Automatic Image Classification Using Neural Networks Increases Accuracy for Allergenic Pollen Monitoring

Marcel Polling (marcel.polling@naturalis.nl)
Naturalis Biodiversity Center, Leiden

Chen Li
Leiden Institute of Advanced Computer Science (LIACS), Leiden

Lu Cao
Leiden Institute of Advanced Computer Science (LIACS), Leiden

Fons Verbeek
Leiden Institute of Advanced Computer Science (LIACS), Leiden

Letty de Weger
Department of Pulmonology, Leiden University Medical Center, Leiden

Jordina Belmonte
Institute of Environmental Sciences and Technology (ICTA-UAB), Universitat Autònoma de Barcelona, Bellaterra, Cerdanyola del Vallès

Concepción De Linares
Institute of Environmental Sciences and Technology (ICTA-UAB), Universitat Autònoma de Barcelona, Bellaterra, Cerdanyola del Vallès

Joost Willemse
Microbial Sciences, Institute of Biology, Leiden

Hugo de Boer
Natural History Museum, University of Oslo

Barbara Gravendeel
Naturalis Biodiversity Center, Leiden

Research Article

Keywords: CNN, Network, information, Urticaceae

Posted Date: December 9th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-116766/v1

Abstract

Monitoring of airborne pollen concentrations provides an important source of information for the globally increasing number of hay fever patients. Airborne pollen are traditionally counted under the microscope, but with the latest developments in image recognition methods, automating this process has become feasible. A challenge that persists, however, is that many pollen grains cannot be distinguished beyond the genus or family level using a microscope. Here, we assess the use of a Convolutional Neural Network (CNN) to increase taxonomic accuracy for airborne pollen. As a case study we use the nettle family (Urticaceae), which contains two main genera (*Urtica* and *Parietaria*) common in European landscapes which pollen cannot be separated by trained specialists. While pollen from *Urtica* species have very low allergenic relevance, those from several species of *Parietaria* are severely allergenic. We collect pollen from both fresh as well as from herbarium specimens and use these to train the CNN model VGG16. The model shows that Urticaceae pollen can be distinguished with 98.3% accuracy. We then apply our model on Urticaceae pollen collected from aerobiological samples and show that the genera can be confidently distinguished, despite the more challenging input images that are often overlain by debris. Our method can also be applied to other pollen families in the future and will thus help to make allergenic pollen monitoring more specific.

Introduction

Pollen allergies are on the rise globally, with worldwide approximately 10 to 30% of adults and 40% of children affected[1,2]. For patients the symptoms include a runny nose, sneezing and itchy eyes, mouth or skin. Control measures and medication are readily available, but to alleviate the symptoms most efficiently, exposure to allergens should be kept to a minimum[3]. Therefore, for more and more people, fast and accurate monitoring of airborne pollen provides an essential early warning system[4,5]. Pollen concentrations in the air are monitored using samplers that collect airborne pollen on sticky tape, e.g. Hirst type samplers[6]. These tapes are microscopically inspected for their pollen content, a process that requires highly trained specialists. Moreover, although the allergenic pollen from some plants can be monitored at the species level (e.g. species of plantain, *Plantago* L.[7]), many other pollen grains cannot be accurately identified to this level. In many taxa, only a genus- or family-level identification is possible using current microscopic methods[8]. This is problematic since different species and even genera within the same family can possess very different allergenic profiles. An extra challenging factor in airborne pollen identification from Hirst samples is that they are collected directly from the air. In contrast to acetolyzed pollen[9], these still contain all organic material, and defining features are less apparent[10].

This 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) have a low allergenic profile[11], while pollen from several species of *Parietaria* L. (pellitory) are a major cause of hay fever and asthma, in particular *P. judaica* L. and *P. officinalis* L.[12,13]. These pellitory species are native to the Mediterranean, but throughout the second half of the twentieth century, a range expansion occurred through north-eastern

Europe, the Americas and Australia as a result of anthropogenic distribution and climate change[14,15]. *Parietaria* sensitization is highly different per geographic area, but has been reported to reach 80% in southern Italy while a value of 13% was found in the United Kingdom[16]. Species of *Parietaria* flower throughout the year but their main flowering peaks occur in May-June and August-October, which overlaps with the flowering season of *Urtica* species (June – October)[17]. Cross-reactivity is present between species of *Parietaria*, but is absent between the genera *Urtica* and *Parietaria*[11,18,19]. *Parietaria* pollen is microscopically indistinguishable from that of *Urtica* and their contribution to the total airborne pollen load is currently not assessed in either native or expanded range[20].

Pollen grains from *Urtica* and *Parietaria* species have a simple morphology: they are small (~11-20µm), rounded to slightly ellipsoidal tri-, tetra- or zonoporate with a psilate to scabrate surface ornament and small pores. Most species have an annulus around the pore, i.e. a thickening of the otherwise very thin exine and a germination area called the oncus (lens-shaped body located in the apertural region)[7]. The only species of Urticaceae that can be distinguished in aerobiological samples is *Urtica membranacea* due to its small size (~10-12µm) and high number of pores (usually more than six[21]. The main difference between the pollen of *Urtica* and *Parietaria* are the slightly smaller size and coarser surface ornamentation of *Parietaria*, and a more angular outline and more pronounced annulus of *Urtica*[22].

