Automated classification of airborne pollen using neural networks

Julian Schiele, Fabian Rabe, Maximilian Schmitt, Manuel Glaser, Franziska Haring, Jens O. Brunner, Bernhard Bauer, Björn Schuller, Claudia Traidl-Hoffmann, Athanasios Damialis

Angaben zur Veröffentlichung / Publication details:

Schiele, Julian, Fabian Rabe, Maximilian Schmitt, Manuel Glaser, Franziska Haring, Jens O. Brunner, Bernhard Bauer, Björn Schuller, Claudia Traidl-Hoffmann, and Athanasios Damialis. 2019. "Automated classification of airborne pollen using neural networks." In *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 23-27 July 2019, Berlin, Germany*, edited by Thomas Penzel, Thomas Lenarz, Mohamad Sawan, and Riccardo Barbieri, 4474–78. New York, NY: IEEE. https://doi.org/10.1109/embc.2019.8856910.

Automated Classification of Airborne Pollen using Neural Networks

Julian Schiele[1], Fabian Rabe[2], Maximilian Schmitt[3], Manuel Glaser[1], Franziska Häring[4], Jens O. Brunner[1], Bernhard Bauer[2], Björn Schuller[3,5], Claudia Traidl-Hoffmann[4], and Athanasios Damialis[4]

***Abstract*— Pollen allergies are considered as a global epidemic nowadays, as they influence more than a quarter of the worldwide population, with this percentage expected to rapidly increase because of ongoing climate change. To date, alerts on high-risk allergenic pollen exposure have been provided only via forecasting models and conventional monitoring methods that are laborious. The aim of this study is to develop and evaluate our own pollen classification model based on deep neural networks. Airborne allergenic pollen have been monitored in Augsburg, Bavaria, Germany, since 2015, using a novel automatic Bio-Aerosol Analyzer (BAA 500, Hund GmbH). The automatic classification system is compared and evaluated against our own, newly developed algorithm. Our model achieves an unweighted average precision of 83.0 % and an unweighted average recall of 77.1 % across 15 classes of pollen taxa. Automatic, real-time information on concentrations of airborne allergenic pollen will significantly contribute to the implementation of timely, personalized management of allergies in the future. It is already clear that new methods and sophisticated models have to be developed so as to successfully switch to novel operational pollen monitoring techniques serving the above need.**

I. INTRODUCTION

A progressive global increase in the burden of allergic diseases across the industrialized world has been reported over the last half century: clinical evidence reveals a general increase in both the incidence and the prevalence of respiratory diseases, including allergic rhinitis and asthma [1]. This may be related to a parallel increase in the amount of airborne allergenic pollen [2]. In developed countries, respiratory allergies can affect more than 20 % of the population; this percentage varies among cities, countries, and continents because of environmental and other factors, sometimes exceeding 40 % [3]. Moreover, from a socio-economic perspective, an average of 13.2 million emergency visits to health practitioners are recorded every year and approximately 4,000 death incidents related to asthma [4]. It is already known that the annual cost for visits to allergy specialists and for drug therapy reaches up to 3.5 billion dollars in the USA [5] for allergic rhinitis alone, not including the most severe form of asthma. If the latter is included, together with indirect effects, e.g., lost working hours, reduced working efficiency etc., the potentially additional cost rises to 16.1 billion dollars per year [4].

Up to date, the first line of allergy management is prophylaxis. Nonetheless, it has been recently revealed that pollen exposure cannot be completely avoided and pollen allergic symptoms be completely 'switched off', as pollen can be found everywhere, even on high elevation sites, i.e., the Alps [6], or near the atmospheric boundary layer, at approximately 2,000 m above ground level [7]. Hence, to diminish exposure to airborne pollen, reliable, accurate, real-time information on the occurrence and abundance of airborne pollen is needed [8]. Towards this aim, airborne pollen has been routinely monitored worldwide, with the biomonitoring gold-standard device being the volumetric Hirst-type sampler [9]. The device involves a fully manual operation, from collection to chemical analysis and microscopic classification of all different types of pollen [10]. Consequently, relevant risk alerts are announced to the public usually with a delay of at least 7 to 10 days. The management of allergies can, however, be effective only if allergen avoidance and prophylaxis are performed in a timely and accurate manner.

