Detection and Classification of Pollen Grain Microscope Images

Sebastiano Battiato
University of Catania
95125 Catania, Italy
battiato@dmi.unict.it

Alessandro Ortis
University of Catania
95125 Catania, Italy
ortis@dmi.unict.it

Francesca Trenta
University of Catania
95125 Catania, Italy
francesca.trenta@unict.it

Lorenzo Ascari
University of Turin
10125 Turin, Italy
lorenzo.ascari@unito.it

Mara Politi
University of Turin
10125 Turin, Italy
mara.politi@unito.it

Consolata Siniscalco
University of Turin
10125 Turin, Italy
consolata.siniscalco@unito.it

Abstract

With the spread of technology in several fields, there is an increasing demand to automate specialized tasks that usually require human involvement in order to maximize efficiency and reduce processing time. Pollen identification and classification is a proper example to be treated in the Palynology field, which has been an expensive qualitative process, involving observation and discrimination of features by highly qualified experts. Although it is the most accurate and useful method, it is a time-consuming process that slowed down the research progress. In this paper, we present a dataset composed of more than 13.000 objects, identified by an appropriate segmentation pipeline applied on aerobiological samples. Besides, we present the results obtained from the classification of these objects by taking advantage of several Machine Learning techniques, discussing which approaches have produced the most satisfactory results, and outlining the challenges we had to face to accomplish the task.

1. Introduction

The monitoring of pollen and fungal spores concentrations allows the detection and the quantification of allergenic species [18] and of potential infectious diseases both on humans [9] and plants [21]. For this reason, the classification of pollen species and types is an important task in many areas such as medicine, biology, and agronomy, among others. However, the long sample processing times and manual counting of relevant entities in microscopy have slowed down the spread of the application of aerobiology in such brand new sectors [1]. Although the recent introduction of devices that allows the identification and classification of pollen grains without the need for end-user intervention [15, 11], manual Hirst-type spore-traps are still predominant [2], thanks to lower costs and established methods for the analysis of samples. In the last few years, modern Machine Learning methods based on deep neural networks have led to an impressive performance on a variety of problems in different fields [14, 19, 7]. One of the key elements of Deep Learning success is the availability of a high amount of annotated data. Indeed, the performances of the Deep Learning techniques scale with the amount of data, promoting the definition of large-scale datasets in different scientific fields. For the above reasons, research on methods for automatic classification of pollen grains will foster the development of tools for aerobiologists that could be used without the need for special equipment other than a common bright-field microscope, a camera for image acquisition and a computer for data processing. These tools would be able to segment pollen grains from images and assist scientists in grouping them by species. In this work, we present a large-scale dataset of microscope pollen grain images, collected from aerobiological samples. Experts in the aerobiology field used Hirst-type spore-traps and standard procedures for the sampling and staining steps. Then, microscope images of the samples have been digitalized and processed through a proper image processing pipeline to detect and extract five classes of objects, including four species of the pollen grain and an additional class that could often be misclassified as pollen (e.g., air bubbles, dust, etc.). The proposed dataset contains more than 13.000 objects derived from microscope images and labeled by experts. This paper is organized as follows. Previous related works exposing dataset that have already employed for automatic pollen grain classification are presented and reviewed in Section 2. In Section 3 we then provide our collected and processed dataset together with a description of the pipeline implementation. Section 4 shows the numerical and graph-

2160-7516/20/$31.00 ©2020 IEEE
DOI 10.1109/CVPRW50498.2020.00498

Dataset	Number of Grains	Image Type	Resolution
Duller's Pollen Dataset [8]	630	Grayscale	25x25
POLEN23E [10]	805	Color	Minimum 250 pixel per dimension
Ranzato et al. [17]	3.686 (1.429 images)	Color	1024x1024 (multiple grains per image)
Proposed Dataset	>12.000 + ∼1.000 examples of debris (e.g., dust, air bubbles)	Color	84x84

Table 1. Comparison between the proposed dataset and the main datasets used in pollen grain classification.

ical results of our experiments. Lastly, a brief exposition on possible future research is presented in Section 6.