Neural networks have been used successfully to manage both the tasks of differentiating pollen from non-pollen debris as well as correctly identifying different taxa (for an overview please refer to [23]). Automatic image recognition can however also be used to improve identification of pollen taxa that are difficult to distinguish using traditional methods. Subtle variations in morphology that are not readily apparent through microscopic investigation may be consistently detected by neural networks. This has for example been shown for the highly similar pollen of black spruce (*Picea mariana* (Mill.) Britton, Sterns & Poggenb.) and white spruce (*Picea glauca* (Moench) Voss) using machine learning [24] and for pollen of 10 species of the thistle genus *Onopordum* L. using an artificial neural network[25]. However, neural networks have so far not been tested for improvement of taxonomic resolution in unacetolyzed pollen in aerobiological samples.

Here we use the Convolutional Neural Network (CNN) model VGG16 to distinguish morphologically similar, unacetolyzed pollen from the nettle family. We collect pollen from all species of Urticaceae present in the Netherlands (*Urtica dioica, U. membranacea, U. urens, Parietaria judaica* and *P. officinalis*), freshly collected as well as from herbaria, and create a pollen image reference dataset. We first optimize the performance of the CNN model using data augmentation and 10-fold cross validation. The model is then applied to Urticaceae pollen from three aerobiological samples with high Urticaceae pollen count. We use one sample from the Leiden University Medical Centre (LUMC), Leiden, the Netherlands as well as one sample each from Lleida and Vielha, Catalonia, Spain (ICTA-UAB). In the Netherlands, stinging nettles (*Urtica*) are highly abundant and therefore it is expected that most Urticaceae pollen will be from this genus. *Urtica* is also expected to be dominant in Vielha, while in the direct surroundings of Lleida, *Parietaria* is very abundant.

The main objectives of this study are (1) to see whether a CNN model can distinguish morphologically similar unacetolyzed pollen of two common genera in the Urticaceae family that have highly differing allergenic profiles; (2) to test whether the trained model can be successfully applied on aerobiological samples containing more complex and for the model before unseen input images.

Results

Model Performance

The trained VGG16 model accurately identified pollen to the genus level for 97.5% of the test images for *Urtica* and 98.5% for *Parietaria* (Figure 1A). For *Parietaria* four images were misclassified (all to *Urtica*), while seven were misclassified for *Urtica* (all to *Parietaria*). The species *Urtica membranacea* was confidently distinguished from all other Urticaceae species (100%), but distinction at the species-level was not possible for any of the other *Urtica* and *Parietaria* species. This is because the distinguishing features of pollen from these species (e.g. exine ornamentation) could not be resolved in the used image projections. The average accuracy for the model using the two genera (*Urtica* and *Parietaria*), and the species *Urtica membranacea* is 98.3%. We find that using a hard voting approach in our model improves the accuracy (Correct Classification Rate, CCR) by 0.58% compared with calculating the average CCR of 10 models (Figure 1B; see Materials and Methods for full explanation). Including data augmentation into the model further improved the accuracy of the model by 0.95% on the test data set.

For all species, pollen were collected from a minimum of four different plants. Looking at the raw pollen images from the different plants, we identified intra-specific differences that result from natural variability within each species. To test whether VGG16 learned the pollen-specific distinguishing features rather than sample-specific details, we produced feature maps (Figure 2). Despite the highly variable input images of unacetolyzed pollen from different plants, the model consistently learned features such as edges in the first convolutional layers, while finer features such as pores and annuli were learned in deeper layers.

Application to Test Cases

Table 1 shows the results of the CNN on Urticaceae pollen from an aerobiological sample from Leiden, the Netherlands, as well as from Lleida and Vielha, Catalonia, Spain. We set the identification threshold at a value of 60% as derived from the model test images, and therefore the CNN also returned unknown images (see Supplementary Table S2 for the full results). For the sample from Leiden, 85.7% of the Urticaceae pollen were identified as *Urtica*, with only minor presence of *Parietaria* (4.5%). The sample from Lleida shows dominance of *Parietaria* pollen grains (81.0%) while 14,3% of the Urticaceae pollen grains were classified as *Urtica*. Finally, for Vielha we find a mixture of ~70% *Urtica* and ~20% *Parietaria.* No *Urtica membranacea* pollen grains were identified in any of the samples. On average, unknown images account for 8.7% of the total images when using 60% identity threshold. When using a stricter

identity threshold (e.g. 70%, see Table 1), the unknown image category increases to an average value of 13.5%.

Discussion

This study demonstrates incorporating neural networks to increase the taxonomic resolution of pollen grains in aerobiological samples. The feature maps in Figure 2 show that the trained deep learning model VGG16 looks at the traditionally used morphological features to distinguish *Urtica* from *Parietaria* pollen. The characteristic thickening of the exine around the pores of *Urtica* shows the highest activation in the deeper convolutional layers. The distinct thickening is missing in *Parietaria* pollen, and the model instead focuses on the pollen outline. As expected, the only species to be distinguished by our model is *Urtica membranacea* which shows a slightly angular outline due to the larger numbers of pores (Figure 2D). For the other species used in this study, no distinction was possible even though it has been shown that pollen from species of *Urtica* (*U. dioica* and *U.* urens) and *Parietaria* (*P.judaica* and *P. officinalis*) can be separated based on differences in their exine ornamentation[22]. These differences can, however, only be imaged using specialized microscopy methods such as SEM or phase-contrast imaging, and are very hard to visualize using brightfield microscopy. Furthermore, these features are obscured when pollen grains are not acetolyzed. For our purposes, this species level distinction is not relevant as no known differences in allergenicity are known between either the species of *Urtica*[11] or *Parietaria*[18].

