ISSN No:-2456-2165

Advances in the Classification of Pollen Grains Images Obtained from Honey Samples of *Tetragonisca angustula* in the Province of Chaco, Argentina

Cristian G. Vizgarra
Laboratorio de Industrias Alimentarias I
Universidad Nacional Del Chaco Austral, (UNCAUS)
Pres. Roque Saenz Peña, Argentina

Matias Roodschild
Grupo de Investigacion en Tecnologias Informaticas Avanzadas (GITIA)
Universidad Tecnologica Nacional (UTN). Tucuman, Argentina

Jorge Gotay Sardiñas
Grupo de Investigacion en Tecnologias Informaticas Avanzadas (GITIA)
Universidad Tecnologica Nacional (UTN). Tucuman, Argentina

Abstract:- Several attempts have been made to automatically identify and classify pollen grains in microscopic images using computer algorithms. However, the success of pollen grain recognition depends completely on the determination of the most important features that can be used to describe it. The process of selecting the relevant characteristics is mostly done by the researcher who manually specifies the input characteristics given to the algorithm destined to solve the problem. In this article, three architectures of artificial neural networks have been selected to identify and classify pollen grains digital images without a priori establishment on the set of fundamental characteristics. For this study, eight different types of pollen grains belonging to the native flora of north-western Argentina have been utilized. The results show that the best neural classifier has an effectiveness of 95.03 % for the recognition of the eight pollen grains species. This percentage demonstrates that the methodology applied is satisfactory.

Keywords:- *Pollen Grains; Classification; Neural Networks; Supervised Learning.*

I. INTRODUCTION

The knowledge of plants preferred by bees to obtain nectar and pollen in a specific region is fundamental to rational planning on the use of the natural resources. The importance of native species in honey composition is undoubtedly due to bees' adaptation in the settlement colony vegetation. The general habit of the species contributes to this pollen diversity; this allows it to visit different sources of tropic resources at distances of 300 to 600 meters favoring their pollination. Although, there are methods for the identification in the recognition of pollen types [1], [2], [3], [4] the microscopic analysis is the most precise method of identifying the origin of pollen grains. This process is developed manually, and it is the most effective procedure nowadays used to identify and classify pollen grains. This is based solely on the observation and discrimination of characteristics of certain features of the pollen grain such as: shape, size and other characteristics of the exine (always maintaining a permanent consultation with the palynological atlas) [5] [6]. In general, this is a slow procedure and depends mainly on the training and expertise of the selected staff to analyze the pollen grain morphology [2], [3], [4], [7]. Besides, several attempts have been made to automatically identify and classify pollen grains in microscopic images using computer algorithms [8]. These approaches are combined with digital image processing techniques and self-learning. In [1], [7] some statistical methods have been studied for the identification of pollen grains by using superficial texture. The grains shape and trimming were analyzed using simple geometric measures [9]. In [10], it showed one of the first works in which texture features and neural networks were used for pollen grains identification task. The results are compared with other statistical classifiers. Although both types of classifiers may work, the neural network was apparently superior to the statistical methods. An increase in the use of the methodology of artificial neural networks for the identification and classification of pollen grains images has been observed in the last decade. In this respect, we can quote the works of [3][4][7][11][12][13]. Also, it is known that many efforts and work have been done in machine learning to obtain better representations of data characteristics using unsupervised algorithms from input data not labeled for higher level tasks, as for example, the classification. Current solutions typically learn multi-level representations by pre-training from several feature layers, one layer at a time, using an unsupervised learning algorithm [14][15][16]. In [17] three types of convolutional neuronal networks were used with a success rate of 97% in a pollen grains database (Pollen23E) corresponding to autochthonous species from the Brazilian zone. Nevertheless, analysts of pollen grains spend a lot of time in the process of manual extraction of the individual characteristics from each one. So, ideally, we would like to have algorithms that can automatically learn representations for even better functions than those created by hand. In this work, we study the effect on the selection of three different neural network training methods for the automatic learning of the characteristics associated to the eight different species of pollen grains which make the classification of digital images. The article organization: Section 2 presents the origin and description of the image

ISSN No:-2456-2165

database of the pollen grains used. Then, we propose an experimental study with the configuration of the neural network architectures used. Section 3 presents the results of the experimental study and its corresponding analysis. Section 4 presents the conclusions of this work.

