Deep Learning Recognition of a Large Number of Pollen Grain Types

Fernando C. Monteiro$^{(\boxtimes)}$, Cristina M. Pinto, and José Rufino

Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto
Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal
`{monteiro,rufino}@ipb.pt`
`https://cedri.ipb.pt`

Abstract. Pollen in honey reflects its botanical origin and melissopalynology is used to identify origin, type and quantities of pollen grains of the botanical species visited by bees. Automatic pollen counting and classification can alleviate the problems of manual categorisation such as subjectivity and time constraints. Despite the efforts made during the last decades, the manual classification process is still predominant. One of the reasons for that is the small number of types usually used in previous studies. In this paper, we present a large study to automatically identify pollen grains using nine state-of-the-art CNN techniques applied to the recently published POLEN73S image dataset. We observe that existing published approaches used original images without study the possible biased recognition due to pollen's background colour or using preprocessing techniques. Our proposal manages to classify up to 97.4% of the samples from the dataset with 73 different types of pollen. This result, which surpasses previous attempts in number and difficulty of pollen types under consideration, is an important step towards fully automatic pollen recognition, even with a large number of pollen grain types.

Keywords: Pollen recognition · Convolutional neural network · Deep learning · Image segmentation

1 Introduction

Food fraud has devastating consequences, particularly in the field of honey production, which the U.S. Pharmacopeia Fraud Database[1] has classified as the third largest area of adulteration, only behind milk and olive oil. Our aim is to find solutions to help solving this problem and prevent its recurrence. The determination of the botanical origin can be used to label honey and the knowledge of the geographic origin is a factor that influences considerably the commercial value of the product and can be used for quality control and to avoid fraud [6].

Although demanding, pollen grain identification and certification are crucial tasks, accounting for a variety of questions like pollination or palaeobotany, but

---

[1] https://decernis.com/solutions/food-fraud-database/.

also for other fields of research, including crime scene investigation [13], allergology studies [7] as well as the botanical and geographical studies concerning origins of honey to prevent honey labelling fraud [15]. However, most of the pollen classification is a time consuming, laborious and a highly skilled work, visually done by human operators using microscopes, trying to identify differences and similarities between pollen grains. These differences are, frequently, imperceptible among pollen grains and may lead to identification errors.

Despite the efforts to develop approaches that allow the automatic identification of pollen grains [11,22], the discrimination of features performed by qualified experts is still predominant [4]. Many industries, including medical, pharmaceutical and honey marketing, depend on the accuracy of this manual classification process, which is reported to be around 67% [19]. A notorious paper [22] from 1996 published a brief summary of the state of the art until then and, more importantly, the demands and needs of palynology to elevate the field to a higher level, thus making it a more powerful and useful tool.

Pattern recognition from images has a long and successful history. In recent years, deep learning, and in particular Convolutional Neural Networks (CNNs), has become the dominant machine learning approach in the computer vision field, specifically, in image classification and recognition tasks. Since the number of annotated pollen images in the publicly available datasets is too small to train a CNN from scratch, transfer learning can be employed. In this paper we propose an automatic pollen recognition approach divided into three steps: initially, the regions which contain pollen are segmented from the background; then, the colour is preprocessed; finally, the pollen is recognized using deep learning.

Most object recognition algorithms focus on recognizing the visual patterns in the foreground region of the images. Some studies indicate that convolutional neural networks (CNN) are biased towards textures [3], whereas another set of studies suggests shape bias for a classification task [2]. However, little attention has been given to analyze how the recognition process is influenced by the background information in the training process.

Considering that there are certain similarities between the layers of a trained artificial network and the recognition task in the human visual cortex, in this study, we hypothesize that if the collected images of a pollen have a unique background colour, different from all the other pollens, it may biases the recognition task, since the recognition could be based only in the background colour. In order to study such influence we trained the CNN with several datasets: one composed with original images, another composed with segmented images (where the background colour was eliminated), and with preprocessed images with histogram equalization and contrast limit adaptive histogram equalization (CLAHE) techniques.

