

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/349467384>

# Pollen Grain Microscopic Image Classification Using an Ensemble of Fine-Tuned Deep Convolutional Neural Networks

Chapter · February 2021

DOI: 10.1007/978-3-030-68763-2\_26

CITATIONS

7

READS

62

4 authors, including:



**Amirreza Mahbod**

Danube Private University

35 PUBLICATIONS 1,210 CITATIONS

[SEE PROFILE](#)



**Rupert C Ecker**

TissueGnostics GmbH

62 PUBLICATIONS 2,391 CITATIONS

[SEE PROFILE](#)



**Isabella Ellinger**

Medical University of Vienna

106 PUBLICATIONS 2,259 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Deep Learning for Automated Microscopic Image Segmentation, Classification and Detection [View project](#)



Transport of IgG across the human placenta [View project](#)

# Pollen Grain Microscopic Image Classification Using an Ensemble of Fine-Tuned Deep Convolutional Neural Networks <sup>★</sup>

Amirreza Mahbod<sup>1</sup>, Gerald Schaefer<sup>2</sup>, Rupert Ecker<sup>3</sup>, and Isabella Ellinger<sup>1</sup>

<sup>1</sup> Institute for Pathophysiology and Allergy Research, Medical University of Vienna,  
Vienna, Austria

<sup>2</sup> Department of Computer Science, Loughborough University, Loughborough,  
United Kingdom

<sup>3</sup> Research and Development Department, TissueGnostics GmbH, Vienna, Austria

**Abstract.** Pollen grain micrograph classification has multiple applications in medicine and biology. Automatic pollen grain image classification can alleviate the problems of manual categorisation such as subjectivity and time constraints. While a number of computer-based methods have been introduced in the literature to perform this task, classification performance needs to be improved for these methods to be useful in practice.

In this paper, we present an ensemble approach for pollen grain microscopic image classification into four categories: *Corylus Avellana* well-developed pollen grain, *Corylus Avellana* anomalous pollen grain, *Alnus* well-developed pollen grain, and non-pollen (debris) instances. In our approach, we develop a classification strategy that is based on fusion of four state-of-the-art fine-tuned convolutional neural networks, namely EfficientNetB0, EfficientNetB1, EfficientNetB2 and SeResNeXt-50 deep models. These models are trained with images of three fixed sizes ( $224 \times 224$ ,  $240 \times 240$ , and  $260 \times 260$  pixels) and their prediction probability vectors are then fused in an ensemble method to form a final classification vector for a given pollen grain image.

Our proposed method is shown to yield excellent classification performance, obtaining an accuracy of 94.48% and a weighted F1-score of 94.54% on the ICPR 2020 Pollen Grain Classification Challenge training dataset based on five-fold cross-validation. Evaluated on the test set of the challenge, our approach achieved a very competitive performance in comparison to the top ranked approaches with an accuracy and a weighted F1-score of 96.28% and 96.30%, respectively.

**Keywords:** Pollen grain images · microscopic images · deep learning · image classification · transfer learning.

---

<sup>★</sup> This research has received funding from the Austrian Research Promotion Agency (FFG), No. 872636. We thank Nvidia corporation for their generous GPU donation.

## 1 Introduction

Palynology is the scientific study of pollen grains which are produced by plants for the purpose of reproduction. Nowadays, pollen grain classification is a valuable tool for various applied sciences including systematics and forensics [16]. In the medical context, pollens are among the most common triggers of seasonal allergies. Allergic diseases have become a major public health problem. The World Health Organization (WHO) estimates that worldwide 400 million people suffer from allergic rhinitis [7]. In Europe, the prevalence of pollen allergy in the general population is about 40% [9]. Clinicians and patients may be able to anticipate the onset of pollen-related allergy symptoms by monitoring pollen levels [2].

The performance of any pollen counter (human or machine) depends on two key tasks: finding the pollen grains by discriminating pollen from non-pollen, and pollen grain classification. Traditionally, pollen grain classification involves observation and discrimination of features by a highly qualified palynologist. While this is an accurate and effective method, it takes considerable amounts of time and resources [36]. Therefore, automatic recognition of pollen species by means of computer vision is of major importance in modern palynology [18].