As the above need has been evident mainly within the last decade, a new generation of automated, near-real-time pollen measurements is currently being developed and has been able in some cases to work on an operational basis. At the moment, only few countries stand out with most well-developed, promising, or already operating automated pollen measuring devices: first, Japan [11], which, however, has been able to provide information only on one pollen type (*Cryptomeria japonica*, Cupressaceae family); second, the USA, which is still under calibration and not in fully operational mode; third, Switzerland [12], which is currently under calibration; fourth, Germany, which is in fully operational mode for the last half decade [13]. Elsewhere, e. g., in Ireland, the process is in more preliminary stages.

The purpose of this paper is to provide a method to automatically classify a given pollen image into the correct pollen taxon. Our contribution to this research is fourfold. First, we describe an automated, near-real-time pollen monitoring system. Second, we manually classify and enhance the existing pollen database. Third, we develop a classification model based on deep neural networks. Fourth, we evaluate the performance of our models and compare it to that from the existing pollen monitoring system.

[1]University Center of Health Sciences at Klinikum Augsburg (UNIKA-T), Chair of Health Care Operations / Health Information Management, Faculty of Business and Economics, University of Augsburg julian.schiele@unikat.uni-augsburg.de

[2]Chair of Software Methodologies for Distributed Systems, Faculty of Computer Science, University of Augsburg fabian.rabe@informatik.uni-augsburg.de

[3]ZD.B Chair of Embedded Intelligence for Health Care and Wellbeing, Faculty of Computer Science, University of Augsburg maximilian.schmitt@informatik.uni-augsburg.de

[4]Chair and Institute of Environmental Medicine, UNIKA-T, Technical University of Munich and Helmholtz Zentrum München thanos.damialis@tum.de

[5]GLAM – Group on Language, Audio & Music, Imperial College London schuller@ieee.org

II. MATERIALS AND METHODOLOGY

In an automated pollen monitoring process, as depicted in Figure 1, air samples of the ambient environment are collected automatically on a probe (step 1), objects within each probe are detected (step 2), objects are classified as either 'no pollen' or as a specific pollen type and counted (step 3), and up-to-date information about the pollen concentration is published online[1] (step 4).

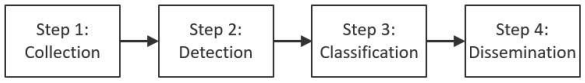

Fig. 1. Automated pollen monitoring process.

A. Materials

Automatic airborne pollen monitoring has been conducted every 3 hours since spring 2015, at the Bavarian Environmental Administration (LfU) in Augsburg, Bavaria, Southern Germany (WGS84 coordinates 48°19'37"N 10°54'7"E, 494 m above sea level), at ground level.

The automated pollen measuring device BAA500[2] continuously samples ambient air at a rate of 60 m^3/h, by use of a 3-stage virtual impactor. Pollen is deposited on a sticky surface and automatically analyzed under a microscope equipped with a camera. Images of the pollen are constructed and compared with a library of known samples. Using an image recognition algorithm, the BAA500 is able to recognize more than 30 pollen taxa [13]. For this study, we consider 15 pollen taxa, the commonest and also most abundant in the atmosphere worldwide [2]. All samples were collected between November 2015 and October 2016 and manually labeled, based on typical morphological features [10]. Table I shows the scientific names (in Latin) and common names (in English) of each pollen taxon, as well as the number of samples used in this study.

TABLE I
POLLEN TAXA CONSIDERED IN THIS STUDY.

Latin	English	Number of total samples
Alnus	Alder	10,063
Betula	Birch	2,370
Carpinus	Hornbeam	8,010
Corylus	Hazel	11,667
Fagus	Beech	728
Fraxinus	Ash	460
Plantago	Plantain	1,721
Poaceae	Grass	3,600
Populus	Poplar	2,065
Quercus	Oak	611
Salix	Willow	526
Taxus	Yew	6,077
Tilia	Linden	180
Ulmus	Elm	339
Urticaceae	Nettle	2,828

[1]https://www.unika-t.de/pollenflug/
[2]https://www.hund.de/en/instruments/pollen-monitor/

The BAA500 is equipped with software that identifies objects in the image and crops them accordingly. Figure 2 shows the cropped image for a *Betula* pollen sample. We are well aware that the provided algorithm is not optimal since we encountered many cases where the cropping was not satisfactory, e.g., objects were not identified or an object was cropped into several images (also see Figure 4 in Discussion).