The acquisition of images usually has some different sources resulting in images with different background, as shown in Fig. 1 from the POLEN73S[2] dataset [1] used in this study. Deep learning based pollen recognition methods focus on learning visual features to distinguish different pollen grains. We

---

[2] https://doi.org/10.6084/m9.figshare.12536573.v1.

observed that existing published approaches used the entire image with the original background, in the training process. Background and foreground pixels in each image contribute with the same influence into the learning algorithms. As each pollen type has a different background from other types, when those trained networks are used to classify the pollens they may be biased by capture relevance from pollen's background which may result in biased recognition.
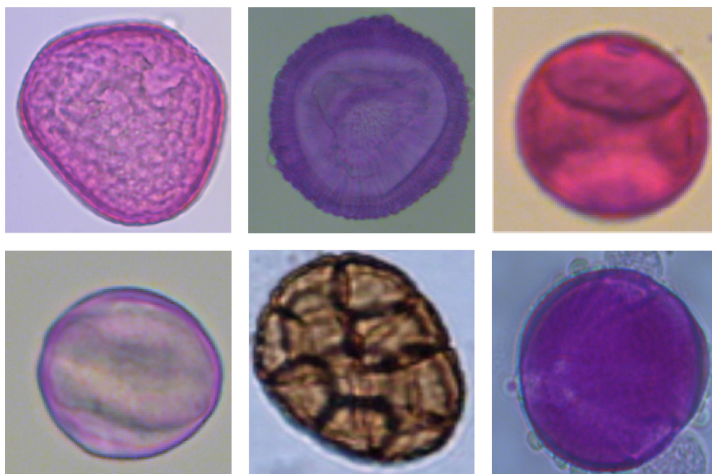

Fig. 1. Pollen dataset samples acquired with different background colours.

In this paper, we investigate the background and colour preprocessing influence by training nine state-of-the-art deep learning convolutional neural networks for pollen recognition. We used a recently published POLEN73S image dataset that includes more than three times as many pollen types and images as the POLEN23E dataset used in recent studies. Our approach manages to classify up to 97.4% of the samples from the dataset with 73 different types of pollen.

The remainder of this paper is organized as follows: Previous related works are presented and reviewed in Sect. 2. In Sect. 3, we describe the used materials and the proposed method. Section 4 presents the results and the discussion of the findings. Finally, some conclusions are drawn in Sect. 5.

2   Related Works

Automatic and semi-automatic systems for pollen recognition based on image features, in particular neural networks and support vector machines, have been proposed for a long time [9,11,17,21]. In general terms, those approaches extract some feature characteristics to identify each pollen type.

Although the classification remains based on a combination of image features, the deep learning CNN approach builds a model determining and extracting the features itself, in alternative of being predefined by human experts. Several

CNN learning techniques have been developed for classifying pollen grain images [1,8,18,19]. In [8], Daood et al. present an approach that learns from the image features and defines the model classifier from a deep learning neural network. This method achieved a 94% classification rate on a dataset of 30 pollen types. Sevillano and Aznarte in [18] and [19] proposed a pollen classification method that applied transfer learning on the POLEN23E dataset and to a 46 different pollen types dataset, achieving accuracies of over 95% and 98%, respectively. In [1], Astolfi et al. presented the POLEN73S dataset and made an extensive study with several CNNs, achieving an accuracy of 95.8%. Despite the importance of their study, we identify two drawbacks in their approach, that influenced the performance: they used different number of samples for each pollen type, and used, for each pollen, an image background that is different from the image background of other pollen types.

3 Experimental Setup

3.1 Pollen Dataset

The automation of pollen grain recognition depends on large image datasets with many samples categorized by palynologists. The results obtained depend on the number of pollen types and the number of samples used. Few samples may result in poor learning models, that are not sufficient to train conveniently the CNN; on the other hand, a small number of pollen types simplifies the identification process making it impractical to be used for recognizing large numbers of pollens usually found in a honey sample.