This is the first time a deep learning model has been used to increase the taxonomic accuracy of unacetolyzed pollen. The model represents a significant improvement of earlier attempts in distinguishing Urticaceae pollen using automatic image classification. In a previous study using shape and texture features, pollen from three Urticaceae species could be distinguished from another with a 89% accuracy[26], though only a small image dataset was used to train the model (i.e. 100 images per species). Similar results were obtained by [27] where shape features were used with a minimum distance classifier to obtain a 86% accuracy between three species of Urticaceae. Because not all species of Urticaceae were included and a low amount of training images was used, these studies have limited applicability to the highly diverse pollen encountered in aerobiological slides.

Deep learning 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[28] or on insect-collected pollen for honey analysis[29,30]. Increasing the taxonomic resolution of pollen grains has been achieved by incorporating an extensively trained deep learning model with super-resolution microscopy on a case study of fossil pollen[31]. Similarly, incorporating SEM images has been found to allow for highly accurate distinction of pollen types[32]. These microscopy methods, however, are often much more expensive than using Light Microscopy and require extensive sample preparation. Moreover, nearly all of these studies work with acetolyzed pollen that allow easier recognition of distinguishing features, and used pollen collected from a single location.

To validate our model, we tested it on Urticaceae pollen from aerobiological samples collected from different locations in Spain and the Netherlands. Most of the pollen grains from the sample from Leiden, the Netherlands were identified by the deep learning model as *Urtica*, with only a low number of images identified as *Parietaria*. While *Parietaria* plants are relatively abundant around the sampling location in Leiden and were flowering on the chosen date, its pollen is most likely simply outnumbered by the much larger number of nettles in the area. For Lleida (Catalonia), where pellitory plants are abundantly present, *Parietaria* pollen grains dominated the assemblage, while the sample from Vielha showed a mixed assemblage. The number of unknown images was the highest for the sample from Vielha (11.5%), which is most likely the result of the presence of more debris on the pollen grains making a certain identification impossible. In all aerobiological slides, debris on top of or below the pollen grains was observed in different focal plains. Nevertheless, the model still successfully classified most of the pollen grains, and in most cases with high confidence (Supplementary Table S2). This shows the potential broad application of this method and opens up opportunities to study both seasonal as well as long-term yearly dynamics of *Parietaria* versus *Urtica* abundance of airborne pollen, as well as using this method to distinguish other morphologically similar species of allergenic importance from different families (e.g. Betulaceae, Amaranthaceae, Oleaceae).

A limitation of our method is that currently pollen from aerobiological slides have to be located manually. In other systems like the commercially available Classifynder system, pollen are automatically located and imaged using darkfield imaging after which a simple neural network classifies the pollen[33]. This is also the case for the BAA500 system used by e.g. Oteros et al.[34], that was particularly developed for recognizing and classifying unacetolyzed airborne pollen for hay fever predictions. While both systems achieve automated and accelerated pollen counting, our method instead particularly increases the accuracy of information useful for allergy prevention by making it more specific.

Conclusions

In conclusion, using a combination of an image-processing workflow and a sufficiently trained deep learning model, we were able to differentiate unacetolyzed pollen grains from two genera and one species 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. Our method can be more broadly applied to distinguish pollen from similarly challenging allergenic plant families and can thus help in producing more accurate pollen spectra to improve the forecasts for allergy sufferers.

Material And Methods

Collection of Pollen

Pollen were collected from all five species of Urticaceae found in the Netherlands. In the genus *Urtica*, the native species *U. dioica* L. (common nettle) and *U. urens* L. (small nettle) are ubiquitous in nitrogen rich

moist areas, ditches, woodlands, disturbed sites and roadsides. The exotic Mediterranean species *U. membranacea* is rarely encountered, though is included in this study since its range is expected to increase due to the effects of global warming. The genus *Parietaria* is represented in the Netherlands by the species *P. judaica* L. (pellitory of the wall) and *P. officinalis* L. (upright pellitory) that both occupy rocky substrates, mainly in the urban environment[15]. Moreover, *P. judaica* has shown a big increase in abundance over the past decades, e.g. in the Netherlands (Supplementary Figure S1), but also in many other parts of the world.

Pollen from all Urticaceae species was either freshly obtained or collected from herbarium specimens (Naturalis Biodiversity Center; see Supplementary Table S1). Fresh material was collected with the help of an experienced botanist in the direct surroundings of Leiden and The Hague during the nettle flowering seasons of 2018 and 2019. Original taxonomic assignments for the herbarium specimens were verified using identification keys and descriptions[35]. A minimum of four different plants were sampled per species, from different geographical locations to cover as much of the phenotypic plasticity in the pollen grains as possible and reflect the diversity found on aerobiological slides.

To produce palynological reference slides, thecae of open flowers were carefully opened on a microscopic slide using tweezers. A stereo microscope was mounted in a fume hood to avoid inhalation of the severely allergenic pollen of *Parietaria* species. Non-pollen material was manually removed to obtain a clean slide. The pollen were mounted using a glycerin:water:gelatin (7:6:1) solution with 2% phenol and stained with Safranin (0.002% w/v). These represent the same conditions also used in airborne pollen collected by a Hirst type sampler. Cover slips were secured with paraffin.

Pollen Image Capture

A total of 6,472 individual pollen grains were scanned, using a minimum of 1000 individual pollen grains for each species, though numbers varied (Supplementary Table S1). The system used for the imaging was a Zeiss Observer Z1 (inverted microscope) linked to a Hamamatsu EM-CCD Digital Camera (C9100), located at the Institute of Biology Leiden (IBL). Grayscale images were used, since the pollen were stained to increase contrast and not for species recognition.