Automatic computer-based pollen grain image classification methods have been proposed in several studies with the aim of reducing human interaction in the analysis procedure [18]. Similar to other medical or natural image classification tasks, machine learning-based approaches have shown a better classification performance for pollen grain image classification compared to standard image processing techniques [1,33].

Most of the classical machine learning-based approaches for pollen grain image classification made use of feature engineering to train a classifier such as multi-layer perceptron or support vector machines. Well-known image features that were used for this task include morphological features [40], texture features [12] or hybrid features [4]. However, due to the similar morphological appearance of pollen grains and various artefacts that may present, deriving well-working hand-crafted features is a challenging task.

Using deep learning approaches and more specifically convolutional neural networks (CNNs) [24] allows for a more suitable solution for distinguishing different types of pollen grains compared to classical machine learning approaches. These deep models can be trained end-to-end without the use of hand-crafted features and can be applied directly on raw or pre-processed images [15,33,1,10].

Since the number of labelled pollen grain images in the publicly available datasets is too small to train a CNN from scratch, transfer learning can be employed. Here, pre-trained CNNs can be used as optimised deep feature extractors or they can be fine-tuned to solve the classification task. Standard pre-trained CNNs such as AlexNet [22] or VGGNet [35] have been used for these purposes [1,33,30,34]. However, the performance of such approaches still needs to be improved to be useful for practical applications.

In this paper, we present an ensemble method for multi-class pollen grain microscopic image classification. Based on our earlier work on various medical im-

age classification tasks and recent advances in transfer learning [28,27,25,42], we develop a classification method which exploits four state-of-the-art pre-trained CNNs fine-tuned with pollen grain images at three different image resolutions. We then fuse the prediction vectors from different models in an ensemble strategy to form the final classification vector. Evaluated on the training set and the test set of the ICPR 2020 Pollen Grain Classification Challenge<sup>4</sup>, our proposed method provides excellent classification performance, achieving weighted F1-scores of 94.54% and 96.30% on the challenge training and test set, respectively.

## 2 Materials and Methods

### 2.1 Datasets

While a number of earlier datasets exist for pollen grain image classification, such as the POLEN23E dataset [33] or the dataset from Ranzato *et al.* [32], which contain 805 and 1,429 pollen grain images, respectively, in this paper, we use the ICPR 2020 Pollen Grain Classification Challenge dataset<sup>5</sup>, which is one the biggest publicly available datasets for this task [1]. This dataset contains 11,279 microscopic training image of size  $84 \times 84$  pixels in RGB format. The images belong to one of four classes, namely *Corylus Avellana* well-developed pollen grain (normal pollen), *Corylus Avellana* anomalous pollen grains (anomalous pollen), *Alnus* well-developed pollen grains (*Alnus*), and non-pollen (debris), example of which are shown in Fig. 1. There are 1566 normal pollen images, 773 anomalous pollen images, 8216 *Alnus* images, and 724 debris images. In addition to the raw images, segmentation masks of the pollen grains are also provided by the challenge organisers, but we did not make use of these. A test set of 1,991 images (without segmentation masks) has also been released but labels for these images are kept private by the challenge organisers.

### 2.2 Pre-processing

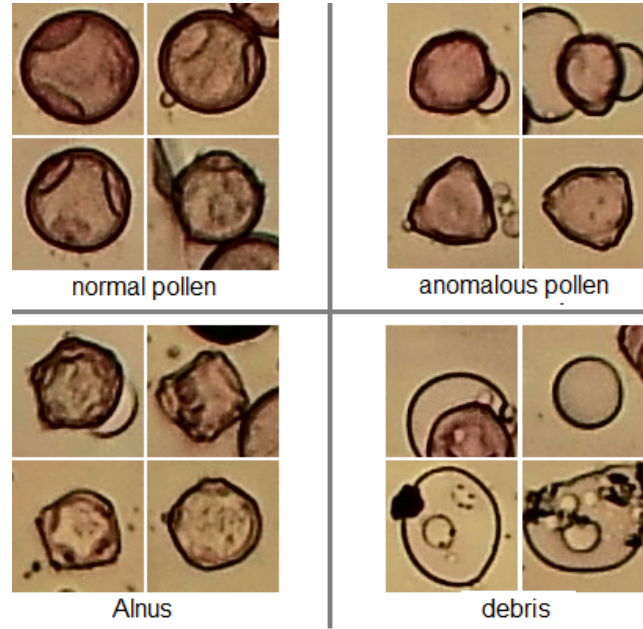
We apply two pre-processing steps. First, we subtract the mean intensity RGB values of the ImageNet dataset [11] from all training and test images. Then we resize all images to a fixed size of  $260 \times 260$  pixels.

