Automatic pollen image identification using neural networks increases accuracy for hay fever forecasts

Marcel Polling[1], Chen Li[2], Lu Cao[2], Fons Verbeek[2], Letty de Weger[3], Joost Willemse[2], Hugo de Boer[4], and Barbara Gravendeel[1]

[1]Naturalis Biodiversity Center
[2]University of Leiden
[3]Leiden University Medical Center
[4]University of Oslo Faculty of Mathematics and Natural Sciences

September 11, 2020

To the Editor,

Pollen allergies are on the rise globally. Therefore, for more and more people, fast and accurate monitoring of airborne pollen provides an essential early warning system. Although the allergenic pollen from some plants can be monitored at the species level (e.g. ragweed, *Ambrosia artemisiifolia* L.[1]), many species cannot be accurately identified to this level based on their pollen. In many taxa, only a genus- or family-level identification is possible using current microscopic methods, even if they possess very different allergenic profiles. An extra challenging factor in airborne pollen identification is that they are collected directly from the air[2]. Because they are not acetolyzed[3], and thus still contain all organic material, defining features are less apparent in these pollen grains[4].

The identification challenge is exemplified in the case of the nettle family (Urticaceae). Pollen grains produced by all species from the genus *Urtica* L. (stinging nettles) are allergenically irrelevant, while pollen from several species of *Parietaria* L. (pellitory) are a major cause of hay fever, in particular *P. judaica* L. and *P. officinalis* L.[5] Both these species of pellitory are native to the Mediterranean, but throughout the second half of the twentieth century, a range expansion occurred through north-western Europe, the Americas and Australia as a result of anthropogenic distribution and climate change[6]. Their pollen is virtually indistinguishable from that of native *Urtica* nettles and their contribution to the total aerial pollen load is currently not assessed.

Automatic image recognition can be used to improve identification of pollen taxa that are difficult to distinguish. Subtle variations in morphology that are not readily apparent through microscopic investigation may be consistently detected by neural networks. Here we use the Deep Convolutional Neural Network (DCNN) model VGG16 to distinguish morphologically similar, unacetolyzed pollen from the nettle family (see online supporting information for details on VGG16).

Pollen from all five species of the nettle family present in the Netherlands was collected from plants in multiple locations (both fresh and herbarium material) to account for potential intraspecific variability. This includes pollen from three species of *Urtica* and two exotic *Parietaria* species (Supplementary Table 1). To train the VGG16 model, a minimum of 1000 pollen images per species were taken at 100x magnification

1

using a Widefield Zeiss observer. We wrote a custom script in ImageJ[7] to automatically extract individual pollen grains from raw multifocal images. Three different image projections were produced to visualize the pollen in two rather than three dimensions, and to emphasize distinguishable features (Figure S3, script available on https://github.com/pollingmarcel/Pollen_Projector). The VGG16 model was pre-trained on the ImageNet dataset and contains 13 convolutional, and three fully connected layers. To mitigate potential overfitting and to increase the size and variability of the training data set, we used 10-fold cross-validation and data augmentation (Figure S4). We used 90% of the pollen images to train the model, and 10% were used for subsequent testing (see online Supplementary Information for a complete description of materials and methods).

The trained VGG16 model accurately identified pollen to the genus level in 95.9% of the time for *Urtica* and 97.8% for *Parietaria*(Figure 1). The species *Urtica membranacea* was confidently distinguished from all other Urticaceae species (99.2%), but distinction at the species-level was not possible for the other *Urtica* and *Parietaria* species. This is because the distinguishing differences between pollen from these species could not be resolved in the image projections used.

From the raw pollen images, we further identified clear intra-specific differences in the pollen grains that result from natural variability within each species. To test whether VGG16 learned the correct distinguishing features rather than sample-specific details, we produced feature maps (Figure 2). Despite the highly variable input images of unacetolyzed pollen, the model learned distinct features such as edges in the first convolutional layers, while finer features, such as pores and annuli, were learned in deeper layers.

This is the first time a DCNN model has been used to increase the accuracy of unacetolyzed pollen. The model represents a significant improvement of earlier attempts in distinguishing Urticaceae using automatic image recognition. In a previous study using shape and texture features, pollen from three Urticaceae species could be distinguished from another (89% accuracy[8]), though only a small image dataset was used. DCNN models have shown similar accuracy rates to ours on larger and more varied pollen datasets as well, but these either focussed on the family level[9] or on insect-collected pollen for honey analysis[10]. Moreover, all of these studies used acetolyzed pollen that allow for easier recognition of distinguishing features, and used pollen collected from only a single location.

In conclusion, using a combination of an image-processing workflow and a sufficiently trained DCNN model, we were able to differentiate unacetolyzed pollen grains from two genera in the nettle family. These are genera that are indistinguishable with current microscopic methods but possess different allergenic profiles, and thus the ability to differentiate them is of medical significance. We expect this method can be more broadly applied to distinguish pollen from similarly challenging plant families and will aid in producing more accurate hay fever forecasts.

