Aerobiologia (2024) 40:109–121
https://doi.org/10.1007/s10453-022-09769-0

SPECIAL ISSUE: AUTOPOLLEN

Comparison of computer vision models in application to pollen classification using light scattering

Gintautas Daunys · Laura Šukienė · Lukas Vaitkevičius · Gediminas Valiulis · Mikhail Sofiev · Ingrida Šaulienė

Received: 7 March 2022 / Accepted: 19 October 2022 / Published online: 4 November 2022

Abstract This study investigates the use of pollen elastically scattered light images for species identification. The aim was to identify the best recognition algorithms for pollen classification based on the scattering images. A series of laboratory experiments with a Rapid-E device of Plair S.A. was conducted collecting scattering images and fluorescence spectra from pollen of 15 plant genera. The collected scattering data were supplied to 32 different setups of 8 computer vision models based on deep neural networks. The models were trained to classify the pollen types, and their performance was compared for the test sub-samples withheld from the training. Evaluation showed that most of the tested computer vision models convincingly outperform the basic convolutional neural network used in our previous studies: the accuracy gain was approaching 10% for best setups. The models of the Weakly Supervised Object Detection approach turned out to be the most accurate, but also slow. However, even the best setups still did not provide sufficient recognition accuracy barely reaching 65%–70% in the repeated tests. They also showed many false positives when applied to real-life time series collected by Rapid-E. Similar to the previous studies, fusion of the new scattering models with the fluorescence-based identification demonstrated almost 15% higher skills than either of the approaches alone reaching 77–83% of the overall classification accuracy.

Keywords Airborne pollen; Image recognition · Flow cytometry · Real-time monitoring

1 Introduction

A significant fraction of the world population suffers from pollen allergy (Pawankar, 2014): rhinitis alone affects between 10 and 30% of the population, being particularly high in Europe and the USA. Continuous real-time monitoring of airborne pollen concentrations is important to provide information to the health professionals and the public about changes in environmental exposure (Tummon et al., 2021). Real-time monitors face the key challenge of identifying the pollen types in the air. The conventional method of collecting pollen by a Hirst-type sampler (Hirst, 1952) and recognizing them manually through microscopic analysis (CEN/EN 16,868:2019) is inapplicable in real time due to its essentially off-line character. It requires skilled personnel and substantial time, at least one full day, to perform the recognition tasks

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10453-022-09769-0.

G. Daunys · L. Šukienė · L. Vaitkevičius · G. Valiulis · M. Sofiev · I. Šaulienė (✉)
Šiauliai Academy, Vilnius University, Šiauliai, Lithuania
e-mail: ingrida.sauliene@sa.vu.lt

M. Sofiev
Finish Meteorological Institute, Helsinki, Finland

and to announce the data. These limitations and substantial uncertainties associated with the Hirst method (Oteros et al., 2017) forced researchers to search for advanced methods of pollen monitoring (Maya-Manzano et al., 2021; Tummon et al., 2021).

Huffman et al. (2020) distinguished 8 different types of the real-time bioaerosol monitoring methods. However, only two of them: (1) fluorescence spectroscopy; (2) image analysis based on scattering, microscopy, and holography, seem to be promising for airborne pollen monitoring.

The first group is based on a technique called laser-induced fluorescence (LIF) (Pöhlker et al., 2012; Pöhlker et al., 2013; O'Connor et al., 2011; O'Connor et al., 2014). It is based on induction of pollen grain fluorescence with monochromatic laser or LED light, which excites organic molecules of the particle topmost layers. Each biomolecule radiates in a specific wavelength range with a characteristic spectrum. Since the signal is a superposition of response of many species comprising the pollen grain, the identification problem gets complicated. Nevertheless, experience shows that in many cases it is possible to distinguish pollen types from each other by their fluorescence (Crouzy et al., 2016; Šaulienė et al, 2019). Field applications of these devices also showed significant potential of the approach (Crouzy et al., 2016; Després et al., 2012; Tešendić et al., 2020).

The second group of methods discriminates pollen by optical properties, which depend on pollen size, form, and surface texture. There are several different approaches inside the group. One approach is to acquire photographic images of collected pollen and recognize them with computer vision algorithms (Oteros et al., 2015; Schaefer et al., 2021). A generalization of this approach uses holographic imaging, which eliminates the problem of obtaining a focused image of a particle flying through detectors in real time (Sauvageat et al., 2020). An earlier method of pollen tracking (Kawashima et al., 2007, 2017) is based on laser optics, where light scattered by a pollen grain is measured at two angles. Despite its simplicity, that method proved to be capable of discriminating two key pollen types in Japan, which was enough to build a real-time monitoring network (Miki et al., 2021).

