dimensional RGB representation. The architecture for the CNN has a first hidden layer containing 32 units, each one equipped with a 3×3 filter for each of the RGB components. Then, the filters' output activate neurons with the hyperbolic tangent function and the dimensionality of the neuron's output is reduced by means of a *max-pooling* layer, here taking the maximum of each 2×2 subset of feature values with a stride of 2.

The configuration for the first layer is replicated for the second layer, and a third layer is added but this time containing 64 units and activating neurons by the scaled exponential linear unit (SELU) function, which helps in controlling the *vanishing gradient problem*, setting its activation parameters to $\alpha \approx 1.6733$ and $\lambda \approx 1.0507$ [4].

Finally the output of the third layer is flattened, obtaining a vector that is then fed into a fully connected layer with 64 neurons, which are activated by a ReLU function. Here we use also try a *dropping out* of 20% of the 64 neurons in each epoch during training, aiming at improving the generalization of the net. The resulting output is finally passed to the final *softmax* layer consisting in four neurons with a sigmoid activation function. We continue to use the categorical cross entropy loss function, and the *Adam* [3] algorithm is implemented for searching the optimal set of parameter estimators, with a learning rate of 0.001 and the β_1 and β_2 coefficients being set at 0.9 and 0.999 respectively (see again [3]).

As mentioned above, CNNs are implemented under a double cross validation scheme. Besides this, we compare the performance of such a convolutional architecture on augmented data, as it has been shown to achieve good results for image classification problems [6,5].

Firstly a pair of seeds are taken out for each class for testing the models, and on the rest of images data augmentation is performed by rotating them on

Table 1: Results for the single hidden layer and the convolutional network architecture under the three scenarios, with the accuracy (Acc) scores for Validation (V) and Test (P)

	Scenario 1		Scenario 2		Scenario 3	
Neurons	Acc(V)	Acc(P)	Acc(V)	Acc(P)	Acc(V)	Acc(P)
1	0.44	0.28	0.52	0.45	0.39	0.41
2	0.55	0.66	0.54	0.68	0.60	0.53
3	0.64	0.77	0.68	0.68	0.63	0.58
4	0.77	0.79	0.70	0.72	0.65	0.63
5	0.63	0.64	0.82	0.86	0.56	0.62
6	0.61	0.68	0.84	0.78	0.68	0.71
7	0.59	0.70	0.82	0.86	0.78	0.75
8	0.67	0.79	0.82	0.94	0.68	0.77
9	0.73	0.81	0.82	0.84	0.75	0.80
10	0.81	0.88	0.78	0.90	0.82	0.78
11	0.81	0.88	0.82	0.90	0.78	0.82
12	0.80	0.90	0.84	0.88	0.69	0.75
13	0.67	0.81	0.84	0.88	0.80	0.80
14	0.82	0.81	0.88	0.88	0.84	0.79
15	0.82	0.88	0.82	0.92	0.82	0.77
CNN	0.93	0.98	0.90	0.92	0.90	0.94

different degrees, in a range of 30 to 160 degrees, also shifting the width and height, and zooming in and out by a rate of 0.1. After the augmentation, we obtain 912, 920, 976 and 1016 images for the four species *Espeletia congestiflora, Bucquetia glutinosa, Calamagrostis effusa* and *Puya santosii*, respectively.

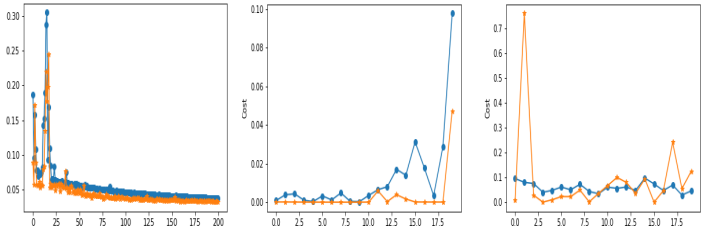
4 Results

After implementing the simulations for the different neural network architectures on the three scenarios described above, the best models can be identified for seed species prediction.

Regarding the single layer networks, the search for the optimal number of neurons in the hidden layer tries from 1 to 15, measuring the loss and the accuracy for the training, validation and test sets. The best configuration is chosen according to the minimum loss and maximum accuracy in validation, running the optimization algorithm up to 200 epochs. On the other hand, the CNNs are also trained under each scenario, without any drop-out, obtaining the results presented in Table 1.

The best result for the single hidden-layer network is obtained under the second scenario with 14 neurons and 197 epochs, achieving 88% accuracy for both validation and test sets (see Figure 2a for its learning behavior), while the best result for the CNN achieves 93% accuracy in validation and 98% in test under the first scenario. This same scenario for the CNN is repeated with 20% drop-out, obtaining a 94% on both validation and test sets with 6 epochs. Its learning

performance can be seen in Figure 2b. The resulting confusion matrices can also be seen in Table 2, where species are codified as E1: Espeletia congestiflora, E2: Bucquetia glutinosa, E3: Calamagrostis effusa, and E4: Puya santosii.