2. Related Works

Existing works on automatic pollen grain detection/classification are trained and evaluated on datasets which include from 65 to about 4.000 number of grains, most of them report results obtained on self-collected databases. Two public databases are the Duller's Pollen Dataset [8] and the POLEN23E [10]. The first, contains a total of 630 grayscale images of size 25x25. Specifically, the database consists of a certain number of images of a few pollen taxa. The raw images were collected using a charge coupled device (CCD) camera with a resolution of 768 by 576 pixels and then processed to identify pollen grains. The authors used an algorithm that takes a series of images at different focal points after detecting a grain via microscope. A database containing 805 color images from 23 different pollen types in the Brazilian savannah, with 35 images for each pollen type has been presented in [10] where the authors proposed an algorithm that combined three types of features: color, shape and texture based features. Moreover, the authors evaluated the performance of computational techniques for the pollen grains classification. Specifically, they used four classifiers: Sequential Minimal Optimization (SMO) and C-Support Vector Classification (C-SVC), a decision tree based classifier and the k-nearest neighbors (KNN). The results show that the computational techniques achieved satisfactory results. Especially, they achieved the highest performance values when using features extracted from texture, color and shape in combination with a Bag of Visual Word (BOW), reporting a Correct Classification Rate (CCR) 64%. In this context, BOW refers to the local information found in specific regions. For what concerns the pollen grain classification, larger datasets have been proposed for the task of pollen grain detection such as the ones used in [17]; however, the number of grains is between 3.000 and 4.000 objects. Furthermore, in [17] the authors proposed a system to recognize biological particles. The proposed system is composed of four stages: First, they detected locations or small regions containing interesting samples by using a feature finder. Second, differential invariants of the brightness are computed at multiple scales of resolution. Third, after applying a point-wise non-linear mappings to a higher dimensional feature space, this information is averaged over the whole region thus producing a vector of features. In the final stage, each sample is classified by a mixture-of-Gaussians generative model. The results shed light on the effectiveness of the proposed solution, achieving a correct classification rate of 83% (accuracy) for 3 categories of particles. In Table 1, we summarized the composition of the datasets herein described, providing a preliminary comparison with the proposed dataset.

(a) (b) (c) (d) (e)

Figure 1. Examples of acquired samples. (a) *Corylus avellana* (well-developed pollen grains), (b) *Corylus avellana* (anomalous pollen grains), (c) *Alnus* (well-developed pollen grains), (d) Debris, (e) *Cupressaceae*.

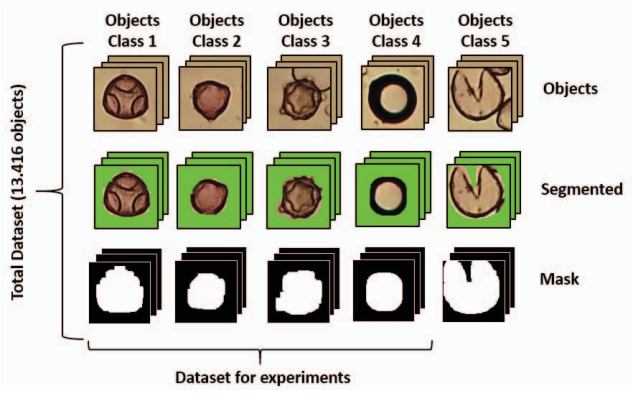


Figure 2. The proposed dataset that includes for each object the related binary mask and the segmented object.

3. Material and Methods

3.1. Dataset

The main contribution of this work is the availability of a novel dataset that includes 13.416 objects of pollen grains. Experts in aerobiology manually labeled each object by using a web-based tool. In order to pre-process and analyze the aerobiological samples, they used segments of tape in which the pollen grains adhered. Each segment of tape was placed on a rotating drum, moved at 2 mm h-1 under a suction hole. The daily segments of the pollen grains have been inspected by making use of a Leitz Diaplan bright-field microscope and a 5 MP CMOS sensor. By the standard procedures, the pollen walls placed on the microscope slides were selectively stained with a mounting medium containing basic fuchsin (0,08 % gelatin, 0,44% glycerin, 0,015% liq-

uefied phenol, 0,0015% basic fuchsin in aqueous solution). The proposed dataset is composed of more than 13.000 objects spanning over 5 different categories: (1) *Corylus Avellana* (well-developed pollen grains), (2) *Corylus Avellana* (anomalous pollen grains), (3) Alnus, (4) Debris, (5) *Cupressaceae* (see Fig. 1). We collected objects for each class alongside the related binary mask and the segmented object with green background. An amount of 63 images in the dataset represents pollen grain objects overlapped with a non-pollen one. To avoid the misclassification of these images, we decided not to insert them into the dataset used for the experiments. Considering the small number of observations related to *Cupressaceae* class (43), we did not include them in the dataset used for the experiments. Therefore, the total number of objects in the dataset for experiments is 13.310 (see Fig. 2).