While a number of earlier datasets have been used for pollen grain classification, such as the POLEN23E[3] dataset [9] or the Pollen Grain Classification Challenge dataset[4], which contain 805 (23 pollinic types) and 11.279 (4 pollinic types) pollen images, respectively, in this paper we use POLLEN73S, which is one of the largest publicly available datasets in terms of pollen types number.

POLLEN73S is an annotated public image dataset, for the Brazilian Savannah pollen types. According to its description in [1] the dataset includes pollen grain images taken with a digital microscope at different angles and manually classified in 73 pollen types, containing 35 sample images for each pollen type, except *gomphrena sp*, *trema micrantha* and *zea mays*, with 10, 34 and 29 samples, respectively. From the results presented in [1], we observed that these small number of samples biased the results. Since CNNs were trained with a smaller number of samples for those types of pollens, this resulted in the worst classification scores relative to the other pollens. To overcome this problem, in our study, several images were generated through rotating and scaling the original images of these pollen types, ensuring the same number of samples for each pollen type, which gives a total of 2555 pollen images. Although the images in the dataset

---

[3] https://academictorrents.com/details/ee51ec7708b35b023caba4230c871ae1fa25 4ab3.

[4] https://iplab.dmi.unict.it/pollenclassificationchallenge/.

have different width and height, they were resized accordingly with the image size of each CNN architecture input.

More datasets were constructed by removing the pollen's background colour (see Fig. 2). Since the images background has medium contrast with the pollen grains, the segmentation process uses just automatic thresholding and morphological operations. We also applied histogram equalization and contrast limit adaptive histogram equalization (CLAHE) to those segmented images. These new datasets allow the independence of training and testing processes from the background colour among the pollen types.
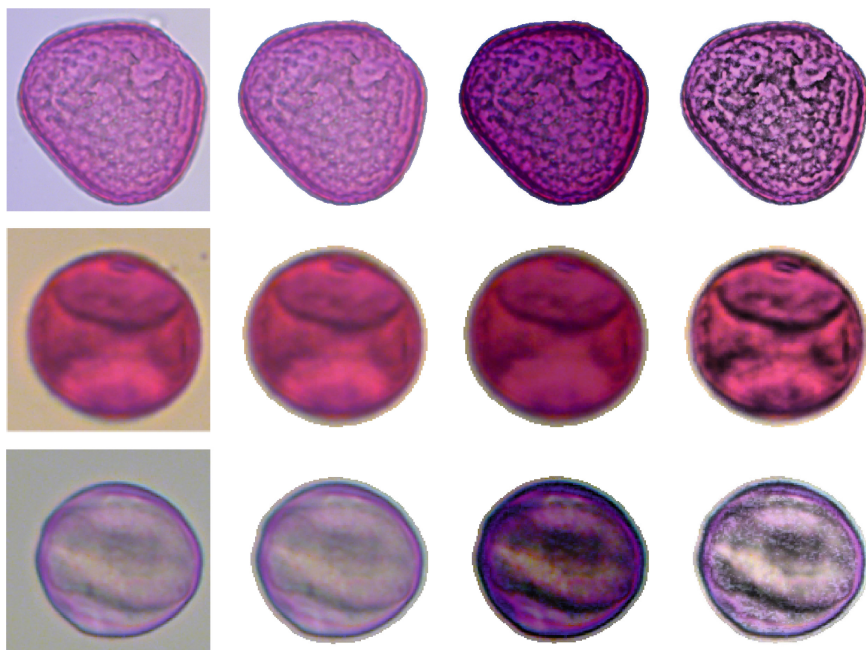

Fig. 2. First column: original images; second column: segmented images with background colour removed; third column: segmented equalized images; fourth column: segmented CLAHE images.