### 2.3 Pre-trained CNNs

There are a number of well-established pre-trained CNNs that can be employed for fine-tuning. Recently developed deep neural networks that are widely used for various transfer learning applications include but are not limited to residual networks and their derivatives such as ResNeXt or wide ResNet [17,41], DenseNets

<sup>4</sup> <https://iplab.dmi.unict.it/pollenclassificationchallenge>

<sup>5</sup> <https://iplab.dmi.unict.it/pollenclassificationchallenge/train.zip>

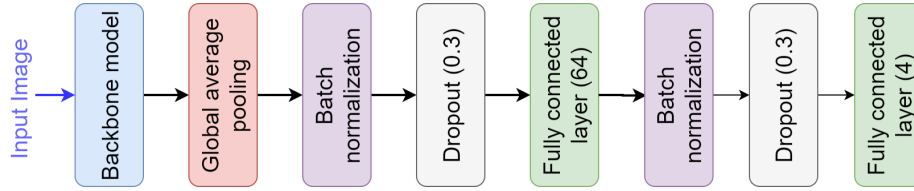


**Fig. 1.** Example images from the ICPR 2020 Pollen Grain Classification Challenge.

with various depths such as Densenet-121 or DenseNet-169 [20], squeeze and excitation networks that can be also combined with other models such as SeResNeXt or SeResNet [19], GoogLeNet and its derivatives such as Inceptions [37,38] and EfficientNet [39] models (the state-of-the-art models for ImageNet classification task). As mentioned above, some earlier pre-trained CNNs such as AlexNet or VGGNet have already been used for pollen grain image classification. However, their classification performance can be improved by employing more advanced architectures which for other tasks have shown superiority over AlexNet or VGGNet.

In our approach, we exploit four recently developed yet already well-established pre-trained CNNs from the EfficientNet [39] and SeResNeXt model [19] families as they have shown excellent performance in, for example, classification of skin lesions or ophthalmological image classification [27,42,29].

The backbone model of the SeResNeXt network is ResNet. ResNet models consist of special building blocks, called residual blocks, that alleviate the vanishing gradient problem by connecting the input and output of each residual block. Thus, the network depth can be increased to yield a better classification performance with deeper models. ResNet-50, ResNet-101, and ResNet-152 are the three most used variations, while ResNeXt is a modified architecture whose residual blocks are wider compared to ResNet through multiple parallel pathways similar to the inception module in the GoogLeNet family. SeResNeXt



**Fig. 2.** Graphic representation of the fine-tuning approach. The backbone model is either EfficientNetB0, EfficientNetB1, EfficientNetB2 or SeResNeXt-50.

incorporates the squeeze and excitation blocks from [19] into the model, while SeResNeXt-50 and SeResNeXt-101 are the best-known models of the family.