The imaging procedure was as follows: on each microscope reference slide containing only pollen of one species of Urticaceae, an area rich in pollen was identified by eye and this area was automatically scanned using multidimensional acquisition with the Zeiss software Zen BLUE. For areas that were very rich in pollen, a user-defined mosaic was created consisting of all the tiles to be scanned (e.g. 20x20 tiles), while a list of XY positions was used for microscopic slides less rich in pollen. Because pollen grains are 3-D shapes, catching all important features can only be achieved using different focal levels, so-called ‘Z-stacks’. A total of 20 Z-stacks were used in this study with a step size of 1.8 µm. The settings used for scanning were a Plan Apochromat 100x (oil) objective and numerical aperture 0.55 with a

brightfield contrast manager. To maintain similar conditions in the image collection process, the condenser was always set to 3.3V with an exposure time of 28ms.

Reference Pollen Image Library

All images were post-processed in ImageJ (Fiji)[36] using the script Pollen_Projector (https://github.com/pollingmarcel/Pollen_Projector). The input for this script is a folder containing all raw pollen images (including all Z-stacks), and the output is a set of projections for each individual pollen grain that are subsequently used as input for the VGG16 deep learning model.

Pollen_Projector identifies all complete, non-overlapping pollen grains and extracts them as stacks from the raw Z-stack. This is achieved using binarization on the raw images to detect only those rounded objects with a circularity >0.3 and a size larger than 5µm. Out-of-focus images within each group of 20 Z-stack slices were removed using a threshold for minimum and maximum pixel values. The conventional input of a convolutional neural network is a three-channel image. In colour images RGB channels are commonly used, but since we use grayscale images, three different Z-stack projections were chosen to represent the three different channels. The projections used are Standard Deviation, Minimum Intensity and Extended Focus. Standard Deviation creates an image containing the standard deviation of the pixel intensities through the stack, where positions with large differences appear brighter in the final projection. Minimum intensity takes the minimum pixel value through the stack and uses that for the projection. Finally, the Extended Focus projection was created using the 'Extended_Depth_of_Field' ImageJ macro of Richard Wheeler (www.richardwheeler.net)[37]. This macro takes a stack of images with a range of focal depths and builds a 2D image from it using only in focus regions of the images. A schematic overview of the processes behind the Pollen_Projector script is shown in Supplementary Figure S2.

Convolutional Neural Network

Convolutional Neural Networks (CNN) are widely used in the field of computer vision for image classification, object detection, facial recognition, autonomous driving, etc. For our work we used the pre-trained VGG16 network[38] in Keras[39]. Compared with traditional neural networks and shallow convolutional neural networks, VGG16 has deeper layers that extract more representative features from images. A feature extractor and classifier are two key structural parts of the CNN that perform the classification task. In order to train a CNN model, a large number of tagged data sets are fed into a model, to train the model to learn more features to be able to distinguish the images.

The VGG16 network contains 13 convolutional layers that form five blocks, which generate features from images in the feature extraction phase. During the training process, parameters of convolutional layers were derived from the pre-trained network on the ImageNet dataset. Subsequently, three fully connected layers were built and added to the convolutional layers to classify the different classes (Supplementary Figure S3). To improve the effectiveness, robustness and generalization ability of the VGG16 model, as

well as to prevent overfitting, 10-fold cross-validation was applied in the training process. The complete pollen image dataset was split into a training data set (90%) and a test data set (10%). The training data set was split into ten subsets and used for 10-fold cross-validation. For each fold, the number of epochs was set to 30. The accuracy of the model converged at this point and the model is therefore found not to be overfitting. For each fold, a model was trained using nine of the folds as training data and validated on the remaining part of the data. After the 10-fold cross-validation, 10 models were obtained and tested on the test dataset. In order to try to improve the accuracy of the model further, we further compared hard voting (or majority voting) to the standard average measure for the ten models. Hard voting is summing the votes for class labels from each model and predicting the class with the most votes.

In order to quantify model accuracy, several commonly used performance measures were used:

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$F1\ score = 2 * \frac{precision * recall}{precision + recall}$$

$$CCR = \frac{TP + TN}{TP + TN + FP + FN}$$

where *TP* refers to true positives, *TN* to true negatives, *FP* to false positives and *FN* to false negatives. Recall is the number of true positives divided by the total number of elements that belong to the correct class, which is the sum of the true positives and false negatives. The F1-score is the weighted average of the precision and recall. The correct classification rate (CCR) reflects the accuracy of the model.

Data augmentation

A large number of images for each class is required to train a deep learning model, as the performance will increase when more variation is fed to the model. Due to the nature of the images investigated in this study, the model was sensitive to small changes, since the differences between the genera are very subtle. Therefore, data augmentation was tested to try to increase the variety of pollen images used as input. We selected the augmentation options brightness and flip. Brightness range was set from 0.1 to 2, with <1 corresponding to a darker image and > 1 to a brighter image. Horizontal- and vertical flip were also applied randomly (Supplementary Figure S4). The results were compared to running the model without data augmentation.

Test Cases

For each aerobiological sample an area representing 10% of the total deposition area was scanned manually for Urticaceae pollen grains (i.e. 8 full transects at 100X magnification) resulting in 112 pollen grains from the sample from Leiden (LUMC), 63 from Lleida and 26 from Vielha (both ICTA-UAB). One aspect of the Catalonian aerobiological samples was the presence of pollen from families that produce similar pollen to Urticaceae, that are rarely encountered in the Netherlands. These included *Humulus lupulus* L. (Cannabaceae) and *Morus* sp. (Moraceae) which were not included in our training dataset. These can be distinguished from Urticaceae, however, in the case of *H. lupulus* by their much larger size (up to 35µm) and the very large onci and, in the case of *Morus* by the more ellipsoidal shape. These pollen grains were removed from the dataset before they were fed to the CNN for classification.