3.2 Convolutional Neural Networks Architectures

CNN is a type of deep learning model for processing images that is inspired by the organization of the human visual cortex and is designed to automatically create and learn feature hierarchies through back-propagation by using multiple layer blocks, such as convolution layers, pooling layers, and fully connected layers from low to high level patterns [2]. This technology is especially suited for image processing, as it makes use of hidden layers to convolve the features with the

input data. The automatic extraction of the most discriminant features from a set of training images, suppressing the need for preliminary feature extraction, became the main strength of CNN approaches.

In this section, we present an overview of the main characteristics of the CNNs used in this study for the recognition of pollen grains types. We choose nine popular CNN architectures due to their performance on previous classification tasks. Table 1 contains a list (chronological sorted) of state-of-the-art CNN architectures, along with a high-level description of how the building blocks can be combined and how the information moves throughout the architecture.

3.3   Transfer Learning

Constraints of practical nature, such as the limited size of training data, degrade the performance of CNNs trained from scratch [18]. Since there is so much work that has already been done on image recognition and classification [10,12,20], in this study we used transfer learning to solve our problem. With transfer learning, instead of starting the learning process from scratch, with a large number of samples, we can use previous patterns that have been learned when solving a similar classification problem.

Transfer learning is a technique whereby a CNN model is first trained on a large image dataset with a similar goal to the problem that is being solved. Several layers from the trained model, usually the lower layers, are then used in a new CNN, trained with sampled images from the current task. This way, the learned features in re-used layers are the starting points for the training process and adapted to classify new types of objects. Transfer learning has the benefit of reducing the training time for a CNN model and can overcome the generalization error due to the small number of images used in the training process when using a network from scratch.

The previous obtained weights, in each layer, may be used as the starting values for the next layers and adapted in response to the new problem. This usage treats transfer learning as a type of weight initialization scheme. This may be useful when the first related problem has a lot more labelled data than the problem of interest and the similarity in the structure of the problem may be useful in both contexts.

3.4   Training Process

In the training process, the CNNs use the fine-tuning strategy, as well as the stochastic gradient descent with momentum optimizer (SGDM) at their default values, dropout rate set at 0.5 and early-stopping to prevent over-fitting, and the learning rate at 0.0001. SGDM is used to accelerate gradients vectors in the correct directions, as we do not need to check all the training examples to have knowledge of the direction of decreasing slope, thus leading to faster converging. Additionally, to consume less memory and train the CNNs faster, we used the CNNs batch size at 12 to update the network weights more often, and trained them in 30 epochs. All images go through a heavy data augmentation

Table 1. Chronological list and descriptions of CNN architectures used in this paper.

VGG 16/19	[20]	Introduced the idea of using smaller filter kernels allowing for deeper networks, and training these networks using pre-training outputs of superficial layers. They have five convolutional blocks where the first two have convolution layers and one max-pooling layer in each block. The remaining three blocks have three fully-connected layers equipped with the rectification non-linearity (ReLU) and the final softmax layer
ResNet 50/101	[10]	Shares some design similarities with the VGG architectures. The batch normalization is used after each convolution layer and before activation. These architectures introduce the residual block that aims to solve the degradation problem observed during network training. In the residual block, the identity mapping is performed, creating the input for the next non-linear layer, from the output of the previous layer
Inception-V3	[24]	This network has three inception modules where the resulting output of each module is the concatenation of the outputs of three convolutional filters with different sizes. The goal of these modules is to capture different visual patterns of different sizes and approximate the optimal sparse structure. Finally, before the final softmax layer, an auxiliary classifier acts as a regularization layer
Inception-ResNet	[23]	Uses the combination of residual connections and the Inception architecture. In Inception networks the gradient is back-propagated to earlier layers, and repeated multiplication may make the gradient indefinitely small, so they replaced filter concatenation stage with residual connections as in ResNet
Xception	[5]	The architecture is composed of three blocks, in a sequence, where convolution, batch normalization, ReLU, and max pooling operations are made. Besides, the residual connections between layers are made as in Resnet architecture
DenseNet201	[12]	Is based on the ideas of ResNet, but built from dense blocks and pooling operations, where each dense block is an iterative concatenation from all previous layers. In the main blocks, the layers are densely connected to each other. Massive reuse of residual information allows for deep supervision as each layer receives more information from the previous layer and therefore the loss function will react accordingly, which makes it a more powerful network
DarkNet53	[16]	It has 53 layers deep and acts as a backbone for the YOLOv3 object detection approach. This network uses successive convolutional layers with some shortcut connections (introduced by ResNet to help the activations propagate through deeper layers without gradient vanishing) to improve the learning ability. Batch Normalization is used to stabilize training, speed up convergence, and regularize the model batch