The above two groups of methods have been combined in several devices, among which two-Rapid-E of Plair S.A. and Poleno of Swisens-provide the most-comprehensive set of parameters for each particle. The current study concentrates on the technology of Rapid-E and continues the developments of Crouzy et al. (2016), Šaulienė et al. (2019), Tešendić et al. (2020), and Daunys et al. (2021).

A specific challenge discovered by the previous study of Rapid-E (Šaulienė et al., 2019), hereinafter referred to as S19, was the strikingly low accuracy of the scattering-based convolutional neural network (44%) despite the rich multi-channel signal provided by the device. A fluorescence-only algorithm scored 67%, whereas the approach using both principles showed an accuracy of 74%. This difference was surprising because the classical manual recognition method is based on just visual analysis of microscopic image (the approach also used by BAA500 of Hund Wetzlar without any fluorescence component—Oteros et al. (2015), Schaefer et al. (2021)). A practical dimension of the problem is that the fluorescence-inducing laser of Rapid-E is the most expensive and the least reliable component of the device. According to our experience, its lifetime is about two years, whereas its replacement requires a full recalibration of the device and re-training of the recognition algorithm.

Motivated by the above problem, this study aimed at optimizing the performance of the scattering-only recognition by employing several modern computer vision models. The second aim was to estimate the added value of combining the new scattering-based model with the fluorescence-based model of S19 and to evaluate the practical feasibility of restricting the real-life device operations to the scattering signal alone.

The paper is organized as follows. The next section outlines the principles of operations of Rapid-E, presents the sample datasets, and lists the analytical approaches tested in the study. The pollen identification outcome is presented in the Results section. The skills of the newly tested methods are compared with previously developed algorithms in Discussion.

2 Materials and methods

2.1 A protocol of the experiment

The experiment was set as an N-fold cross-validation of a series of modern and classical image analysis

methods. To maintain the amount of computations under control, just two folds were made. However, we selected the sub-samples in a way to explore the maximum diversity between the folds.

All tried methods were trained with a reference set of scattering images of known pollens (about 10,000 of each pollen type) and subsequently tested with the pollen grains withheld from the training. The skills of each method were recorded in a form of confusion matrix and multilabel classification accuracy (further in text referenced as accuracy) calculated from it as a percentage of true classifications from all items.

The second step involved the neural network designed in S19 for the fluorescence signal. Its results were merged with the scattering recognition by a simple summation of probabilities. The procedure was repeated for both folds revealing the uncertainty of the comparison.

Upon completion of the individual-methods experiments, the best approaches were selected and fused together in various combinations.

2.2 Plair Rapid-E device

Rapid-E of Plair S.A. makes use of two physical principles, scattering of a laser beam and a laser-induced fluorescence, to describe each particle that passes through the inspection camera (Kiselev et al., 2011, 2013).

In theory, the multi-angle scattering images strongly depend on the particle morphology, such as size and shape, and thus, can identify the pollen grain if a sufficient number of channels and view angles are available. Rapid-E has 24 high-frequency sensors, which register the scattered light while particle passes through the laser beam (Fig. 1, left-hand panel). The passage time depends on particle size, position in the air jet, and orientation. Therefore, the height of the image is always 24 pixels, while its width is specific to each particle (see Fig. 1, left panel, and examples in Šaulienė et al., 2019). Particles do not move through the beam in a perfectly consistent position and can even be spinning, which is arguably the most-complicated part of the problem, potentially causing the low identification skills in scattering-based classifiers. The mean scattering image depends on the size of the specific pollen and primary scattering directions (Fig. 1, middle panel). Importantly, this signal is not identical across the Rapid-E devices: even small difference in the laser alignment and sensors sensitivity leads to noticeable differences in the images (Fig. 1, right-hand panel shows the mean scattering for the same *Betula pendula* pollen sample recorded by the Rapid-E of Finnish Meteorological Institute). These issues largely complicate the development of unified recognition algorithms applicable to several devices and reiterate the necessity of using the modern computer vision techniques as they might be more robust than the classical convolution neural network.