3.2. Processing pipeline

This section introduces the designed pipeline which consists of three main stages: pre-processing, segmentation and mask post-processing.

Pre-processing. As reported earlier, the automatic detection of pollen grains is a challenging task due to the presence of heavy background noise in the digitalized slides scans, which can drastically affect the performance of segmentation methods. For this reason, we designed a proper pre-processing pipeline with the aim of reducing background noise and improving the high segmentation quality in the presence of several problematic conditions deriving either from the manual sectioning of the aerobiological sample itself (i.e., debris and dust and fungal spores) or from the mounting technique (air bubbles). The first step of the proposed pipeline consists of applying the OpenCV implementation of mean shift algorithm [1]. The output of this procedure is an image with color gradients and fine-grain texture flattened. In the second stage, we split the pipeline into two different steps. First, we developed a procedure to smooth background artifacts to improve object detection in the foreground. In order to discriminate the foreground from the background of the mean-shifted image, we used Otsu's method [16] in combination with a binary threshold, set to 127. Under the supervision of experts in aerobiology, we observed that pollen objects are usually higher than 500 in diameter. Based on these assumptions, we removed all the objects with a size smaller than this by using connected components with eight neighbors. Finally, we applied the resulting binary mask to the input image and changed the color of the detect contours in yellow. The second step of this stage consists of applying an 11×11 kernel Gaussian filter to the input image in order to blur the artifacts in the background. In the final stage, we combined both the output images from previous steps. This step is essential not only to distinguish objects of the foreground from the other ones in the background but also to guarantee that the following image processing will not damage the object contours, influencing the effectiveness of the image segmentation pipeline. In Fig. 3, we illustrated the described steps to highlight objects contours.

Segmentation pipeline. The object segmentation pipeline aims to partition an image by maximizing the pollen grains detection alongside reducing the detection of non-pollen ones (i.e., dust, artifacts, etc.). With this aim, after changing the color space of the output image of the previous stage from RGB to HSV, we transformed the image in grayscale and applied a binary threshold to it. In order to reduce the noise background generated by previous image processing steps, we applied a closing operator followed by dilation using a 3×3 kernel for both operations. In this instance, we implemented the flood fill algorithm intending to distinguish the foreground from the background by reassigning the values of all neighboring pixels of a given point with a required uniform color. Therefore, all the objects of interest have been filled with black color, whereas the background has been filled with white color. Through analyzing connected components in the image, we also removed objects with a size smaller than 100 pixels in diameter.

Mask post-processing pipeline. The developed procedure for mask post-processing is fundamental to improve the image quality of the output image derived from the previous steps. The main idea is to generate a binary mask for each object, developing a closed outline for the object's boundary in a manner that it outlines the object contours tightly and also minimize the presence of background noise. To this aim, we applied a pool of effective image processing functions. After applying a mean shift algorithm, we used an adaptive threshold function. Formally, the function converts a grayscale image into a binary image by using a given threshold. The threshold value is calculated individually for each pixel in the input image. Specifically, we used an adaptive Gaussian threshold. In the final step, a procedure based on selecting connected components (objects) with a size greater than 150 pixels of diameter is developed in order to reduce background noise. Once applied the flood fill algorithm, we used a dilation operator with a 3×3 kernel with full of ones to increase the size of the object in binary mask image, iterating the process for 5 times. In Fig.4, we reported the overall pipeline.

4. Experiments

This section presents the results obtained for a pool of Machine Learning algorithms, comparing the performance of the techniques herein proposed for pollen grains classification. Also, we addressed the problem of the imbalanced dataset and how we deal with it.

[1] https://docs.opencv.org/2.4/modules/imgproc/doc/filtering.html?highlight=meanshiftfiltering

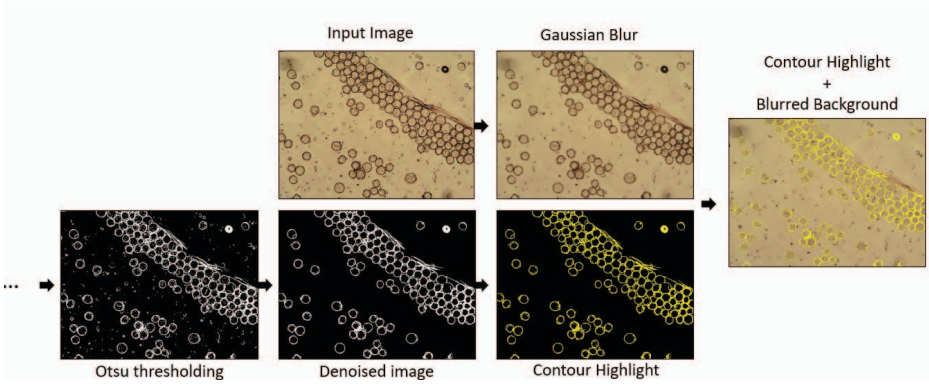


Figure 3. Pipeline used to highlight contours objects.