which includes horizontal and vertical flipping, 360° random rotation, rescaling factor between 70% and 130%, and horizontal and vertical translations between -20 and $+20$ pixels. The CNNs were trained using the Matconvnet package for Matlab$^{\circledR}$ on a node of the CeDRI cluster with two NVIDIA RTX 2080 Ti GPUs.

As in [1], we used 5-Set Cross-Validation, where in each set the images were split on two subsets, 70% (1825 images) for training and 30% (730 images) for testing, allowing the CNN networks to be independently trained and tested on different sets. Since each testing set is build with images not seen by the training model, this allows us to anticipate the CNN behaviour against new images. The four datasets (original, segmented, segmented with equalization and segmented with CLAHE) were trained and tested in an independent way.

4   Results and Discussion

Other works use different evaluation metrics like Precision, Recall, F1-score [1] or correct classification rate (CCR) [18]. However, those metrics use the concept of true negative and false negative. As in this type of experiments we only obtain true positives or false positives, we evaluate the results with Accuracy (Precision gives the same score), which relates true positive with all possible results.

The evaluation results for the nine CNN architectures considered, with different colour pre-processing techniques, are presented in Table 2. The numbers exhibited in bold indicate the best Accuracy result obtained for each network.

Table 2. Classification results (in percentage) on the test set for the different CNNs and preprocessing techniques considered.

CNN	Original	Segmented	Seg. Equal.	Seg. CLAHE
VGG16	94.2	**94.3**	88.7	92.6
VGG19	91.9	**92.4**	92.1	92.1
ResNet50	95.1	**95.2**	93.2	94.8
ResNet101	94.5	94.6	93.0	**95.2**
Inception V3	92.3	92.8	**94.0**	90.8
Inception-ResNet	92.3	89.9	**92.4**	90.3
Xception	**89.7**	87.8	87.5	88.1
DenseNet201	96.7	**97.4**	96.6	95.9
DarkNet53	95.8	**95.9**	95.1	94.9

Based on the results of Table 2, we can conclude that segmenting pollen grains images improves the classification performance for the majority of CNN models allowing the DenseNet201 to achieve an accuracy of 97.4%. Only the Xception network produces better results for the original images. The Inception architectures achieve the best performance with segmented histogram equalized

images. The remain architectures achieved the highest performance when using segmented images without any colour processing.

The DenseNet201 classified correctly all the images for 65 pollen grain types out the 73 types of the dataset. For the other 9 types, it misclassified up to two images, with a total of eleven false positives in the 730 tested images. The lowest accuracy result of the DenseNet201 was achieved with the pollen types *dipteryx alata* and *myrcia guianensis*. These pollen types have predominantly rounded shapes and high texture, that are normally learned in the first CNN layers. Since the transfer learning process changes only a set of the last CNN layers it does not change those learned features during the training process with our images, producing some misclassified results.

The accuracy rates achieved by the DenseNet201 network are relevant due to the amount of pollen types in the POLEN73S dataset, since Sevillano et al. [19] obtained a higher accuracy in a dataset containing only 46 pollen types. That shows that DenseNet201 presented an important performance on POLEN73S.