The scattering signal allows for construction of integral morphological features suggested by Crouzy et al. (2016) and used by Šaulienė et al. (2019), Tešendić et al. (2020) and Boldeanu et al. (2021). These studies employed deep neural networks (DNN) with convolutional blocks for analysis of this dataflow.

The second information channel is the fluorescence recordings. The fluorescence data are collected in two ways: for 32 wavelengths at 8 consecutive time moments and for 4 high-frequency channels, which register the fluorescence evolution with time (Šaulienė et al., 2019). At the first and, sometimes, the second time moments, the fluorescence signal could get saturated. However, we left such particles for processing.

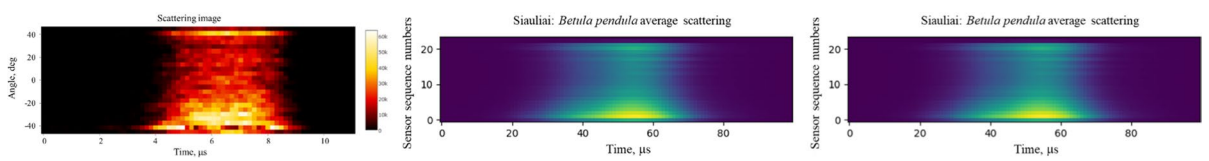


Fig. 1 Example of a scattering image of a single pollen particle (left panel, courtesy of J. Palamarchuk), mean scattering image over > 1000 particles (middle panel), mean scattering image over > 1000 particles for another Rapid-E device (right-hand panel, Helsinki, FMI)

2.3 Experimental data

We used an extension of the S19 dataset, which was obtained following the procedure described in Šaulienė et al. (2019). The Rapid-E device was supplied with samples of 15 pollen taxa (Table 1; Valiulis et al., 2020, 2021). In comparison with the S19 study, we added 5 new species (groups of species) marked with an asterisk in Table 1.

Laboratory work of supplying known pollens to the device was supplemented with ambient air observations to collect recordings for a non-pollen class of particles. This class consisted of ambient-air particles, which were registered by the device during days when pollen concentrations in the air (according to the standard pollen trap of Hirst type) were much lower (< 1%) than the total coarse particle concentrations observed by the device. Therefore, the vast majority of the "non-pollen" class are the coarse particles of various origin (over 5 µm in diameter, filter applied by Rapid-E at the time of collection).

To cope with a large variety of these particles and a possible small fraction of pollen in this class, the number of samples in it was made very large (Table 1). In total, 16 classes (15 genus pollen morphotypes and one class of non-pollen) comprised the datasets for identification models.

Table 1 Number of images per pollen taxa used for model training and evaluation

Plant family	Species	Number of particles
Betulaceae	*Alnus glutinosa*	10,447
	Betula pendula	11,838
	Corylus avellana	10,763
Oleaceae	**Fraxinus excelsior*	12,181
Fagaceae	*Quercus robur*	8384
Salicaceae	*Populus tremula*	10,807
	**Salix caprea*	13,403
Aceraceae	*Acer negundo*	8546
Cupressaceae	*Juniperus communis*	10,015
Pinaceae	**Picea abies*	5276
	Pinus sylvestris	8293
Asteraceae	**Ambrosia artemisiifolia*	12,677
	Artemisia vulgaris	13,235
Poaceae	**Dactylis glomerata*	10,442
	Festuca pratensis	7611
Non-pollen	*any particle*	326,577

The dataset was split to training and testing subsets. For the fold-1, the last 1000 particles of each type supplied to the device were withheld from the training subset. For the fold-2, the first 1000 particles were taken for testing. Being a deviation from the classical N-fold evaluation procedure, which requires random splits, this method ensures the largest possible difference between the testing sets of the folds. Despite cleaning the device by isolating it from the ambient air and leaving it idle for a while before starting the next pollen type, contamination by pollen from the previous round was still possible (also noticed by other groups working with Rapid-E). As a result, the samples might have a slightly inhomogeneous quality, where the first fold might be more contaminated and the second fold less contaminated.

2.4 Pollen classification algorithms

We used classification models from two families of neural networks: Convolutional Neural Networks and Vision Transformers. Experiments were performed with different architectures within each family, selected in accordance with the model performance. Each architecture had several setups with varying number of layers and number of neurons within each layer (marked as "a," 'b," etc., indices added to the method abbreviation).

To compare the results with the previous studies and reveal the added value of the new methodologies, the identification has been done twice for each method: once based on the scattering images alone and once fusing the result of the scattering and the S19 fluorescence model.