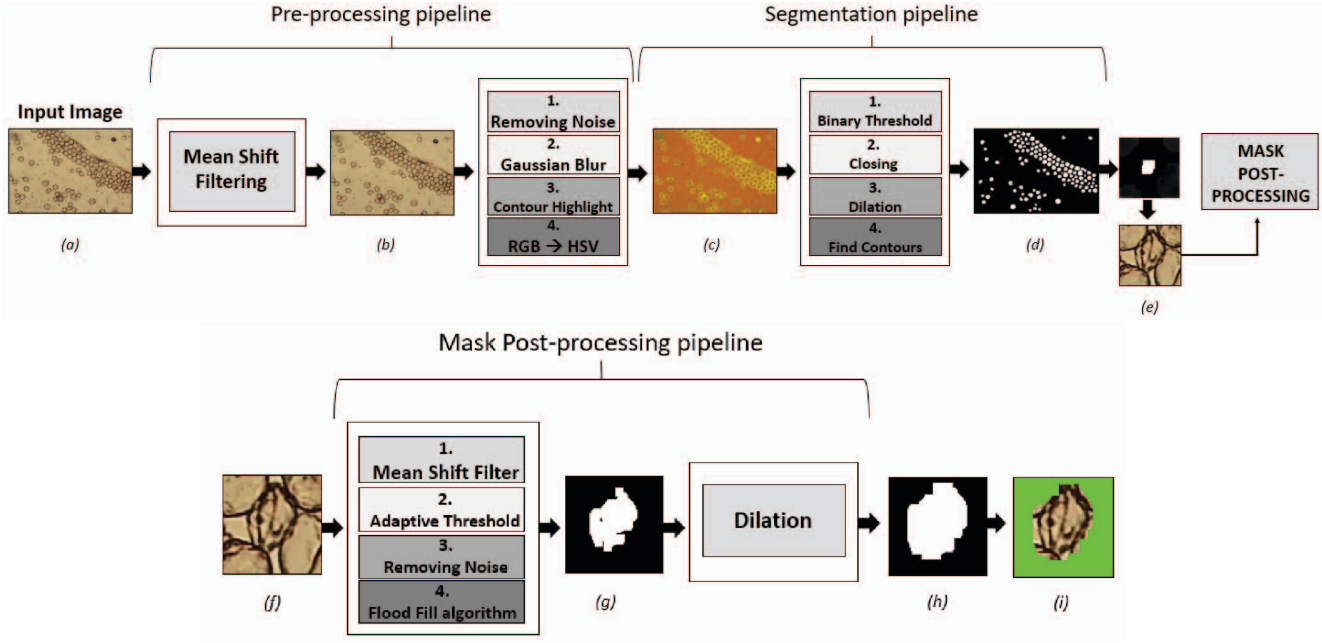


Figure 4. The overall pipeline. (a) Image of an aerobiological sample, (b) image after applying mean shift filtering function, (c) the resulting image after converting color space from RGB to HSV derived from previous steps, (d) the mask generated by applying binary threshold, closing and dilate operators, (e) image after applying binary mask to input image (f) the detected object, (g) the resulting binary mask after applying a mean shift filter and adaptive threshold, (h) the obtained binary image and (i) the related segmented object from original patch.

4.1. Experiments with LBP and HOG features

In this stage, we carried out our experiments via classical Machine Learning techniques, such as Random Forest, Support Vector Machines (SVM), AdaBoost, among others [3]. In order to introduce the data to a classifier, we first needed to extract features from the raw data. Specifically, we investigated object texture by using Local Binary Pattern (LBP) [13] and Histogram of Oriented Gradient (HOG) [6]. In order to perform classification, we divided the data into a training set and a test set. To do so, we have considered the first 85% of the data as the training set and the remaining 15% as test set. In this stage, the classification is performed with the dataset composed of segmented images with green background in order to consider the only texture of the detected object. Also, we carried out our experiments by using the following models: Linear Support Vector Machine (SVM), RBF SVM, Random Forest, AdaBoost, Multi-Layer Perceptron (MLP) [3]. To handle the problem of imbalanced data, we first selected optimal hyperparameters by applying the Grid Search algorithm [2]. and computing the average accuracy at each run. Therefore, we evaluated the performance of each classifier by using the proposed imbalanced dataset, relying on optimal hyperparameters that have been selected from the previous step. To