Fig. 2. *Betula* pollen sample.

Those cases are not considered for this study, as we focus only on the classification (step 3 of the automated process in Figure 1). We rely on the cropped images provided by the BAA500 and use them as basis for our classification model.

B. Methodology

The classification of each pollen sample into one of the 15 considered pollen classes is done by a deep neural network model learned using 60 % of the described data set as training data. The field of deep learning has made significant progress in the past decade and – combined with today's computing capacity – offers entirely new possibilities. A comprehensive overview on deep learning is provided in [14] and [15]. Convolutional neural networks (CNNs) are a state-of-the-art method in image classification tasks, such as visual object recognition, optical character recognition, or image based medical diagnostics [16]. They usually consist of a cascade of so-called convolutional layers and one or several fully-connected neural layers. Convolutional layers are specialized for the task of feature extraction (such as edges or dots in images) by sharing the weights of their two-dimensional filter kernel across the whole input image. This makes the outputs of the single kernels shift invariant. To further reduce information and scale of the input, a maximum-pooling step is sometimes done after each convolutional layer, propagating only the largest output value of a certain neighbourhood to the next layer. Daood et al. propose a system using a cascade of pre-trained CNNs and recurrent neural networks to classify pollen [17]. They achieve a quite good performance (F-measure of 89 %) since they consider only 10 different classes and work in almost perfect conditions with 392 well-segmented, high-resolution, multi-focal image sequences obtained by a microscope. In a more complicated study design than the above, we consider 15 classes and use data from a series-production pollen monitoring system with an automated segmentation algorithm containing only one focus per pollen instance.

Before training our model, we apply some preprocessing to the cropped images provided by the BAA500. Each pollen image is embedded in the center of a black background frame, sized 256×256 pixels. The data set is converted to

tfrecords and split into three sets: the training set containing 60 % of all samples of each class, the development set with 20 %, and the test set containing the remaining 20 %.

Inspired by classic architectures, such as the MNIST standard Tensorflow model [3], we experimented with different network architectures and tuned the hyperparameters on the development set until we ended with a powerful model. A deep neural network, consisting of three convolutional layers, each followed by a maximum-pooling step, a fully-connected layer with dropout, and an output layer, is trained on the raw pixel values of the embedded images. Dropout is a technique to regularize the model and improve the ability to generalize, especially on limited data sets, by randomly setting the output of certain neurons to zero. The output layer contains one neuron for each class. As activation functions, rectified linear units (ReLU) are used for all convolutional layers and the fully-connected layer, and a softmax function is used for the output layer.

We use the *Adam* optimizer with *cross-entropy loss* to train the network [18]. Since the classes in the data set are imbalanced, the gradients are multiplied with a class-specific weight, according to the inverse of the ratio of the instances in each class with the instances of the majority class (see Eq. (1)). By doing so, we also consider the underrepresented classes with an equal impact on the loss function during training.

$$\text{Weight of class } c = \frac{\text{Max. number of samples in all classes}}{\text{Number of samples in class } c} \tag{1}$$

We use the unweighted average of the class-specific F-measures (F1) to decide when to stop the training process and which model configuration to choose. The F-measure is the harmonic mean between recall and precision for each class. The class-specific precision is defined as the share of true positives, i. e., all samples of the respective class that were correctly labeled, and all predicted positives, i. e., including samples that were wrongly assigned to the respective class in Eq. (2).

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \tag{2}$$

The class-specific recall is defined as the share of true positives and all positives, i. e., including samples that were wrongly not assigned to the respective class in Eq. (3).

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \tag{3}$$

All hyperparameters and architectural parameters are shown in Table II. They were optimized on the development set. Training was stopped after 500,000 training steps and the weights were restored from the iteration where the maximum unweighted average F-measure was reached on the development set (after 438,000 iterations).