II. EXPERIMENTAL SETTING AND DATASETS

A. Origin and Types of Pollen Grains

For the research, an image database of eight pollen grain species belonging to five botanical families of the northwest region of Argentina was used. It should be noted that all these species were found in stingless bees' honey samples (*Tetragonisca angustula*). They correspond to botanical species generally visited by stingless bees.

Seventy five images for each of the eight species were captured with a digital camera (Canon Powershot G10), which was connected to a microscope (Carl Zeiss Primo Star Mod. 415500). Another 75 images of each species were generated with random preprocessing operations (reflection, rotation and scaling) to increase the size of the database to 1200 images. On the other hand, each of the 256 x 256 color image were placed in gray scales (256 levels of gray) in a matrix of 128 x 128 pixels. Finally, each matrix is presented as a vector of 16.384 components, which constitute the input to the neural network.

Class	Family	Specie	Common name	Size (range of variation)
1	Euphorbiaceae	Sapium haematospermium	curupi	Medium(25-50 micra)
2	Malvaceae	Sphaeralcea bonariensis	malvavisco	Medium(25-50 micra)
3	Myrthaceae	Eucalyptus sp	Eucalipto	Small (10-25 micra)
4	Asteraceae	Vernonia chamaedrys	escobadura	Small (10-25 micra)
5	Asteraceae	Eupatorium Inolifolium	Santa misa	Small (10-25 micra)
6	Euphorbiaceae	Croton bonplandianus	Comida de paloma	Large (50-100 micra)
7	Malphigiaceae	Heteropterys glabra	Tilo de campo	Medium(25-50 micra)
8	Myrthaceae	Eugenia uniflora	Ñangapiri, pitanga	Small (10-25 micra)

Table 1:- Information about Families and Species

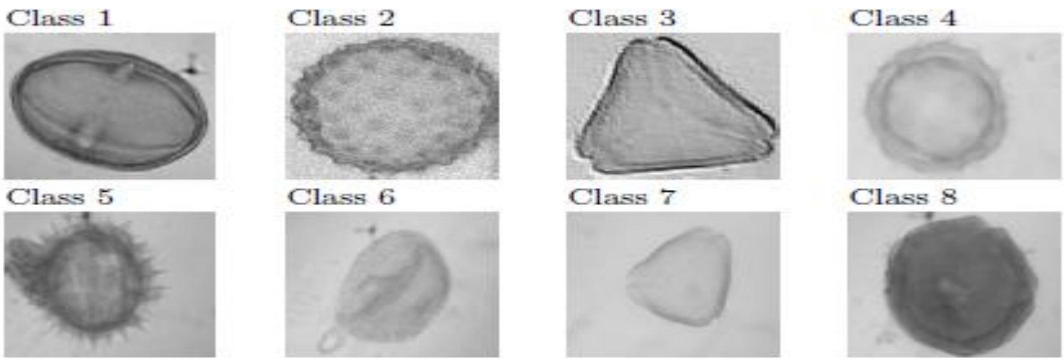

Fig 1:- Sample of Images of Each of the Eight Species of Pollen Grains

B. Data Sets

A k-fold cross-validation with k=5 at the data base of 1200 images were used in such a way that the vectors from the different references were divided into five data sets: Set A, Set B, Set C, Set D and Set E. To test the reliability of the methodology, we selected four sets as a training set, and the remaining sets were used to test the accuracy of the method. Each test set contains 240 images, 30 images of each of the respective 8 species.

Each training set contains 960 images, 120 images of each of the respective 8 species.