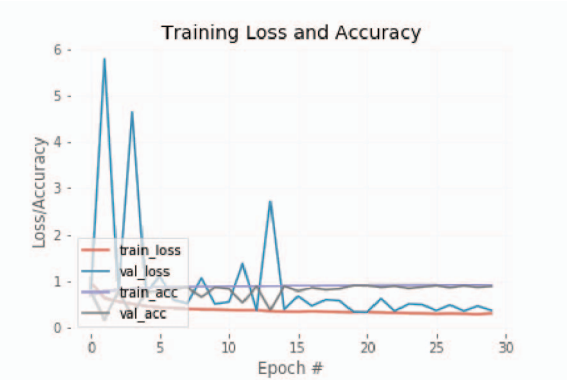
[2] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

carry out our experiments, we used penalized classification models for SVM and Random Forest algorithms. Penalized models help focusing on the minority class by adjusting weights inversely proportional to class frequencies in the training data. The major issue of the class imbalanced dataset is the lack of samples of a given class, which could lead to poor results. Also, we performed a stratified train-test splitting to ensure that both training and test set featuring the same percentage of classes. To evaluate the performance of each classifier, we used the weighted F1 score for quantitative evaluation [4], which represents a more reliable performance metrics than accuracy in this context. The weighted F1 score function calculates the F1 metrics for each class, and their average weighted by support (i.e., the number of true instances for each class). With regard to LBP, Table 2 reports the experiments related to the LBP feature representation for the input pollen images, considering the evaluated parameters for each employed classification algorithm. Results show that MLP leads to the best results in terms of accuracy (0.8002) and F1 score (0.7764) with an alpha value equal to 0.0001 and a number of estimators equal to 500. We also observed that Linear SVM, Random Forest, and AdaBoost performed accurate results yielding an accuracy and F1 score of over greater than 0.70, whereas SVM with RBF kernel [5] showed worst performance. The classifier achieved an accuracy of 0.6430 and an F1 score of 0.6714, considering a gamma value of 1.0 and a C value of 1000. Also, Table 2 reports the experimental results obtained using the HOG features. In this instance, SVM with RBF kernel achieved the highest value in terms of accuracy (0.8658) and F1 score (0.8566), considering a gamma value of 0.1 and a C value of 1000. This specific setting also outperforms the performances of the approaches that take the LBP representation. In particular, all the evaluated approaches detailed in Table 2 achieved accuracy and F1 scores higher than 0.70.

Methods	**Parameters**		**HOG**	
			Accuracy	**F1**score
LINEAR SVM	$C = 1000$		0.7646	0.7673
RBF SVM	G = 0.1	C = 1000	**0.8658**	**0.8566**
RANDOM FOREST	$EST = 10$		0.7616	0.7124
ADABOOST	LR = 0.5	EST = 500	0.7752	0.7627
MLP	a = 0.1	EST = 300	0.8493	0.8431
Methods	**Parameters**		**LBP**	
			Accuracy	**F1**score
LINEAR SVM	$C = 100$		0.7446	0.7439
RBF SVM	G = 1.0	C = 1000	0.6430	0.6714
RANDOM FOREST	$EST = 1000$		0.7792	0.7387
ADABOOST	LR = 1.0	EST = 100	0.7722	0.7487
MLP	a = 0.0001	EST = 500	0.8002	0.7764

Table 2. Comparison between the best results by using HOG and LBP features.

(a)



(b)

Figure 5. AlexNet training loss and accuracy. (a) Loss/accuracy without using augmented dataset (b) Loss/accuracy with using augmented dataset.