KEYWORDS

automatic image detection – deep convolutional neural networks – hay fever – nettle family – pellitory – pollen – Urticaceae – VGG16

CONFLICT OF INTEREST

The authors declare no conflict of interest

Marcel Polling[1,2] and Chen Li[3]Lu Cao[3]Fons Verbeek[3]Letty de Weger[4]Joost Willemse[5]Hugo de Boer[1,2]Barbara Gravendeel[1,6]

[1]*Naturalis Biodiversity Center, Leiden, The Netherlands*[2]*Natural History Museum, University of Oslo, Norway*[3]*Leiden Institute of Advanced Computer Science, Leiden, The Netherlands*[4]*Department of Pulmonology, Leiden University Medical Center, Leiden, The Netherlands*[5]*Microbial Sciences, Institute of Biology, Leiden, The Netherlands*[6]*IWWR, Radboud University, Nijmegen, The Netherlands*

Correspondence Marcel Polling, Naturalis Biodiversity Center, Leiden, The Netherlands Email: marcel.polling@naturalis.nl

ORCID

Fons Verbeek https://orcid.org/0000-0003-2445-8158 Letty de Weger https://orcid.org/0000-0002-9067-7407 Hugo de Boer https://orcid.org/0000-0003-1985-7859 Barbara Gravendeel https://orcid.org/0000-0002-6508-0895

REFERENCES

1. Sikoparija B, Galán C, Smith M. Pollen-monitoring: between analyst proficiency testing. *Aerobiologia.* 2017;33(2):191-199.

2. Hirst JM. An automatic volumetric spore trap. *Annals of applied Biology.* 1952;39(2):257-265.

3. Erdtman G. The acetolysis method-a revised description. *Sven Bot Tidskr.* 1960;54:516-564.

4. Jones GD. Pollen analyses for pollination research, acetolysis.*Journal of pollination Ecology.* 2014;13(21):203-217.

5. Ciprandi G, Puccinelli P, Incorvaia C, Masieri S. *Parietaria*Allergy: An Intriguing Challenge for the Allergist. *Medicina.*2018;54(6):106.

6. Bass D, Bass D. *Parietaria judaica* L. A cause of allergic disease in Sydney. A study of habit and spread of the weed. *Review of Palaeobotany and Palynology.* 1990;64(1-4):97-101.

7. Rasband WS. ImageJ. US National Institutes of Health, Bethesda, MD, U.S.A. http://rsb.info.nih.gov/ij/. In:1997-2006.

8. Rodriguez-Damian M, Cernadas E, Formella A, Fernández-Delgado M, De Sa-Otero P. Automatic detection and classification of grains of pollen based on shape and texture. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews).* 2006;36(4):531-542.

9. Daood A, Ribeiro E, Bush M. Classifying pollen using robust sequence alignment of sparse Z-stack volumes. Paper presented at: International Symposium on Visual Computing2016.

10. Sevillano V, Holt K, Aznarte JL. Precise automatic classification of 46 different pollen types with convolutional neural networks.*bioRxiv.* 2020.

SUPPORTING INFORMATION

Additional supporting information can be found in the online Supplementary Information file

Figure 1. Results
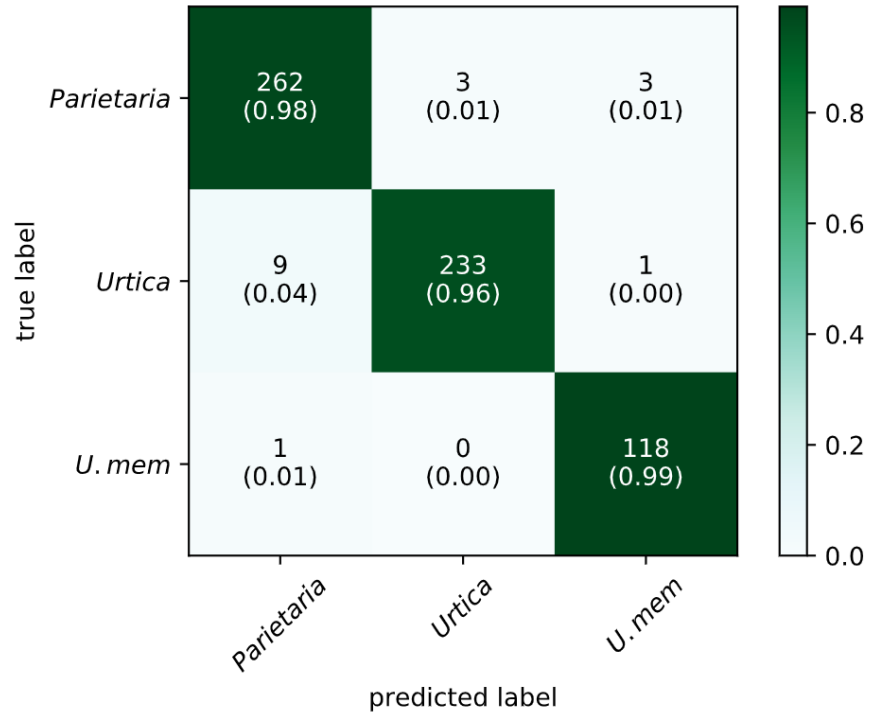
	Training	Test
CCR	99.0%	97.3%
Precision	0.990	0.973
Recall	0.990	0.973
F1 score	0.990	0.973

Figure 2 Feature Maps