Declarations

Data Availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files).

Author Contributions Statement

MP: Conceptualization, Methodology, Visualization, Formal analysis, Writing – original draft **CL:** Methodology, Software, Formal analysis, Data Curation, Investigation **LC:** Resources, Formal analysis, Software, Supervision **FV:** Validation, Supervision, Software **LdW:** Resources, Validation, Writing – Review & Editing **JB:** Resources, Validation, Writing – Review & Editing **CDL:** Resources, Validation **JW:** Software, Methodology **HdB**: Funding acquisition, Writing - Review & Editing **BG:** Conceptualization, Supervision, Project administration, Funding acquisition

Additional Information

Conflict of Interest

The authors declare no conflict of interest

Funding Information

This work was financially supported by the European Union's Horizon 2020 research and innovation programme under H2020 MSCA-ITN-ETN grant agreement No 765000 Plant.ID.

References

1. Bieber, T. *et al.* Global Allergy Forum and 3rd Davos Declaration 2015: Atopic dermatitis/Eczema: challenges and opportunities toward precision medicine. *Allergy***71**, 588-592 (2016).
2. Reitsma, S., Subramaniam, S., Fokkens, W. W. & Wang, D. Y. Recent developments and highlights in rhinitis and allergen immunotherapy. *Allergy***73**, 2306-2313 (2018).
3. Johansson, S. & Haahtela, T. World Allergy Organization guidelines for prevention of allergy and allergic asthma. *International archives of allergy and immunology***135**, 83-92 (2004).
4. Karatzas, K. D., Riga, M. & Smith, M. in *Allergenic pollen* 217-247 (Springer, 2013).
5. Geller-Bernstein, C. & Portnoy, J. M. The clinical utility of pollen counts. *Clinical reviews in allergy & immunology***57**, 340-349 (2019).
6. Hirst, J. M. An automatic volumetric spore trap. *Annals of applied Biology***39**, 257-265 (1952).
7. Beug, H.-J. r. *Leitfaden der Pollenbestimmung für Mitteleuropa und angrenzende Gebiete*. (Verlag Dr. Friedrich Pfeil, 2004).
8. Ziello, C. *et al.* Changes to airborne pollen counts across Europe. *PLoS One***7**, e34076, doi:10.1371/journal.pone.0034076 (2012).
9. Erdtman, G. The acetolysis method-a revised description. *Sven Bot Tidskr***54**, 516-564 (1960).
10. Jones, G. D. Pollen analyses for pollination research, acetolysis. *Journal of pollination Ecology***13**, 203-217 (2014).
11. Tiotiu, A. *et al.* Urtica dioica pollen allergy: clinical, biological, and allergomics analysis. *Annals of Allergy, Asthma & Immunology***117**, 527-534 (2016).
12. D'Amato, G. & Liccardi, G. Pollen-related allergy in the European Mediterranean area. *Clinical & Experimental Allergy***24**, 210-219, doi:10.1111/j.1365-2222.1994.tb00222.x (1994).
13. Ciprandi, G., Puccinelli, P., Incorvaia, C. & Masieri, S. *Parietaria* Allergy: An Intriguing Challenge for the Allergist. *Medicina***54**, 106 (2018).
14. 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***64**, 97-101 (1990).
15. Fotiou, C., Damialis, A., Krigas, N., Halley, J. M. & Vokou, D. Parietaria judaica flowering phenology, pollen production, viability and atmospheric circulation, and expansive ability in the urban environment: impacts of environmental factors. *Int J Biometeorol***55**, 35-50, doi:10.1007/s00484-010-0307-3 (2011).
16. D'Amato, G., Ruffilli, A., Sacerdoti, G. & Bonini, S. Parietaria pollinosis: a review. *Allergy***47**, 443-449, doi:10.1111/j.1398-9995.1992.tb00661.x (1992).
17. Guardia, R. & Belmonte, J. Phenology and pollen production of Parietaria judaica L. in Catalonia (NE Spain). *Grana***43**, 57-64, doi:10.1080/00173130410019244 (2010).
18. Corbi, A. L., Pelaez, A., Errigo, E. & Carreira, J. Cross-Reactivity between *Parietariajudaica* and *Parietaria officinalis*. *Annals of Allergy***54**, 142-147 (1985).