4.2. Experiments with Convolutional Neural Networks

For the sake of comparison, we also performed object classification through the use of a Deep Convolutional Neural Network (i.e., AlexNet [12]). Considering that CNNs and Deep Learning models take advantage of a huge amount of data, we trained two CNN standard architectures (AlexNet and SmallerVGGNET) considering two different settings for the training data. First, we trained the CNNs considering the standard dataset, composed of the patches depicting the pollen objects. In a second stage, we augmented the dataset by including the segmented version of the training patches obtained by applying the segmentation mask and padding the background with all green pixels. This can be considered as an additional approach for data augmentation, which helps the CNN to focus on the pollen grain in the image, as well as ignoring the remaining elements present in the background. For both experiments, we also performed additional data augmentation by performing

Epoch	AlexNet - SD		SmallerVGGNET - SD	
	Accuracy	**F1_score**	**Accuracy**	**F1_score**
10	0.8067	0.7709	0.3831	0.4463
20	0.6830	0.7130	0.8428	0.8156
30	0.7992	0.7447	0.8277	0.7976
Epoch	AlexNet - AD		SmallerVGGNET - AD	
	Accuracy	**F1_score**	**Accuracy**	**F1_score**
10	0.8528	0.8507	0.8618	0.8465
20	0.8943	0.8890	**0.8973**	**0.8914**
30	**0.8963**	**0.8897**	0.8408	0.8139

Table 3. Classification performances of AlexNet and SmallerVGGNet by using Standard Dataset (SD) and Augmented Dataset (AD).

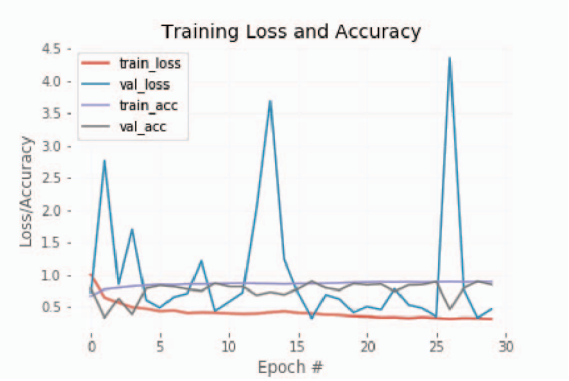


Figure 6. SmallerVGGNET training loss and accuracy. (a) Loss/accuracy without using augmented dataset (b) Loss/accuracy with using augmented dataset.

random horizontal and vertical flipping and random rotation range of 25° in training data. Moreover, we set the base learning rate to 0.001, the batch size to 64, and a number of epochs to 1000. To manage the overfitting problem, we introduced an Early stopping function, which allows us to stop training once the model performance stops improving on a hold out validation set. After every 10 epoch, we evaluated the network performance on the test set. AlexNet yielded an average F1 score of 0.87 using the augmented dataset. With regard to the standard dataset, it could be observed that AlexNet achieved an average F1 score of 0.74, as reported in Table 3. Fig.5 (a) and Fig.5 (b) show the performance plots related to the training of AlexNet considering the standard and the augmented dataset, respectively. In particular, each plot shows the loss and accuracy of the training and validation data over 30 epochs, when Early Stopping occurred. It can be observed that the use of the augmented dataset, besides improving the model performances, helps in maintaining more stable the loss and accuracy fluctuation over the epochs after just 5 epochs. Table 3 details the accuracy and F1 results measured on the test set every 10 epochs, considering the standard (SD) and the augmented dataset (AD), respectively. AlexNet outperforms the standard ML approaches only when the augmented dataset is considered. For the sake of completeness, we carry out our experiments by also using a different CNN architecture, which could represent a more suitable solution for classifying images with a small size. In our case, we implemented a SmallerVGGNet, which is a variant of Very Deep Convolutional Networks (VGGNet) [20] by employing the same parameters used for AlexNet. SmallerVGGNet achieved an average F1 score of 0.85 using an augmented dataset and an average F1 score of 0.69 using images from the standard dataset. Fig.6 (a) and Fig.6 (b) show the performance plots related to the training of SmallerVGGNET considering the standard and the augmented dataset, respectively. Also, in this case, the use of the augmented dataset improves the model performances and helps in maintaining more stable the loss and accuracy fluctuation over the epochs. Table 3 details the accuracy and F1 results measured on the test set by using the standard and the augmented dataset, respectively. SmallerVGGNET trained on the augmented dataset outperforms all the previous approaches reaching an accuracy of 0.8973 and an F1 score of 0.8914 after 20 epochs (see Table 3). In our experiments,