A Pytorch deep learning framework was used for training and evaluation of all models. They were trained on one GPU using Adam optimizer with a learning rate scheduler.

2.4.1 Computer vision models based on convolutional neural networks

ConvNets (Convolutional neural networks) were first introduced in the 1990s (LeCun et al., 1998) and became popular after 2012 when AlexNet (Krizhevsky et al., 2012) won the 2012 ImageNet competition. Many ConvNets modifications have been developed, but most of them have not been used to classify

scattering images of airborne particles of biological origin.

A simple ConvNets structure with 2 convolutional blocks used in the S19 study formed a starting point for the experiment. That model is marked as version "ConvNets_a" in Table 2. Recently, it was demonstrated by Boldeanu et al. (2021) that the architecture with 3 layers gives better accuracy-the version "ConvNets _b" in Table 2. On the other hand, it is well known that increase in the network layers can lead to worse results because of overfitting (Bishop, 1995). Training the deep neural networks is also challenging because of the problems of vanishing or exploding gradient. Therefore, we limited the experiment with these two setups.

RepVGG architecture has been introduced quite recently (Ding et al., 2021). In comparison with the simple ConvNets, it has several advantages: the RepVGG topology is feed-forward without any branches, the convolutional part uses only 3×3 convolutions and ReLU activation. We tested the RepVGG_A0 configuration (https://github.com/DingXiaoH/RepVGG) changing the input channels from three (RGB image) to one (monochromatic image). Several configurations were tested with different number of layers in blocks and width multipliers.

ResNet (He et al., 2015) architecture introduces shortcut connections, which help solving the issue of the vanishing gradient. The 1×1 convolution allowed not only to reduce the number of parameters in the network but also to improve the network ability to handle nonlinearities ResNet has models with a different number of layers (18, 34, 50, 101, 152). Because the scattering image is quite small, we tried Resnet-18 and ResNet-34 architectures. Models were built using Pytorch ResNet implementation (https://pytorch.org/vision/0.8/_modules/torchvision/models/resnet.html).

EfficientNet (Tan & Lee, 2019; Tan & Lee, 2021) is an easily scalable neural network structure. For our experiments, we used the EfficientNetV2-S configuration (https://github.com/d-li14/efficientnetv2.pytorch).

2.4.2 Transformer-based architectures

Vision transformers have been successfully employed for natural language processing tasks. Recently, there have been many attempts to apply them to object recognition in computer vision tasks.

ViT Vision Transformer (Dosovitskiy et al., 2020) was selected for its high capacity and several successful modifications suitable for image analysis (https://github.com/lucidrains/vit-pytorch). Its application involved several steps: (i) the input image is divided to fixed-size parts, (ii) their feature vectors (called tokens) are calculated, (iii) feature vectors related to position–position embeddings are calculated, (iv) tokens and position embeddings are supplied to the transformer.

The method, however, has two significant drawbacks (Yuan et al., 2021a): (i) the direct token calculation from the input images by a hard split makes ViT unable to model the image local structures like edges and lines, thus requiring much larger samples; (ii) the attention backbone of ViT is redundant and leads to limited feature richness and difficulties in the model training.

ViT-CCT, ViT Compact Convolutional Transformer (Hassani et al., 2021), is one of the recently proposed ViT modifications (Khan et al., 2021). The introduced modifications reduced the needs in large databases for the model training. The model eliminates the requirement for positional embeddings through a novel sequence pooling strategy and use of convolutions.

VOLO is another transformer-based model (Yuan et al., 2021b). It addresses the low ViT efficacy in encoding the fine-scale features and contexts into the token representations. The modification used small image patches for token calculations: 8×8 instead of 16×16 pixels. It also involved a structure called Outlooker to generate more expressive token representations at the fine level. In the second stage of the algorithm, another patch embedding module is utilized to down-sample the tokens. A sequence of transformers is then adopted to encode global information. The authors proposed five versions of the model: VOLO-D1—VOLO-D5, where VOLO-D1 is the simplest. We used VOLO-D1 and its even further simplified versions (https://github.com/lucidrains/vit-pytorch).