19. Bousquet, J., Hewitt, B., Guerin, B., Dhivert, H. & Michel, F. B. Allergy in the Mediterranean area II: cross-allergenicity among Urticaceae pollens (*Parietaria* and *Urtica*). *Clinical & Experimental Allergy***16**, 57-64 (1986).
20. D'Amato, G. *et al.* Allergenic pollen and pollen allergy in Europe. *Allergy***62**, 976-990, doi:10.1111/j.1398-9995.2007.01393.x (2007).
21. Rodríguez, A. M., Palacios, I. S., Molina, R. T. & Corchero, A. M. Urtica membranacea and the importance of its separation from the rest of the Urticaceae in aeropalynological studies carried out in the Mediterranean region. *Plant Biosystems***140**, 321-332 (2006).
22. Punt, W. & Malotaux, M. Cannabaceae, moraceae and urticaceae. *Review of Palaeobotany and Palynology***42**, 23-44 (1984).
23. Holt, K. A. & Bennett, K. D. Principles and methods for automated palynology. *New Phytol***203**, 735-742, doi:10.1111/nph.12848 (2014).
24. Punyasena, S. W., Tcheng, D. K., Wesseln, C. & Mueller, P. G. Classifying black and white spruce pollen using layered machine learning. *New Phytol***196**, 937-944, doi:10.1111/j.1469-8137.2012.04291.x (2012).
25. Kaya, Y., Mesut Pınar, S., Emre Erez, M., Fidan, M. & Riding, J. B. Identification ofOnopordumpollen using the extreme learning machine, a type of artificial neural network. *Palynology***38**, 129-137, doi:10.1080/09500340.2013.868173 (2014).
26. 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)***36**, 531-542 (2006).
27. De Sá-otero, M. P., González, A., Rodríguez-Damián, M. & Cernadas, E. Computer-aided identification of allergenic species of Urticaceae pollen. *Grana***43**, 224-230, doi:10.1080/00173130410000749 (2004).
28. Daood, A., Ribeiro, E. & Bush, M. in *International Symposium on Visual Computing.* 331-340 (Springer).
29. Sevillano, V., Holt, K. & Aznarte, J. L. Precise automatic classification of 46 different pollen types with convolutional neural networks. *bioRxiv* (2020).
30. Menad, H., Ben-Naoum, F. & Amine, A. in *JERI.*
31. Romero, I. C. *et al.* Improving the taxonomy of fossil pollen using convolutional neural networks and superresolution microscopy. *Proceedings of the National Academy of Sciences***117**, 28496-28505 (2020).
32. Daood, A., Ribeiro, E. & Bush, M. in *International Symposium on Visual Computing.* 321-330 (Springer).
33. Holt, K., Allen, G., Hodgson, R., Marsland, S. & Flenley, J. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany and Palynology***167**, 175-183 (2011).

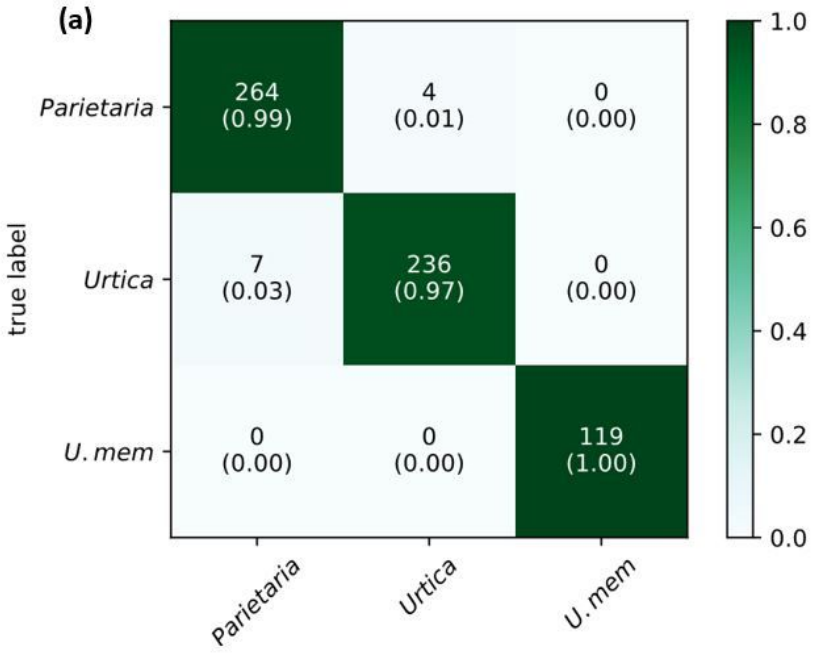
34. Oteros, J. *et al.* Automatic and Online Pollen Monitoring. *Int Arch Allergy Immunol***167**, 158-166, doi:10.1159/000436968 (2015).
35. Duistermaat, L. Heukels' flora van Nederland. (2020).
36. Rasband, W. S. ImageJ. *US National Institutes of Health, Bethesda, MD, U.S.A. http://rsb.info.nih.gov/ij/* (1997-2006).
37. Wheeler, R. *Extended Depth of Field*, <http://www.richardwheeler.net> (
38. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
39. Chollet, F. Keras. *https://github.com/fchollet/keras* (2015).

Tables

Table 1 Results of the deep learning model on Hirst-type aerobiological samples from Leiden (the Netherlands), Lleida and Vielha (both Catalonia, Spain). The threshold for identification was tested at 60% and 70%. Image that were classified below this level were classified as unknown.

Sample Location	Date collected	No. Pollen	% *Urtica*	% *Parietaria*	% *U. mem.*	% Unknown	Identity Threshold
Leiden, NL	23/08/2019	112	85.7	4.5	0	9.8	
Lleida, SP	16/06/2019	63	14.3	81.0	0	4.8	**60%**
Vielha, SP	09/08/2019	26	69.2	19.2	0	11.5	
Leiden, NL	23/08/2019	112	83.0	3.6	0	13.4	
Lleida, SP	16/06/2019	63	12.7	79.4	0	7.9	**70%**
Vielha, SP	09/08/2019	26	69.2	11.5	0	19.2	