(a) (b) (c)

Fig. 2: Loss behavior on training ("o") and validation ("*") for the best (a) single-layer, (b) CNN with 20% drop-out and (c) the CNN with data augmentation.

Table 2: Confusion matrix for the best models: single hidden-layer with 14 neurons ($14-NN$), the CNN with 20% drop-out (CNN_0) and the CNN with data augmentation (CNN_a) in Validation and Test

	$14-NN$								CNN_0								CNN_a							
Pred/	Validation				Test				Validation				Test				Validation				Test			
Real	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4	E1	E2	E3	E4
E1	13	1	0	1	14	0	0	1	23	1	0	1	16	0	0	0	13	0	0	0	8	0	1	0
E2	1	7	1	1	0	9	0	2	1	21	0	0	2	17	0	1	0	9	0	0	0	4	0	0
E3	0	1	14	0	0	1	9	1	0	1	10	0	0	0	23	0	0	0	14	0	0	0	6	2
E4	0	0	0	11	0	1	0	13	0	0	0	18	1	0	0	17	0	0	0	28	0	0	0	10

The results for the different scenarios show a stable performance among the different partitions for train, validation and test sets. This is evidenced by the fact that on the three double cross-validation settings, the single hidden-layer network always achieves a maximal accuracy with 14 neurons.

Together with these experiments on double cross validation, a CNN was built after augmenting the image data base. In this way, controlling the risk of overfitting by training and validating on a bigger sample and leaving out a pair of seeds for each species for test. The results for the first 10 learning epochs can be seen in Table 3.

The respective learning-loss behavior of this CNN can be seen in Figure 2c, minimizing the validation-loss with 16 epochs, and its confusion matrix appears under (CNN_a) in Table 2, where the performance in validation is computed for a batch of size 64, and for test there are 31 available images.

Table 3: Results for the CNN architecture after data augmentation, with the accuracy scores for Training (T), Validation (V) and Test (P) for up to 10 epochs

Epoch	Acc(T)	Acc(V)	Acc(P)
1	0.78	0.89	0.75
2	0.88	0.86	0.82
3	0.96	0.96	0.95
4	0.89	0.91	0.89
5	0.95	1.00	0.92
6	1.00	0.96	0.92
7	0.94	1.00	0.97
8	0.97	0.97	0.95
9	0.98	1.00	1.00
10	0.97	0.95	0.94

5 Conclusion

The neural network approach presented in this study for seed species prediction achieved a high performance. On the one hand, under a double cross-validation scheme, a 14-neuron single hidden-layer feed-forward neural network obtained an 88% test accuracy, while a CNN achieved 94%. On the other hand, after exploring data-augmentation techniques by rotation, a CNN obtained 100% accuracy on validation and a small control-test set. Therefore, suggesting a promising methodology for seed species prediction based on optical RGB images.

The statistical experiments following a double cross validation scheme allowed controlling the risk of over-fitting with a small data set, also reducing the bias estimation for the loss in prediction. In this way, the reported performance also demonstrated a stable behavior through the different validation scenarios, where the single hidden-layer network achieving a maximal accuracy always consisted in holding 14 neurons. Besides, the greater performance of the CNNs could be confirmed after the data-augmented techniques, achieving a perfect ability to discriminate between the four species considered in this study.

For future research, more samples of seeds can be gathered for testing the models presented in this initial study, and they could also be tested under natural conditions for seed species recognition. Having in mind that the practical purpose for these models consists in detecting native seeds directly at the paramo environment, the proposed methodology should allow recognizing visible seeds lying on any surface.

References

1. Ashqar, B., Abu-Nasser, B., Abu-Naser, S.: Plant seedlings classification using deep learning. International Journal of Academic Information Systems Research **3**, 7–14 (2019)
2. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT press (2016)
3. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. arXiv:1412.6980, 2017

4. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-Normalizing Neural Networks, pp. 971–980. Advances in Neural Information Processing Systems 30, Curran Associates, Inc. (2017)
5. Wang, J., Perez, L.: The effectiveness of data augmentation in image classification using deep learning. arXiv:1712.04621, 2017
6. Wong, S., Gatt, A., Stamatescu, V., McDonnell, M.: Understanding data augmentation for classification: when to warp?, pp. 1–6. 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), IEEE Xplore, Gold Coast, QLD (2016)
7. Yang, X., Hong, H., You, Z., Cheng, F.: Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. Sensors **15**, 15578–15594 (2015)
8. Zhang, X., Liu, F., He, Y., Li, X.: Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. Sensors **12**, 17234–17246 (2012)
9. Zhao, Y., Zhang, C., Zhu, S., Gao, P., Feng, L., He, Y.: Non-destructive and rapid variety discrimination and visualization of single grape seed using near-infrared hyperspectral imaging technique and multivariate analysis. Molecules **23**, 1352 (2018)