[3] https://github.com/tensorflow/models/blob/master/official/mnist/mnist.py

TABLE II
HYPERPARAMETERS.

Parameter	Value
Convolutional layer 1	(4×4), 32 filters
Convolutional layer 2	(4×4), 64 filters
Convolutional layer 3	(4×4), 64 filters
Fully-connected layer	1,024 neurons
Dropout	0.25
Activation function	ReLU
Optimizer	Adam
Learning rate	10^{-5}
Batch size	200
Training steps	max. 500,000

The described neural network model to classify pollen images has been implemented with the Tensorflow library (version 1.7.0) in Python. For data processing, we use Pandas, NumPy, SciPy, and Scikit-learn. The training of the neural network was performed on a dedicated GPU cluster equipped with 16 Nvidia Titan X (Pascal) cards. Each network architecture and configuration during optimization was trained and evaluated on a single GPU card.

III. RESULTS

We use unweighted average precision (UAP), unweighted average recall (UAR), and unweighted average F-measure (UAF1) to evaluate the performance of our classification model. Since most of the pollen taxa included in the current study are considered allergologically important, we choose the unweighted scores as performance metrics. We achieve a UAP of 83.0 %, a UAR of 77.1 %, and a UAF1 of 79.1 % on the test set. Table III shows the performance of our model for the development set and the test set.

TABLE III
PERFORMANCE OF OUR MODEL.

Performance measure [%]	Development set	Test set
Unweighted average precision	83.2	83.0
Unweighted average recall	75.5	77.1
Unweighted average F1	78.4	79.1

The normalized confusion matrix depicted in Figure 3 shows the performance of our model for the individual classes. Each row represents the share of instances in an true class and each column represents the instances in a predicted class. The values are normalized per true label (row) and the diagonal shows the recall. *Taxus* and *Fagus* are predicted particularly well showing a recall of 96 % and 92 %, respectively. This seems reasonable as both are relatively simple to identify due to their unique pollen morphology and bigger size, respectively. However, *Quercus* and *Fraxinus* show a rather bad performance (35 % and 43 %). This is mostly due to the lower abundance of pollen samples for the respective taxa. For example, *Tilia* pollen is easier to recognize thanks to the morphologically unique colpopori whereas the finely reticulate outer surface of *Fraxinus* pollen is still a challenge for automated classification algorithms as well as for human experts. A look at Figure 3 reveals that *Fraxinus* is often confused with *Alnus*, *Betula*, and *Taxus*;

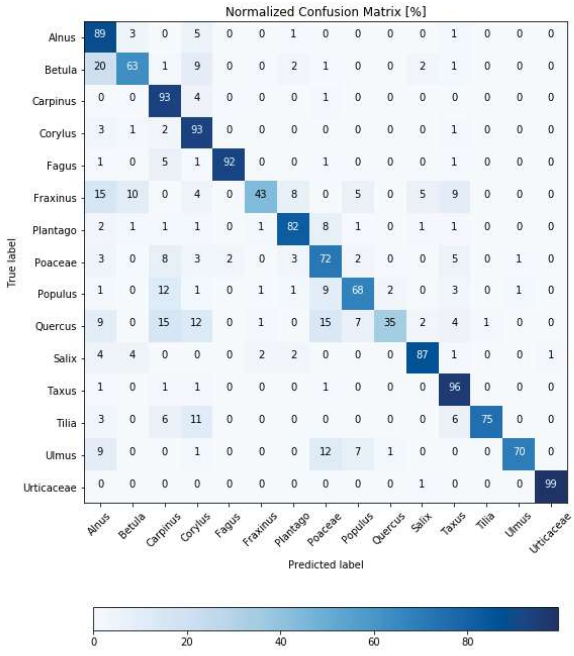

Fig. 3. Normalized confusion matrix for our model. The percentage of possible instances is shown as color scale and noted in each square.

taxa that are represented with sample sizes of a magnitude higher. Similarly, *Quercus* is confused with well-represented *Carpinus* and Poaceae.