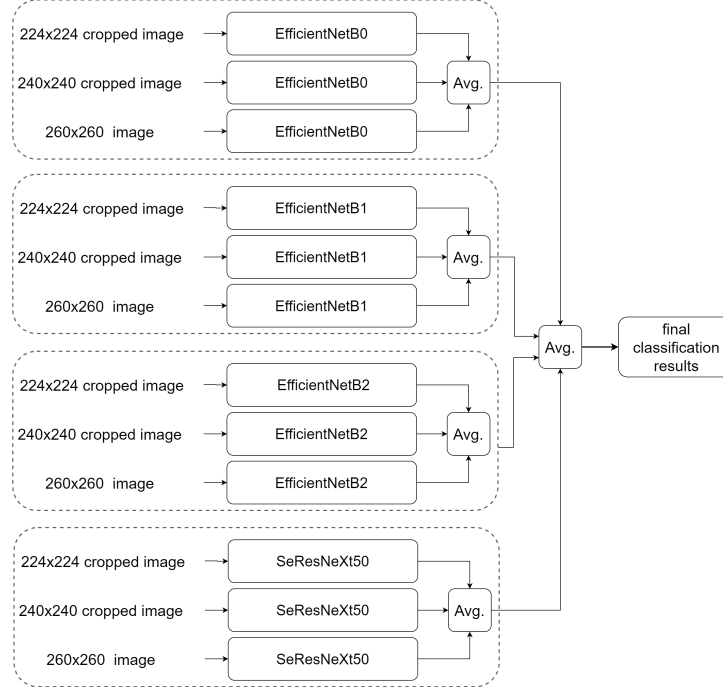
In general, by solving the vanishing gradient issue by some techniques such as skip or dense connections, the performance of a CNN can be improved by increasing the network depth, width, or input image resolution. In the design of many former CNNs such as ResNet or wide ResNet, only one aspect of the model is increased which leads to a better classification performance (e.g. DenseNet-169 has a better classification score in comparison to DenseNet-121 as it is deeper). In contrast, in the EfficientNet architecture, all three aspects of the models are increased systematically and by a constant factor. This leads to generation of eight different architectures, EfficientNetB0 to EfficientNetB7, that have various depths, widths and pre-defined input image sizes (from  $224 \times 224$  to  $600 \times 600$  pixels) while being computationally less expensive compared to other well-known CNNs.

We use three variations of EfficientNet, namely EfficientNetB0, EfficientNetB1, and EfficientNetB2, as well as the SeResNeXt-50 model for fine-tuning. We choose the shallower versions of the EfficientNet family networks to prevent over-fitting to the limited training data available. All four utilised networks are initially trained on natural images with image sizes between  $224 \times 224$  to  $260 \times 260$  pixels

## 2.4 Fine-tuning

We apply a similar approach for fine-tuning the networks as in our work on skin lesion classification [27,28,26]. The general scheme for fine-tuning the pre-trained networks is depicted in Fig. 2.

To fine-tune the pre-trained networks, first fully connected (FC) layers of the original networks are removed. Then, we apply two blocks of batch normalisation, drop out layers (with a drop factor of 0.3) and FC layers (with 64 and 4 nodes in the first block and second block, respectively). We train the models with three different images sizes, namely  $224 \times 224$ ,  $240 \times 240$ , and  $260 \times 260$  pixels. For the two smaller sizes (i.e.  $224 \times 224$  and  $240 \times 240$  pixel images), random cropping within the  $260 \times 260$  pixel images is performed during training and testing. We use a global average pooling layer just before the first batch normalisation layer to prevent dimensionality mismatch for different image sizes.



**Fig. 3.** Graphic representation of the proposed ensemble strategy.

The weights of the newly added FC layers are initialised with the Xavier method [14] and the weight factors are kept 10 times larger compared to all other learnable weights in the models. We fine-tune the network with the Adam optimiser [21] with an initial learning rate of 0.001. We halve the learning rate after every seven epochs and train the networks in total for 40 epochs. To deal with the imbalanced dataset set, we use a weighted focal loss function [23] with weights of 7.20, 14.59, 1.37, and 15.57 for the normal pollen, anomalous pollen, *Alnus* pollen, and debris, respectively.

As in [27], we use various augmentation techniques in the training and testing phases to prevent over-fitting. For training augmentation, we use random cropping (for the models trained with  $224 \times 224$  and  $240 \times 240$  pixels images), adaptive histogram equalisation (with a probability of 0.1), random cut-out (with a probability of 0.1), random brightness and contrast shifts (with a probability of 0.4), random horizontal and vertical flipping (with a probability of 0.5), and random rotations (0, 90, 180 and 270 degrees, with a probability of 0.5).

In the inference phase, we use 25-folds test time augmentation with the same augmentation methods as described for training. We monitor the prediction probability vectors for each augmented image. If the maximum value in a prediction vector is below 0.5, we disregard that prediction vector and take the average over the rest of the prediction vectors for a specific test image.