2.4.3 Weakly Supervised Object Detection

Today's state-of-the-art object detector can achieve near-perfect performance with fully supervised settings, i.e., Fully Supervised Object Detection

Table 2 Setups tested in the experiments. The starting point, the setup used in S19 study is ConvNet_a, highlighted with bold font and a shadow

	Layers	Width multipliers
RepVGG		
a	[2, 4, 14, 1]	[0.75, 0.75, 0.75, 2.5]
b	[2, 4, 6, 1]	[0.75, 0.75, 0.75, 2.5]
c	[2, 3, 3, 1]	[0.75, 0.75, 0.75, 2.5]
d	[1, 2, 2, 1]	[0.75, 0.75, 0.75, 2.5]
e	[1, 2, 2, 1]	[0.5, 0.5, 0.5, 1.0]
f	[1, 2, 2, 1]	[0.5, 0.5, 0.5, 0.5]
g	[1, 1, 1, 1]	[0.75, 0.75, 0.75, 2.5]
h	[1, 1, 1, 1]	[0.5, 0.5, 0.5, 1.0]
i	[1, 1, 1, 1]	[0.5, 0.5, 0.5, 0.5]
j	[2, 3, 3, 1]	[0.5, 0.5, 0.5, 1.0]
k	[2, 3, 3, 1]	[0.5, 0.5, 0.5, 0.5]
	Channels	Squeeze-excitation
EfficientNet		
a	[24, 48, 64, 128, 160, 256]	[0,0,0,1,1,1]
b	[12, 24, 32, 64, 80, 128]	[0,0,0,1,1,1]
c	[24, 48, 64, 128, 160, 256]	[0,0,0,0,0,0]
	Patch size	MLP dim
ViT		
a	[12, 12]	128
b	[6, 12]	128
c	[6, 12]	256
d	[12, 12]	256
e	[24, 2]	256
	Num. heads	Embed. Dims
VOLO		
a	[3, 6, 6, 6]	[192, 384, 384, 384]
b	[3, 6, 6, 6]	[192, 192, 192, 192]
c	[3, 3, 3, 3]	[192, 384, 384, 384]
	Nbr of conv. blocks	
ConvNet		
a	2	
b	3	
	Layers	
ResNet		
a, ResNet-18	[2, 2, 2, 2]	
b	[1, 2, 2, 1]	
c	[1, 1, 1, 1]	
d, ResNet-34	[3, 4, 6, 3]	

Table 2 (continued)

	Num. layers
ViT-CCT	
a	14
b	12
c	10
Wildcat (no parameters to alter)	

(FSOD). Unfortunately, these methods suffer from two inevitable limitations: (i) the large-scale instance annotations are difficult to obtain and labor intensive, (ii) labelling the input data may inadvertently introduce annotation biases.

To avoid these problems, the community starts to solve the object detection with weakly supervised settings, i.e., Weakly Supervised Object Detection (WSOD). WSOD classifies and locates object instances using only image-level labels in the training phase. We tested the WSOD Wildcat method (Durand et al., 2017).

2.4.4 Tested setups

A full set of tested setups is summarized in Table 2. Each method has up to two parameter(s) to alter—shown in the table for each method.

3 Results

The main results of the experiment are presented in Fig. 2. The computations revealed a few poorly working models and setups but for every method at least one setup was reaching or exceeding the level of 65% for scattering-only skills of the fold-1 (a few % less for the fold-2). Both folds show the same relations between the different classification models, i.e., the small difference was only in the absolute skills of the methods, not in their relation to each other. Therefore, the obtained ranking of the models and setups is robust.

The second important result is that the S19 algorithm, the left-most bar in Fig. 2, has been substantially outperformed by most of the modern approaches. The best model, Wildcat, outperformed it by almost 10%.

From a practical standpoint, however, the higher skill of the scattering-based classification was still way below the dual-channel model quality. For the fusion of scattering- and fluorescence-based methods, the maximum skills were almost 15% higher than those for scattering: close to 83%.

Looking at the details of the methods performance (see the supplemental information for the quantitative scores), there is a tendency of lower model performance when the number of layers or general complexity of the network is reduced. Such reduction in scores is significant for scattering-only analysis and only partially alleviated by the dual scattering-fluorescence fused model. It also suggests that complexity of some models might be even increased with minor risk of over-fitting.