We compared our results with the BAA500 internal classification algorithm as installed in the default configuration and without location-specific modifications. The built-in algorithm is based on a mathematical model calculating features such as area, perimeter, eccentricity, and roundness for each pollen sample. This presumably limits the model to round objects only. Moreover, features like probable season start and end seem to be included as well, despite being highly variable per year and location of sampling, which explains some of the classification errors, e. g., *Alnus*, *Betula*, and *Corylus* are not identified in October to December and are mainly classified as 'Varia' pollen. Since the BAA500 algorithm is able to predict at least 34 pollen taxa and we currently only predict 15 taxa, we present different comparisons.

First, we compare the predictions of the BAA500 to the manual annotations. We ignore samples where the annotator assigned a class which the BAA500 algorithm is not able to predict to avoid introducing a new class to the unweighted measurements, where the algorithm would receive the lowest possible score. This approach rather overestimates the performance of the BAA500 for two reasons. By dropping samples where the annotator disagreed, we are ignoring some false positive samples. Additionally, we are currently not able to provide a statement about the samples where the BAA500 predicted 'no pollen', since these are not yet manually labeled. As there are roughly five times more samples with 'no pollen' in our data set than all other classes combined, it is not unlikely that during manual annotation many false negatives would be identified in this share of the data set.

Second, we calculated the BAA500 performance on the test set containing the 15 pollen taxa which have also been used for the training and evaluation of our model. However, we are aware that this comparison rather discriminates BAA500's performance, since it has been designed for a greater variety of options (34 classes) than the one included in the data set (15 classes).

TABLE IV
COMPARISON BETWEEN BAA500 AND OUR MODEL.

Performance measure [%]	BAA500 (34 classes)	BAA500 (15 classes)	Our model (15 classes)	Our model (5 classes)
UAP	66.6	59.4	83.0	88.2
UAR	62.3	54.5	77.1	87.0
UAF1	60.1	56.4	79.1	87.5

The results are summarized in Table IV. Overall, our model is able to predict less classes, but shows superior performance than the BAA500 model. Even for 34 classes, we expect our model to still outperform the BAA500 since the current performance difference is quite large and we do not expect significant performance decrease by expanding to additional classes (which was not possible at the moment due to insufficient number of training samples). The performance of our model for 5 classes (see Table IV), i.e., *Alnus*, *Betula*, *Carpinus*, *Corylus*, and Poaceae, underlines this hypothesis.

IV. DISCUSSION AND OUTLOOK

We have developed a deep neural network based model to classify different types of airborne pollen and presented numerical results. We have shown that our model yields promising predictions for the 15 most abundant and allergenic pollen types worldwide [2], even outperforming the existing built-in algorithm of the automatic monitoring system (within the boundaries of this comparison). Moreover, our developed model consistently exhibits a high classification performance for the 5 most allergenic pollen types worldwide (Table IV), i.e., UAP and UAR of more than 85 %.

The model developed here certainly still presents specific drawbacks and some issues have to be resolved, such as expanding to an even larger number of pollen classes as the original built-in algorithm. Nonetheless, if we take into account that the huge class of 'no pollen' has been excluded from the manual classification and the comparison between the two models, it is expected that further training with correctly manually labeled samples will even further enhance our model.

Overall, automated classification of allergenic pollen is a very demanding and ambitious task. First, pollen grains exhibit a large size variability, from 10 μm to 150 μm, and infinite variations in morphological characteristics due to the 3-dimensionality in real-life. At the same time, they co-occur in the atmosphere (and therefore in pollen samples) together with other airborne particles, like fungal spores, bacteria,

viruses, insects, plant fragments, dust, and so on, which makes classification more complicated.

We are confident that our model performs better than the one currently used in BAA500, as it originally tries to eliminate flaws or shortcomings from the very beginning of the modeling concept. For instance, the BAA500 algorithm struggles to identify non-round objects as illustrated in Figure 4. Our future work will also include the development

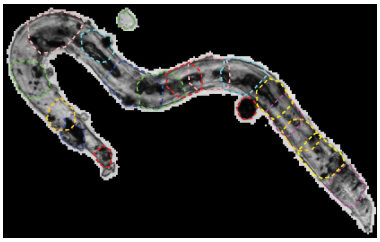

Fig. 4. Badly cropped object. Each colored region shows where the BAA500 identifies a potential sample, which is later individually classified.

and implementation of an own object detection algorithm (step 2 in Figure 1), which is expected to leverage additional performance gains by detecting more pollen samples and cropping them correctly. We also expect to be able to detect non-round objects, which would allow us to expand to automated monitoring of all other air particles as well, e. g., fungal spores.