## 2.5 Fusion

To improve the classification performance, we employ a straightforward ensemble strategy as shown in Fig. 3.

In our fusion scheme, first, the results from the five folds for each image size and for each network are fused. Then, the results from three different image sizes for a specific model architecture are combined and finally the classification probability vectors of the four utilised model architectures (i.e. EfficientNetB0, EfficientNetB0, EfficientNetB0, and SeResNeXt-50) are fused to form the final classification predictions. From the final prediction for each image, we chose the element with maximum probability to determine the image class. Fusion of multiple models is performed by taking the average over their classification prediction vectors.

## 2.6 Evaluation

As evaluation measures, we use accuracy, balanced accuracy [3], which is defined as the average recall over all classes, and the weighted F1 score [5], which is calculated as the F1 score for each class weighted by its number of true samples. The latter is also the main evaluation measure that is used in the ICPR challenge.

## 2.7 Implementation

We use Keras<sup>6</sup> and Tensorflow<sup>7</sup> deep learning frameworks for algorithm development. All experiments are conducted using a single workstation with an Intel Core i7-8700 3.20 GHz CPU, 32 GB of RAM and a Titan V Nvidia GPU card with 12 GB of installed memory.

## 3 Results

Since the ground truth of the test data is kept private by the challenge organisers, in developing our model and for showing the effectiveness of the utilised ensemble approach, we use five-fold cross-validation (5CV) based on the ICPR challenge training data. That is, the training data is split into five folds and four of these are used for training while the fifth is utilised for testing. This is repeated five times so that each test fold is used once and the obtained results are averaged.

Table 1 shows the 5CV results for each network architecture and each input image size.

The results in Table 2 give the performance of the classification models obtained by fusing networks trained at the three different image sizes. The final row in Table 2 shows the final fusion performance, derived from fusion of 12 (3 image sizes  $\times$  4 architectures) sub-models.

<sup>6</sup> <https://keras.io/>

<sup>7</sup> <https://www.tensorflow.org/>



**Table 1.** 5CV results [%] based on different input crop sizes and different network architectures.

network	image size	accuracy	balanced accuracy	weighted F1-score
EfficientNetB0	$224 \times 224$	92.48	92.20	92.66
EfficientNetB0	$240 \times 240$	91.96	92.22	92.18
EfficientNetB0	$260 \times 260$	93.07	92.18	93.20
EfficientNetB1	$224 \times 224$	93.01	92.11	93.16
EfficientNetB1	$240 \times 240$	94.06	92.68	94.13
EfficientNetB1	$260 \times 260$	93.20	92.60	93.34
EfficientNetB2	$224 \times 224$	92.74	92.03	92.91
EfficientNetB2	$240 \times 240$	92.68	92.38	92.84
EfficientNetB2	$260 \times 260$	93.18	92.17	93.31
SeResNeXt-50	$224 \times 224$	92.00	91.84	92.20
SeResNeXt-50	$240 \times 240$	92.40	91.81	92.57
SeResNeXt-50	$260 \times 260$	91.50	91.49	91.74

**Table 2.** 5CV results [%] based on network fusion with different input image sizes and fusion of all architectures.

network	image size	accuracy	balanced accuracy	weighted F1-score
EfficientNetB0	all sizes	93.35	92.75	93.47
EfficientNetB1	all sizes	94.26	92.83	94.34
EfficientNetB2	all sizes	93.61	92.52	93.72
SeResNeXt-50	all sizes	93.45	92.38	93.56
all networks	all sizes	94.48	93.22	94.54

We also compare the obtained results with other state-of-the-art classification models for pollen grain images. In particular, we compare with the results reported in [1] where both classical feature extraction-based methods and deep learning models were evaluated. The investigated techniques include histogram of oriented gradients (HOG) [8] and local binary pattern (LBP) [31] features, and multi-layer perceptron (MLP) [13] and support vector machine (SVM) [6] classifiers, while the explored deep learning models include AlexNet [22] and VGG [35].