C. Architecture and Parameters of the Neural Network

Three neural networks with different architectures were used. The first one **(Net1)**, have two hidden layers with sigmoid neurons and an output layer with eight neurons using the softmax function. The second one **(Net2)**, have two hidden layers with sigmoid neurons too and in the output layer a support vector machine (SVM) was used for eight classes. Both networks (**Net1** and **Net2**) were pre-trained using two sparse autoencoders [18]. The sigmoid function was used to perform the coding and the linear function for decoding. In both cases, a one percent activation of the 50 and 50 neurons in the first and second hidden layer respectively. The third network (**Net3**) is a Convolution Neural Network (**CNN**) that is described in table 2 and figure 2.

Name	Type	Learnables
input	Images 128 x 128 x 1	
Conv1	3 x 3 x 64	Weigths 3 x 3 x 64 Bias 1 x 1 x 64
BN1	Batch Normalize	Two parameters
Pool1	Max Pooling 2 x 2 stride 2	
Relu1	ReLU	3 x 3 x 64
Conv2	3 x 3 x 128	Weigths 3 x 3 x 128 Bias 1 x 1 x 128
BN2	Batch Normalize	Two parameters
Pool2	Max Pooling 2 x 2 stride 2	
Relu2	ReLU	3 x 3 x 128
Conv3	3 x 3 x 96	Weigths 3 x 3 x 128 Bias 1 x 1 x 128
BN3	Batch Normalize	Two parameters
Pool3	Max Pooling 2 x 2 stride 2	
Relu3	ReLU	3 x 3 x 96
fc	Fully Connected Layer	Bias 8 x 1
Softmax	Softmax	1 x 1 x 8
Output	Output classification	

Table 2:- Information about Net3(Convolution Neural Network)

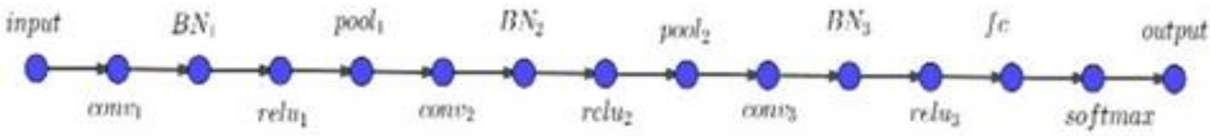

Fig 2:- Sample of structure of Convolution Neural Network (CNN)

III. RESULTS AND ANALYSIS

Training Results

When the database was divided into five data sets (cross-validated k-fold, with k = 5), 100% successful recognition was achieved in all types of neural network architectures in training sets. However, for the test sets the average success recognition was 84.02% and 73.40% and 95.03% for Net1 and Net2 and Net3 architecture respectively. Table 3 shows the results of the accuracy and testing errors of each of the subsets considered for each neural network. By using the third neural network architecture (Net3), which is formed by three convolution layers, the results in the classification were consistently superior in each test set. The first Neuronal network (Net1) had a better performance than the second neural network (Net2) as indicated in Table 2. In Figure 3, the best results of the neural classifier Net3 are shown in the set of tests. It is observed that species 1,2,4,5,6 and 8 are recognized with 100 % success. However, the classes 3 and 7, two species were erroneously placed in classes 6 and (4 and 6) respectively.

Net	Set A	Set B	Set C	Set D	Set E
1	85% 15 %	85 % 15 %	86,3 % 13.7 %	86,3 % 13.7 %	77,5 % 22.5 %
2	73 % 27 %	73 % 27 %	75 % 25 %	75 % 25 %	71 % 29 %
3	94.66% 5.34%	93.8% 6.2%	98,8% 1.2%	95.6% 4.4%	92.3% 7.7%