Also, the performance of the classification itself will be further improved, in particularly by generation of additional training data, e. g., by rotation of the pollen images, and labeling of new samples. We are currently manually classifying the 'no pollen' class, which is approximately five times the size of the existing database and contains all different types of air particles, to further improve the training of the deep neural network.

By integrating all of the aforementioned features in a single classification model, it is expected that its performance will be boosted and given its already reasonably high performance, it will most probably contribute to the development of a truly operational automatic pollen classification system. Practically, this will provide the basis for accurate, real-time, automatic dissemination of allergenic pollen information. In combination with novel and fully operational pollen apps, an infrastructure for the benefit of allergic individuals will be created, providing the necessary environmental health service required as the first line of allergy management.

References

[1] Joakim Bunne et al. "Increase in Allergic Sensitization in Schoolchildren: Two Cohorts Compared 10 Years Apart". In: *The Journal of Allergy and Clinical Immunology: In Practice* 5.2 (2017), pp. 457–463.

[2] Chiara Ziello et al. "Changes to Airborne Pollen Counts across Europe". In: *PLoS ONE* 7.4 (2012).

[3] Jan L. Brożek et al. "Allergic Rhinitis and Its Impact on Asthma (ARIA) Guidelines-2016 Revision". In: *The Journal of Allergy and Clinical Immunology* 140.4 (2017), pp. 950–958.

[4] MaryJane K. Selgrade et al. "Induction of Asthma and the Environment: What We Know and Need to Know". In: *Environmental Health Perspectives* 114.4 (2006), pp. 615–619.

[5] William Storms et al. "The Economic Impact of Allergic Rhinitis". In: *Journal of Allergy and Clinical Immunology*. The Prevalence and Medical and Economic Impact of Allergic Rhinitis in the United States 99.6, Part 2 (1997), pp. 820–824.

[6] Athanasios Damialis et al. "Human Exposure to Airborne Pollen and Relationships with Symptoms and Immune Responses: Indoors versus Outdoors, Circadian Patterns and Meteorological Effects in Alpine and Urban Environments". In: *Science of The Total Environment* 653 (2019), pp. 190–199.

[7] Athanasios Damialis et al. "Estimating the Abundance of Airborne Pollen and Fungal Spores at Variable Elevations Using an Aircraft: How High Can They Fly?" In: *Scientific Reports* 7 (2017).

[8] Anna Muzalyova et al. "Pollen Allergy and Health Behavior: Patients Trivializing Their Disease". In: *Aerobiologia* (2019).

[9] Jeroen T. M. Buters et al. "Pollen and Spore Monitoring in the World". In: *Clinical and Translational Allergy* 8 (2018).

[10] British Aerobiology Federation. *Airborne Pollens and Spores: A Guide to Trapping and Counting*. National Pollen and Hayfever Bureau, 1995.

[11] Shigeto Kawashima et al. "Automated Pollen Monitoring System Using Laser Optics for Observing Seasonal Changes in the Concentration of Total Airborne Pollen". In: *Aerobiologia* 33.3 (2017), pp. 351–362.

[12] Benoît Crouzy et al. "All-Optical Automatic Pollen Identification: Towards an Operational System". In: *Atmospheric Environment* 140 (2016), pp. 202–212.

[13] Jose Oteros et al. "Automatic and Online Pollen Monitoring". In: *International Archives of Allergy and Immunology* 167.3 (2015), pp. 158–166.

[14] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. The MIT Press, 2016.

[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. "Deep Learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[16] Andre Esteva et al. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks". In: *Nature* 542.7639 (2017), pp. 115–118.

[17] Amar Daood, Eraldo Ribeiro, and Mark Bush. "Sequential Recognition of Pollen Grain Z-Stacks by Combining CNN and RNN". In: *The Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS-31)*. USA, 2018.

[18] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv:1412.6980* (2014).