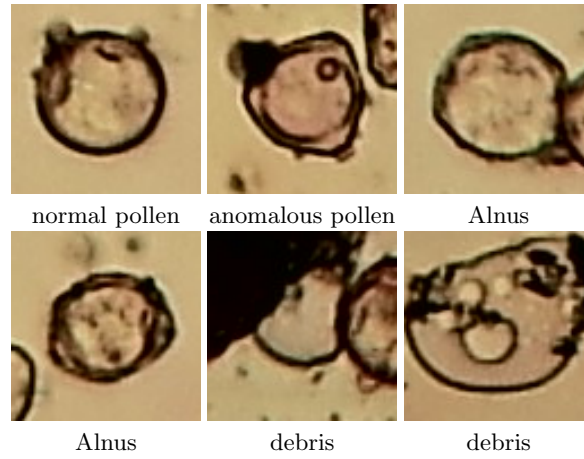
The results are listed in Table 3. It should be noted that the results in [1] are obtained based on a fixed split into 85% training data and 15% test data and that this split is not publicly available. In contrast, we perform 5CV to evaluate our method, and thus use slightly less training data while ensuring that all available data is used once for testing.

Examples of correctly or incorrectly classified images in the training data (by the full fusion approach) are shown in Fig. 4 and Fig. 5, respectively.

As the final experiment, we extend our fusion approach to combine the output of all 5CV results in the ensemble strategy to investigate the performance of the model for the 1,991 test images of the ICPR challenge. Hence, instead of combining 12 sub-models, we fuse the results of all 60 sub-models ( $5CV \times 3$

**Table 3.** Comparison to other state-of-the-art methods from [1] based on the ICPR 2020 Pollen Grain Classification Challenge training data.

approach type	features/model	accuracy	weighted F1-score
classical machine learning	HOG features + RBF SVM	86.58	85.66
classical machine learning	HOG features + MLP	84.93	84.31
classical machine learning	LBP features + MLP	80.02	77.64
deep learning	AlexNet + augmentation	89.63	88.97
deep learning	small VGG + augmentation	89.73	89.14
deep learning	our approach	94.48	94.54

**Fig. 4.** Examples of correctly classified images.**Table 4.** Results on ICPR Challenge test data of our extended fusion approach and the top three performers of the challenge.

User name	accuracy	weighted F1-score
Zhangbaochang	97.53	97.51
Fang Chao	97.34	97.30
Penghui Gui	97.29	97.26
our approach	96.28	96.30

image sizes  $\times$  4 architectures). The results for this (evaluated by the challenge organisers) and comparison to the top three performers of the ICPR challenge for the test data are shown in Table 4.

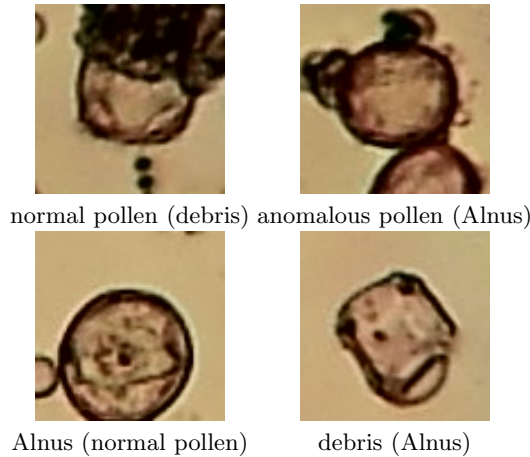
## 4 Discussion

The results in Table 1 show that all networks deliver very good classification performance, while EfficientNetB1 is slightly superior compared to the rest and SeResNeXt-50 is slightly worse.

Looking at Table 2, we can see that fusing the results from different image sizes improves the classification performance for all four models. The performance is further slightly improved when the results from different model architectures are also fused as shown in the bottom row of Table 2.

The comparison between our proposed method and other state-of-the-art techniques in Table 3 demonstrates a clear superiority of our approach over the other classification models. While, as mentioned, the training/test configuration we employ is somewhat different from that of the other methods, it represents arguably a harder test, yet as the results in the table clearly indicate we achieve significantly better results both in terms of accuracy and weighted F1 score. In addition, the techniques reported in [1] utilise the segmentation masks (which are only available for the training dataset of the challenge) to obtain segmented images. In contrast, our approach uses only the raw images as input and is segmentationless.