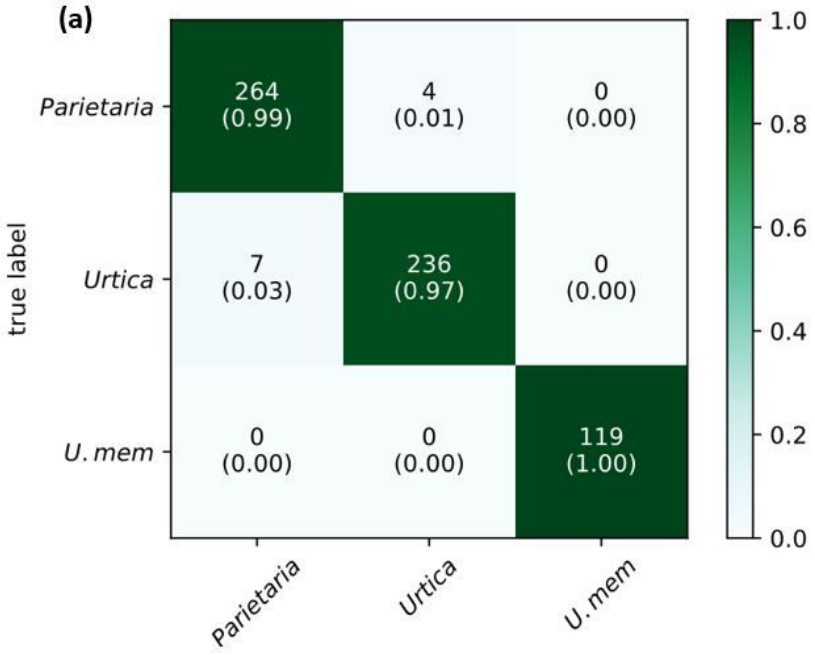
Figures



(b)	Data Augmentation	No data Augmentation
Hard voting CCR	98.3%	97.3%
Average CCR	97.7%	96.5%
Precision	0.983	0.973
Recall	0.983	0.973
F1 score	0.983	0.973

Figure 1

Model performance. (a) Confusion matrix of results using 90% of the images for training and 10% for testing. Numbers represent the actual number of correctly recognized images while those between brackets represent the ratio of correctly classified images (b) Results compared with a hard voting and an average CCR approach as well as with and without data augmentation. Abbreviations used: CCR = correct classification rate (accuracy), U.mem = Urtica membranacea.



(b)	Data Augmentation	No data Augmentation
Hard voting CCR	98.3%	97.3%
Average CCR	97.7%	96.5%
Precision	0.983	0.973
Recall	0.983	0.973
F1 score	0.983	0.973

Figure 1

Model performance. (a) Confusion matrix of results using 90% of the images for training and 10% for testing. Numbers represent the actual number of correctly recognized images while those between brackets represent the ratio of correctly classified images (b) Results compared with a hard voting and an average CCR approach as well as with and without data augmentation. Abbreviations used: CCR = correct classification rate (accuracy), U.mem = Urtica membranacea.