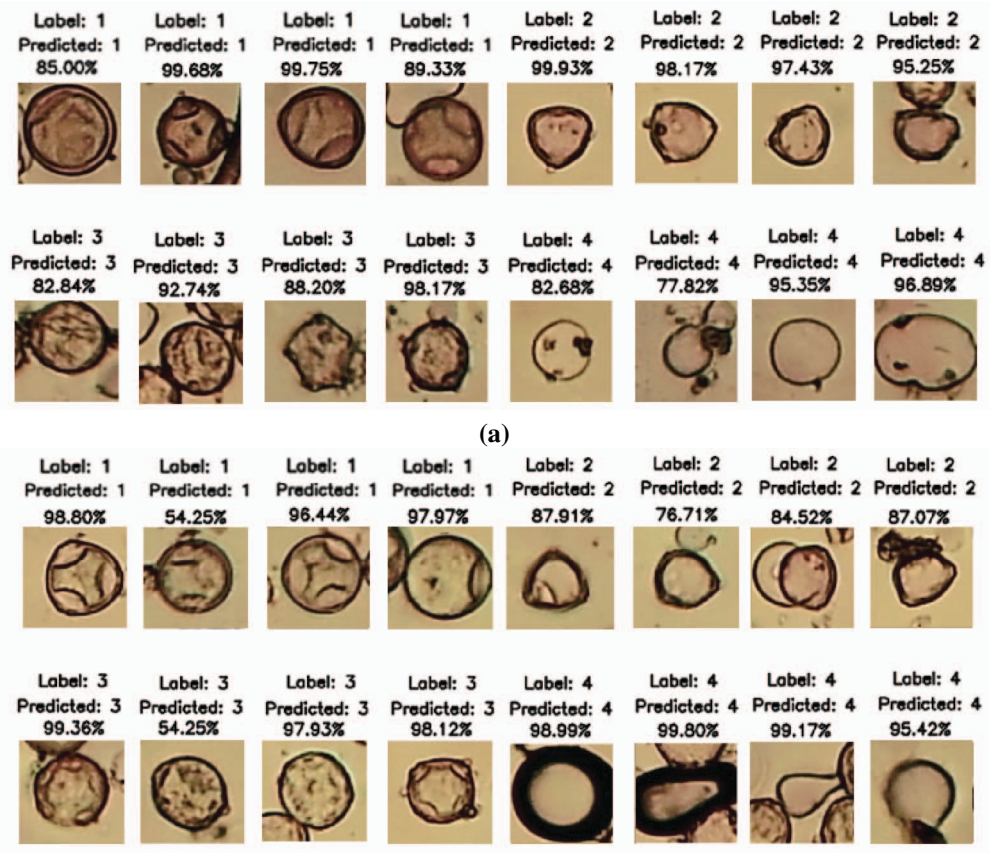


Figure 7. Example of good classification performed by (a) AlexNet and (b) SmallerVGGNET. Each raw includes 8 examples of each involved objects correctly classified together with its confidence score.

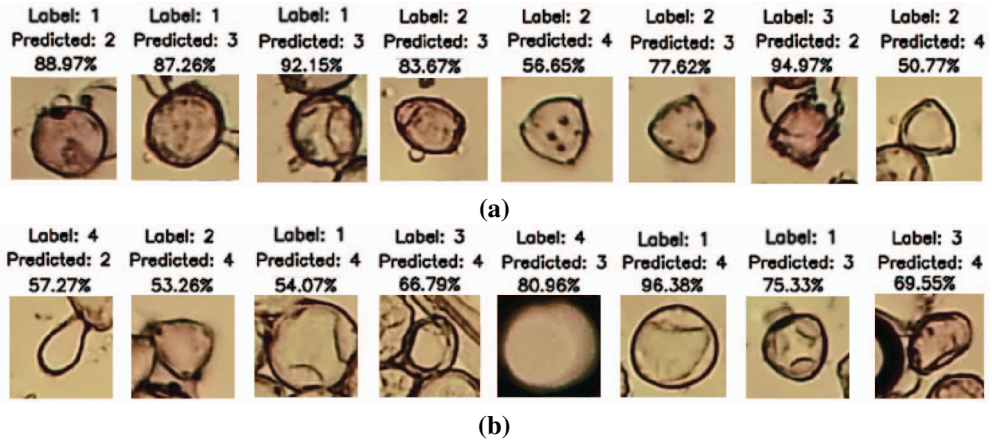


Figure 8. Example of bad classification performed by (a) AlexNet and (b) SmallerVGGNET, together with the confidence score.