Finally, the results in Table 4 show the performance of our proposed method on the actual test data of the ICPR challenge. By comparing the 5CV results (last row in Table 3) and the test results (last row of Table 4), performance improvement in both accuracy and weighted F1 score can be observed. Compared to the top performers of the ICPR challenge our approach delivers very competitive performance with slightly inferior results (by 1.25% and 1.21% in terms of accuracy and weighted F1 score, respectively). Further details of the



**Fig. 5.** Examples of incorrectly classified images. The ground truth class is mentioned first and the predicted class in (brackets).

test set results can be found on the challenge website<sup>8</sup>. Note that the details of the other approaches are not disclosed at the time of writing this manuscript.

## 5 Conclusions

In this paper, we have proposed an effective approach to classify pollen grain microscopic images to four pollen grain types. We exploit several fine-tuned CNNs at various image sizes and combined them in a simple yet effective fusion framework. Our approach does not require segmentation and we have demonstrated excellent classification results on the challenge’s training and test data sets. The performance can be potentially further improved by incorporating other network architectures or other image resolutions.

## References

1. Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., Siniscalco, C.: Detection and classification of pollen grain microscope images. In: Conference on Computer Vision and Pattern Recognition Workshops (2020)
2. Biedermann, T., Winther, L., Till, S.J., Panzner, P., Knulst, A., Valovirta, E.: Birch pollen allergy in Europe. *Allergy* **74**(7), 1237–1248 (2019). <https://doi.org/10.1111/all.13758>
3. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: International Conference on Pattern Recognition. pp. 3121–3124 (2010). <https://doi.org/10.1109/ICPR.2010.764>
4. Chica, M.: Authentication of bee pollen grains in bright-field microscopy by combining one-class classification techniques and image processing. *Microscopy Research and Technique* **75**(11), 1475–1485 (2012). <https://doi.org/10.1002/jemt.22091>
5. Chinchor, N.A., Sundheim, B.: Message understanding conference (MUC) tests of discourse processing. In: Spring Symposium on Empirical Methods in Discourse Interpretation and Generation. pp. 21–26 (1995)
6. Cortes, C., Vapnik, V.: Support-vector networks. *Machine learning* **20**(3), 273–297. <https://doi.org/https://doi.org/10.1007/BF00994018>
7. Cruz, A.A.: Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach. World Health Organization (2007)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Society Conference on Computer Vision and Pattern Recognition. vol. 1 (2005)
9. D’Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., Liccardi, G., Popov, T., Van Cauwenberge, P.: Allergenic pollen and pollen allergy in Europe. *Allergy* **62**(9), 976–990 (2007). <https://doi.org/10.1111/j.1398-9995.2007.01393.x>
10. Daoud, A., Ribeiro, E., Bush, M.: Pollen grain recognition using deep learning. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., Isenberg, T. (eds.) *Advances in Visual Computing*. pp. 321–330. Springer International Publishing, Cham (2016)