Table 3:- Results of Cross-Validation in the Test Sets

ISSN No:-2456-2165

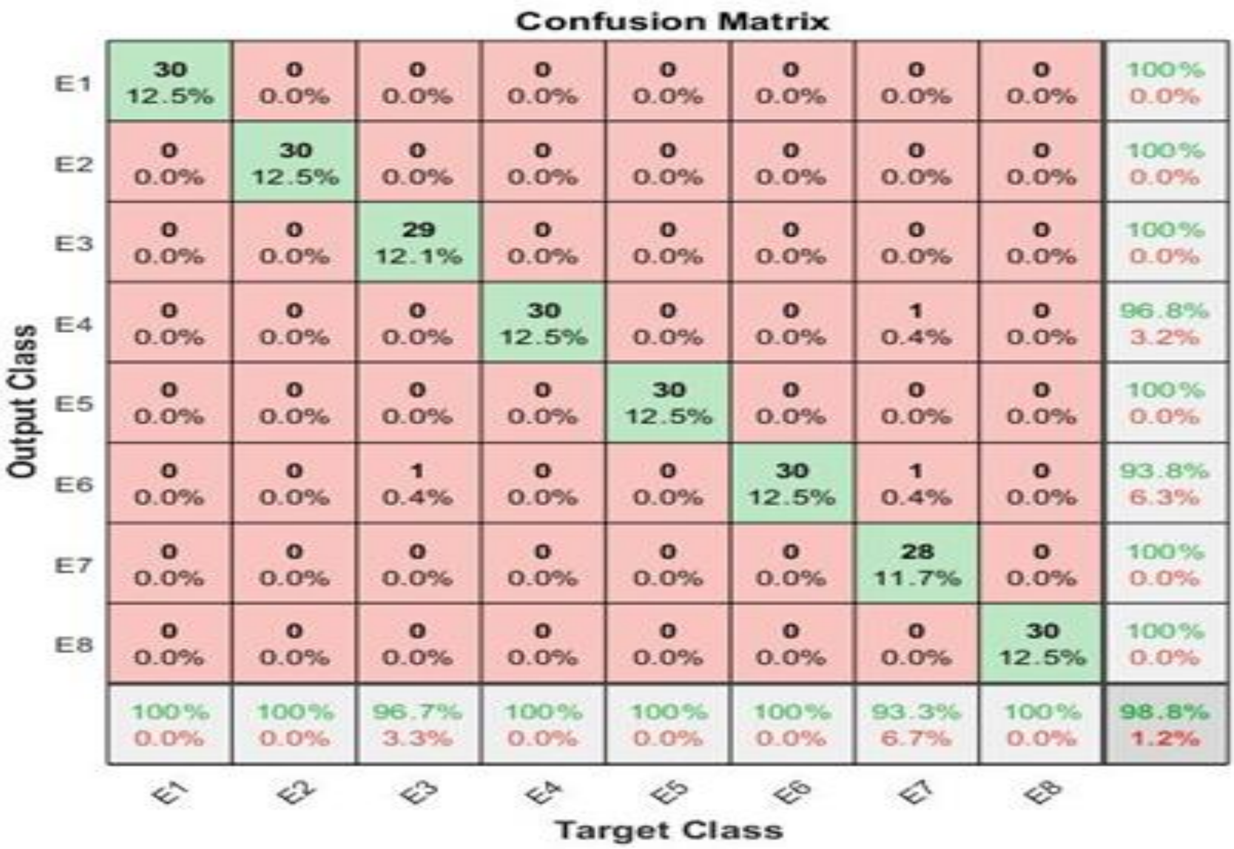

Fig 3:- Best Results of the Neural Classifier (Net3)

IV. CONCLUSIONS

All in all, three architectures of neural networks were used to classify pollen grains from digital images, without specifying a priori the set of most important characteristics which are critical for a successful classification. In all types of architectures, the neural classifiers offered performances that reached 84%, 73.4% and 95.03% of successful recognition in the Net1, Net2 and Net3 respectively.

The use of convolution neural networks (CNN) represents a more efficient way to automatically extract the characteristics of pollen grains images to make a classification from levels of abstraction higher than the one used to make a classification based on pixel data without processing.. The convolution neural networks come to solve the problem that ordinary neural networks do not scale well for much defined images. The proposed approach, without a doubt, provides a new expectation in the simplification of the arduous, tedious and complex task of the identification and classification of the pollen grains that biologists and palynologists must perform.