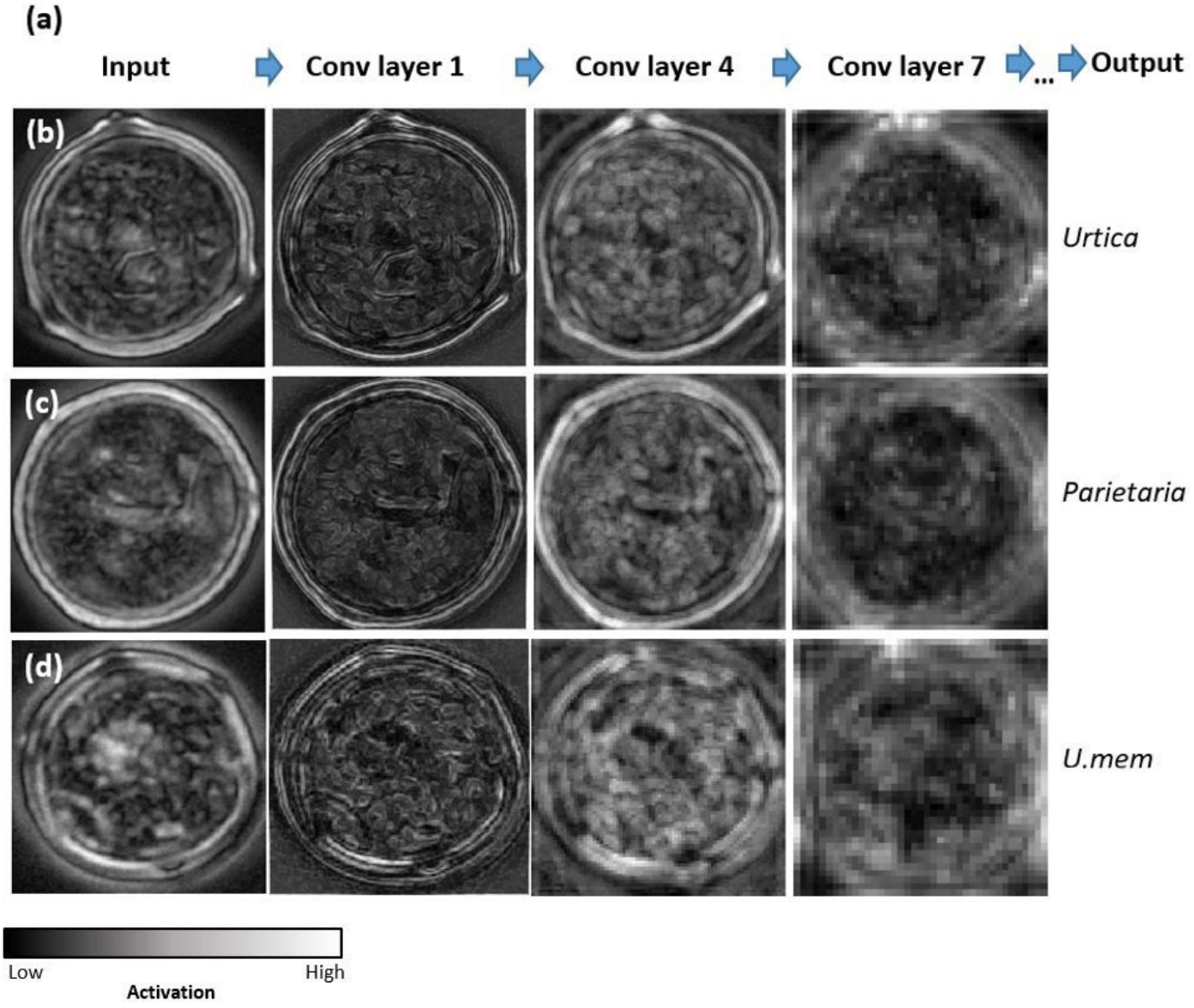


Figure 2

Feature maps. (a) simplified view of the VGG16 model showing three convolutional layers. (b-d) Feature maps of Urticaceae pollen grains from the standard deviation projection that were confidently distinguished by VGG16. (b) Urtica urens, (c) Parietaria judaica and (d) Urtica membranacea. Activation levels are indicated by white indicating high activation and black very low/no activation. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country,

territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

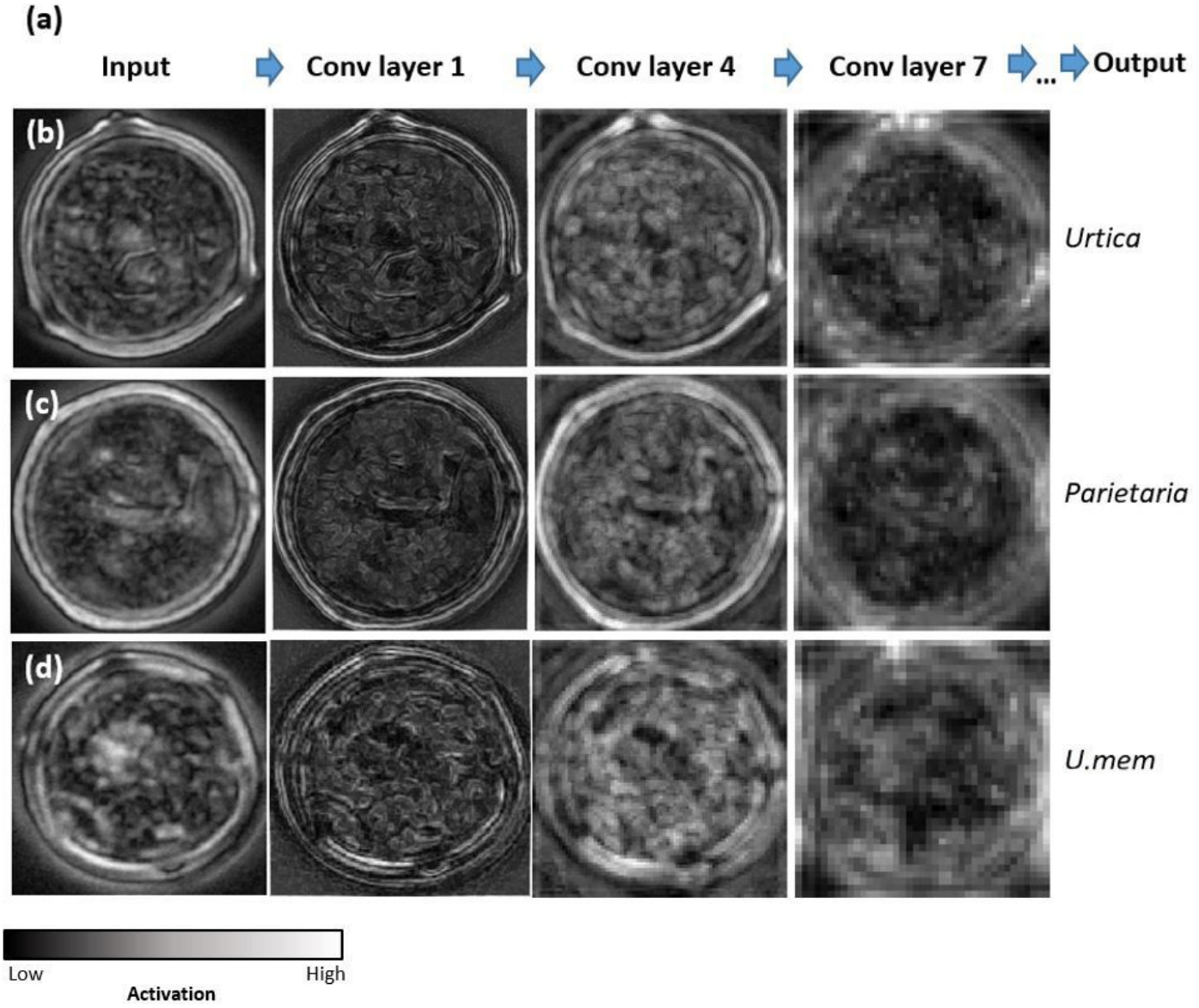


Figure 2

Feature maps. (a) simplified view of the VGG16 model showing three convolutional layers. (b-d) Feature maps of Urticaceae pollen grains from the standard deviation projection that were confidently distinguished by VGG16. (b) Urtica urens, (c) Parietaria judaica and (d) Urtica membranacea. Activation levels are indicated by white indicating high activation and black very low/no activation. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PollingandLiSupplementary.docx
- PollingandLiSupplementary.docx