<sup>8</sup> <https://iplab.dmi.unict.it/pollenclassificationchallenge/results>

11. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A large-scale hierarchical image database. In: Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009). <https://doi.org/10.1109/CVPR.2009.5206848>
12. Fernandez-Delgado, M., Carrion, P., Cernadas, E., Galvez, J., Sa-Otero, P.: Improved classification of pollen texture images using SVM and MLP. In: International Conference on Visualization, Imaging and Image Processing. vol. 2 (2003)
13. Gardner, M.W., Dorling, S.: Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment* **32**(14-15), 2627–2636 (1998)
14. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: International Conference on Artificial Intelligence and Statistics. pp. 249–256 (2010)
15. Goncalves, A.B., Souza, J.S., Silva, G.G.d., Cereda, M.P., Pott, A., Naka, M.H., Pistori, H.: Feature extraction and machine learning for the classification of Brazilian Savannah pollen grains. *PloS One* **11**(6), e0157044 (2016). <https://doi.org/https://doi.org/10.1371/journal.pone.0157044>
16. Halbritter, H., Ulrich, S., Grímsson, F., Weber, M., Zetter, R., Hesse, M., Buchner, R., Svojtka, M., Frosch-Radivo, A.: Palynology: History and Systematic Aspects, pp. 3–21. Springer International Publishing, Cham (2018)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
18. Holt, K.A., Bennett, K.D.: Principles and methods for automated palynology. *New Phytologist* **203**(3), 735–742 (2014). <https://doi.org/10.1111/nph.12848>
19. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Conference on Computer Vision and Pattern Recognition. pp. 7132–7141 (2018)
20. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 4700–4708 (2017)
21. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
22. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. pp. 1097–1105 (2012)
23. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: International Conference on Computer Vision. pp. 2980–2988 (2017)
24. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. *Medical Image Analysis* **42**, 60–88 (2017)
25. Mahbod, A., Ellinger, I., Ecker, R., Smedby, Ö., Wang, C.: Breast cancer histological image classification using fine-tuned deep network fusion. In: Image Analysis and Recognition. pp. 754–762. Springer International Publishing (2018)
26. Mahbod, A., Schaefer, G., Ellinger, I., Ecker, R., Pitiot, A., Wang, C.: Fusing fine-tuned deep features for skin lesion classification. *Computerized Medical Imaging and Graphics* **71**, 19–29 (2019). <https://doi.org/https://doi.org/10.1016/j.compmedimag.2018.10.007>
27. Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., Ellinger, I.: Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. *Computer Methods and Programs in Biomedicine* p. 105475 (2020). <https://doi.org/https://doi.org/10.1016/j.cmpb.2020.105475>

28. Mahbod, A., Schaefer, G., Wang, C., Ecker, R., Dorffner, G., Ellinger, I.: Investigating and exploiting image resolution for transfer learning-based skin lesion classification. arXiv preprint arXiv:2006.14715 (2020)
29. Mahbod, A., Tschandl, P., Langs, G., Ecker, R., Ellinger, I.: The effects of skin lesion segmentation on the performance of dermatoscopic image classification. *Computer Methods and Programs in Biomedicine* **197**, 105725 (2020). <https://doi.org/https://doi.org/10.1016/j.cmpb.2020.105725>
30. Menad, H., Ben-Naoum, F., Amine, A.: Deep convolutional neural network for pollen grains classification. In: 3rd Edition of the National Study Day on Research on Computer Sciences. CEUR Workshop Proceedings, vol. 2351 (2019)
31. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7) (2002)
32. Ranzato, M., Taylor, P., House, J., Flagan, R., LeCun, Y., Perona, P.: Automatic recognition of biological particles in microscopic images. *Pattern Recognition Letters* **28**(1), 31–39 (2007)
33. Sevillano, V., Aznarte, J.L.: Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. *PloS One* **13**(9), e0201807 (2018). <https://doi.org/https://doi.org/10.1371/journal.pone.0201807>
34. Sevillano, V., Holt, K., Aznarte, J.L.: Precise automatic classification of 46 different pollen types with convolutional neural networks. *PLoS One* **15**(6), e0229751 (2020). <https://doi.org/https://doi.org/10.1371/journal.pone.0229751>
35. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
36. Stillman, E., Flenley, J.R.: The needs and prospects for automation in palynology. *Quaternary Science Reviews* **15**(1), 1–5 (1996)
37. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: *Conference on Computer Vision and Pattern Recognition*. pp. 1–9 (2015)
38. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Conference on Computer Vision and Pattern Recognition*. pp. 2818–2826 (2016)
39. Tan, M., Le, Q.V.: EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
40. Travieso, C.M., Briceño, J.C., Ticay-Rivas, J.R., Alonso, J.B.: Pollen classification based on contour features. In: *International Conference on Intelligent Engineering Systems* (2011). <https://doi.org/10.1109/INES.2011.5954712>
41. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: *Conference on Computer Vision and Pattern Recognition*. pp. 5987–5995 (2017)
42. Zhang, Z.: Deep-learning-based early detection of diabetic retinopathy on fundus photography using efficientnet. In: *International Conference on Innovation in Artificial Intelligence*. pp. 70–74 (2020). <https://doi.org/10.1145/3390557.3394303>