we observed that both AlexNet and SmallerVggnet classify pollen objects with a high confidence score. Regarding AlexNet, it tends to classify better the *Corylus Avellana* objects of category 1 (well-developed pollen) and 2 (anomalous pollen). Although AlexNet is able to classify the majority of the pollen grains accurately, it tends to misclassify pollen objects of *Corylus Avellana* (well-developed pollen) class with a texture similar to the object of *Alnus*

class. Furthermore, the highest error occurred in AlexNet when the objects present features (texture) different from the peculiar ones of their class (see Fig. 8 (a)). In this instance, AlexNet is not able to classify them accurately. Compared to AlexNet, SmallerVGGnet achieved lower accuracy and F1 score when applied to objects of categories 1 and 2. However, under challenging conditions where a pollen object is overlapped with another one (pollen or Debris), SmallerVGGNET achieved impressive results. In general, the objects of class 2 and 4 are the most confused by the model. In Fig. 7, examples of good classification performed by AlexNet and SmallerVGGNET are reported, respectively. In Fig. 8 we reported examples of misclassification performed by both CNN models. Each example details the true and predicted class, as well as the confidence of the model.

5. Acknowledgements

The research has been carried out thanks to the collaboration with Ferrero HCo, that financed the project and allowed the collection of aerobiological samples from hazelnut plantations.

6. Conclusions and Future Works

In this paper we presented a large-scale pollen image dataset. We processed microscope images to detect and extract the depicted objects. Indeed, the project involved experts on Agronomy, Aerobiology and Computer Vision. In this paper, we investigated the performance of common Machine Learning approaches on the pollen grain classification task. The presented work is meant to support the research on automatic pollen grain classification systems, that can now leverage on the first large-scale annotated dataset of microscope pollen grain images. In future studies, we aim to extend our approach by defining an appropriate Deep Learning architecture to improve the reliability of the classification results.

References

[1] Principles and methods for automated palynology. *New Phytol*, (1996):1–8, 2014. 1

[2] Pollen and spore monitoring in the world. *Clinical and Translational Allergy*, 8(1):1–5, 2018. 1

[3] C. M. Bishop. *Pattern Recognition and Machine learning*. Springer, 2006. 4

[4] N. A Chinchor and B. Sundheim. Message understanding conference (muc) tests of discourse processing. In *Proc. AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation*, pages 21–26, 1995. 5

[5] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995. 5

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005. 4

[7] Arwinder Dhillon and Ashima Singh. Machine learning in healthcare data analysis: A survey. 2019. 1

[8] A. Duller, G. Guller, I France, and H Lamb. A pollen image database for evaluation of automated identification systems. *Quaternary Newsletter*, pages 4–9, 1999. 2

[9] A. Fernstrom and M. Goldblatt. Aerobiology and Its Role in the Transmission of Infectious Diseases. *Journal of Pathogens*, 2013:1–13, 2013. 1

[10] A. B. Gonçalves, J. S. Souza, G. G. Da Silva, M. P. Cereda, A. Pott, M. H. Naka, and H. Pistori. Feature extraction and machine learning for the classification of Brazilian Savannah pollen grains. *PLoS ONE*, 11(6):1–20, 2016. 2

[11] S. Kawashima, M. Thibaudon, S. Matsuda, T. Fujita, N. Lemonis, B. Clot, and G. Oliver. Automated pollen monitoring system using laser optics for observing seasonal changes in the concentration of total airborne pollen. *Aerobiologia*, 33(3):351–362, 2017. 1

[12] A. Krizhevsky, I. Sutskever, and G. E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012. 5

[13] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. 4

[14] A. Ortis, G. M. Farinella, and S. Battiato. A survey on visual sentiment analysis. *IET Image Processing*, 2020. 1

[15] J. Oteros, G. Pusch, I. Weichenmeier, U. Heimann, R. Möller, S. Röseler, C. Traidl-Hoffmann, C. Schmidt-Weber, and J. T.M. Buters. Automatic and online pollen monitoring. *International Archives of Allergy and Immunology*, 167(3):158–166, 2015. 1

[16] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. 3

[17] M. Ranzato, P. E. Taylor, J. M. House, R. C. Flagan, Y. LeCun, and P. Perona. Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters*, 28(1):31–39, 2007. 2

[18] H. Ribeiro, S. Morales, C. Salmerón, A. Cruz, L. Calado, M. I. Rodríguez-García, J. D. Alché, and I. Abreu. Analysis of the pollen allergen content of twelve olive cultivars grown in Portugal. *Aerobiologia*, 29(4):513–521, 2013. 1

[19] F. Rundo, F. Trenta, A. L. di Stallo, and S. Battiato. Machine learning for quantitative finance applications: A survey. *Applied Sciences*, 9(24):5574, 2019. 1

[20] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 6

[21] J. S. West and R. B E Kimber. Innovations in air sampling to detect plant pathogens. *Annals of Applied Biology*, 166(1):4–17, 2015. 1