The network trained and tested using the segmented images produced false positives results that are misclassified as pollens that have high similarity with the tested ones. Figure 3 shows some of those false positive examples.
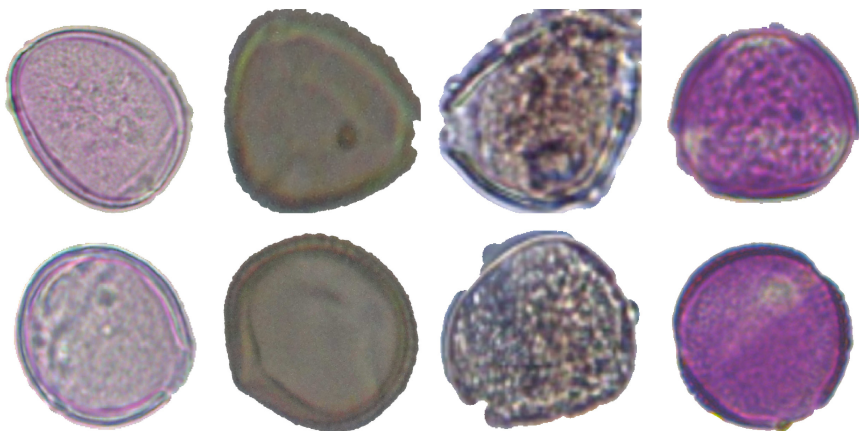

Fig. 3. First row: segmented tested pollens (*magnolia champaca, myrcia guianensis, dipteryx alata, arachis sp*); second row: misclassified pollens (*ricinus communis, schizolobium parahyba, zea mays, myracroduon urundeuva*).

In networks trained and tested with segmented images the background colour bias information was removed, and so the pollen is classified using only the grain pollen information, correcting some of the false positives of the network trained with original images, where the background colour was used as a feature in the classification process.

The high values for the evaluation metric in all CNNs show that the number of correctly identified pollens is high when compared to the number of tested

Table 3. Comparison with previous attempts at pollen classification of more than 20 pollen types using a CNN classifier, with number of types and the highest reported accuracy.

Authors	#Types	Accuracy
Sevillano et al. [18]	23	97.2
Menad et al. [14]	23	95.0
Daood et al. [8]	30	94.0
Sevillano et al. [19]	46	97.8
Astolfi et al. [1]	73	95.8
Our approach	**73**	97.4

images. We believe that an accuracy over 97% is enough to build an automatic classification system of pollen grains, since the visual classification performed by human operators is a hard and time consuming task with a lower performance.

4.1   Comparison with Other Studies

We compared our results with other automatic approaches, from the current literature, that used a CNN classifier. Previous deep learning approaches have shown similar or higher accuracy rates to ours, but these studies were conducted with a small number of pollen types. Table 3 provides a summary table of previous studies, including class sizes and accuracy/success rates against our result. All the literature reviewed, except [1], used a significantly smaller image dataset, in terms of pollen types, than the one used in this paper.

Although the work of Sevillano et al. [19], with forty six types of pollen, achieved a slightly higher performance than our study, as the number of pollen types is directly related to the classification performance of the CNNs, the results must be evaluated taking into account this difference in the number of pollen types between the work presented in [19] and ours.

In short, it can thus be concluded that training a network with the attention focused on the object itself by removing the background dissimilarities can improve the performance of CNN model in pollen classification problem.

5   Conclusion

The usual method for pollen grains identification is a qualitative approach, based on the discrimination of pollen grain characteristics by an human operator. Even though this manual method is quite effective, the all process is time consuming, laborious and sometimes subjective. Creating an automatic approach to identify the grains, in a precise way, thus represents a task of utmost interest.

In this study, an automated pollen grain recognition approach is proposed. We investigate the influence of background colours and colour pre-processing in

the recognition task using nine state-of-the-art CNN topologies. Using a combination of an image-processing workflow and a sufficiently trained deep learning model, we were able to recognize pollen grains from seventy three pollen types, one of the largest number of pollen types studied until now, achieving an accuracy of 97.4% that represents one of the best success rate so far (when weighted for the number of pollen types used in this work).