ACKNOWLEDGMENT

We would like to thank all the staff of the Food Industries Laboratory I and in general, the staff and authorities of Universidad Nacional del Chaco Austral (UNCAUS) for the unconditional support provided in this research.

REFERENCES

[1]. I. France, A. Duller and H. Lamb. “A new approach to automatic pollen analysis”. Quatenary Science Reviews, (19):pages 537-546. 2000.
[2]. A. Boucher, P. Hidalgo, M. Thonnat, J. Belmonte, C. Galan, P . Bonton and R. Tomczak. “Development of a semi-automatic system for pollen recognition”. Aerobiologia, pages 195-201. 2002.
[3]. M. Rodriguez-Damian, E. Cernadas, A. Formella, M. Fernandez-Delgado and P. Sa-Otero. “Automatic detection and classification of grains of pollen based on shape and texture”. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, (536):pages 531-542. 2006.
[4]. N. Khorissi, A. Mellit, A. Guessoum and A. Mesaouer. “Ga-based feed-forward neural network for image classification: Application for grains of pollen”. Applied Computer Science, (17(2)):pages 83-96. 2009.
[5]. LBenitez Bosco and C. Fernandes Pinto da Luz. “ Pollen analysis of Atlantic forest honey from the Vale do Ribeira Region, state of São Paulo,Brazil, Grana. (2018).
[6]. A. Volkan & Karahan, F. Pınar & Karahan and M. Ozturk. Pollen analysis of honeys from Hatay/Turkey. 11. 209-222. 2018).
[7]. M. Chica and P. Campoy. “Discernment of bee pollen loads using computer vision and one-class classification techniques”. J. Food Eng, (112):pages 50-59. 2012.

[8]. M. Fernandez-Delgado, P. Carrion, E. Cernadas, J. Galvez and P. Otero. "Improved classification of pollen texture images using svm and mlp". In benalmadena, editor, Proceedings of international conference on visualization, imaging, and image processing, volume 2. 2003.

[9]. W. Treloar, G. Taylor and J. Flenley. "Towards automation of palynology 1: Analysis of pollen shape and ornamentation using simple geometric measures, derived from scanning electron microscope images". Journal of Quaternary Science, (19(8)):745-754. 2004.

[10]. P. Li and J. Flenley. "Pollen texture identification using neural networks". Grana, (38):59-64. 1999.

[11]. G. P. Allen, R. M. Hodgson, S. R. Marsland and J. R. Flenley. "Machine vision for automated optical recognition and classification of pollen grains or other singulated microscopic images". Fifteenth International Conference on Mechanotronics and Machine Vision in Practice. 2008.

[12]. K. Holt, G. Allen, R. Hodgson, S. Marsland and J. Flenley. "Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory". Review of Palynology and Palaeobotany, (167):175-183. 2011.

[13]. J. Oteros, H. Garca-Mozo, C. Hervas-Martinez and C. Galan."Year clustering analysis for modelling olive owering phenology". Int. J. Biometeorol, (57):545-555. 2013.

[14]. G. Hinton, S. Osindero, and Y. Teh. "A fast learning algorithm for deep belief nets". Neural Computation, 18(7):1527 -1554. 2006.

[15]. K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun "What is the best multi-stage architecture for object recognition?".International Conference on Computer Vision. 2009.

[16]. H. Lee, C. Ekanadham, and A.Y. Ng "Sparse deep belief net model for visual area". v2. NIPS. 2008.

[17]. V. Sevillano and J. L. Aznarte . "Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks". PLOS ONE. https://doi.org/10.1371/journal.pone.0201807. 2018.

[18]. N. Nguyen, M. Donalson-Matasci and M. Shin. "Improving Pollen Classification with Less Training Effort". Applications of Computer Vision (WACV). 2013.