This study proves that using deep learning CNN architectures for the pollen grain recognition task allows good classification results when using a transfer learning approach. In the future, we plan to combine the features from several CNNs enhancing the effectiveness of deep learning approaches in pollen grain recognition.

References

1. Astolfi, G., et al.: POLLEN73S: an image dataset for pollen grains classification. Ecol. Inform. **60**, 101165 (2020)
2. Baker, N., Lu, H., Erlikhman, G., Kellman, P.J.: Local features and global shape information in object classification by deep convolutional neural networks. Vis. Res. **172**, 46–61 (2020)
3. Bianco, S., Cusano, C., Napoletano, P., Schettini, R.: Improving CNN-based texture classification by color balancing. J. Imaging **3**(33) (2017)
4. Buters, J., et al.: Pollen and spore monitoring in the world. Clin. Transl. Allergy **8**(9) (2018)
5. Chollet, F.: Xception: deep learning with depthwise separable convolutions. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1800–1807 (2017)
6. Corvucci, F., Nobili, L., Melucci, D., Grillenzoni, F.V.: The discrimination of honey origin using melissopalynology and raman spectroscopy techniques coupled with multivariate analysis. Food Chem. **169**, 297–304 (2015)
7. D'Amato, G., et al.: Allergenic pollen and pollen allergy in Europe. Allergy **62**(9), 976–990 (2007)
8. Daood, A., Ribeiro, E., Bush, M., et al.: Pollen grain recognition using deep learning. In: Bebis, G. (ed.) ISVC 2016. LNCS, vol. 10072, pp. 321–330. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-50835-1_30
9. Gonçalves, A.B., et al.: Feature extraction and machine learning for the classification of Brazilian savannah pollen grains. PLoS ONE **11**(6), e0157044 (2016)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
11. Holt, K.A., Bennet, K.D.: Principles and methods for automated palynology. New Phytol. **203**(3), 735–742 (2014)
12. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2261–2269 (2017)
13. Laurence, A.R., Bryant, V.M.: Forensic Palynology, pp. 1741–1754. Springer, New York, NY (2014)
14. Menad, H., Ben-Naoum, F., Amine, A.: Deep convolutional neural network for pollen grains classification. In: JERI (2019)

15. Ponnuchamy, R., et al.: Honey pollen: using melissopalynology to understand for-
    aging preferences of bees in tropical south India. PLoS ONE **9**(7), e101618 (2014)
16. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. ArXiv p.
    1804.02767 (2018)
17. Rodriguez-Damian, M., Cernadas, E., Formella, A., Fernandez-Delgado, M., Pilar
    De Sa-Otero: Automatic detection and classification of grains of pollen based on
    shape and texture. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **36**(4),
    531–542 (2006)
18. Sevillano, V., Aznarte, J.L.: Improving classification of pollen grain images of the
    POLEN23E dataset through three different applications of deep learning convolu-
    tional neural networks. PLoS ONE **13**(9), e0201807 (2018)
19. Sevillano, V., Holt, K., Aznarte, J.L.: Precise automatic classification of 46 differ-
    ent pollen types with convolutional neural networks. PLoS ONE **15**(6), e0229751
    (2020)
20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale
    image recognition. In: International Conference on Learning Representations, pp.
    1–14 (2015)
21. Sobol, M.K., Finkelstein, S.A.: Predictive pollen-based biome modeling using
    machine learning. PLoS ONE **13**(8), e0202214 (2018)
22. Stillman, E., Flenley, J.: The needs and prospects for automation in palynology.
    Quat. Sci. Rev. **15**(1), 1–5 (1996)
23. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.: Inception-v4, inceptionresnet and
    the impact of residual connections on learning. In: Proceedings of the Thirty-First
    AAAI Conference on Artificial Intelligence, pp. 4278–4284 (2017)
24. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the incep-
    tion architecture for computer vision. In: 2016 IEEE Conference on Computer
    Vision and Pattern Recognition (CVPR), pp. 2818–2826 (2016)