The efficiency of the methods was measured as time needed for processing one image with a single CPU core (Fig. 3). The absolute values strongly depend on nuances of the hardware and software configuration and therefore are only indicative. But their ratios show the relative costs of each method compared to others (can still be influenced by cache size, speed of memory vs CPU efficiency, etc.). Figure 3 plot shows very strong differences between the models. Processing time varies from about 1 ms (milliseconds) for ConvNet_a up to over 40 ms for EfficientNet_c. Two tendencies can be noticed: (i) simpler methods are faster but less accurate, e.g., ConvNet_a is the fastest and the worst; (ii) for sophisticated methods, the higher complexity does mean low speed but does not necessarily correspond to proportional gain in accuracy. In particular, the most-accurate WildCat, while slow, is by no means the slowest: 19 ms is the 4-th slowest result among the tested algorithms. A graphical representation of tradeoff between model accuracy and its processing time is presented in Fig. 4. There are included only

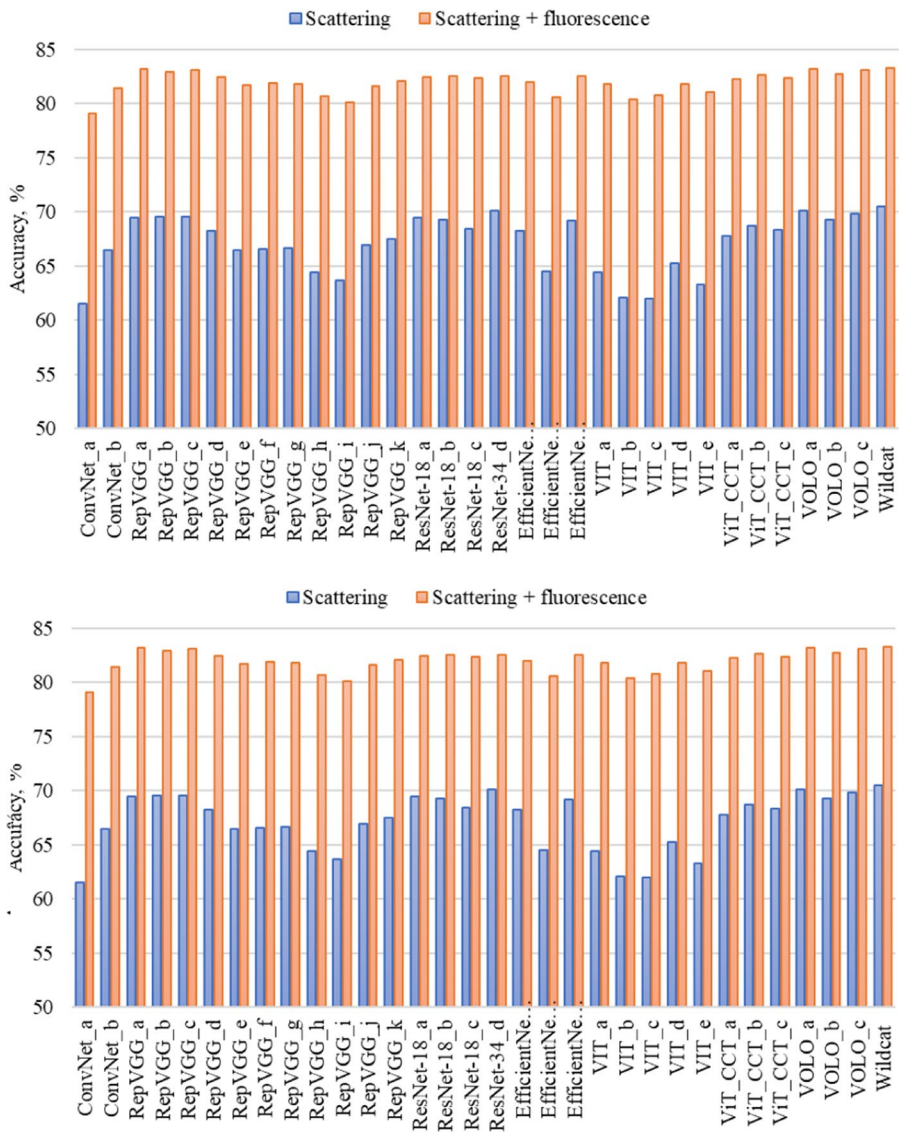


Fig. 2 Skills of different approaches applied to the extended pollen set of S19. Upper panel: fold-1, lower panel: fold-2

models which accuracy is higher 68% and processing time is less 20 ms.

Selecting an optimal combination of acceptable speed (Supplement 1) and high accuracy (Supplement 1, according to scattering + fluorescence), three models-RepVGG_c, ResNet_b, and VOLO_a-were taken to the next round of the fusion experiments. The models were fused pairwise in all possible pairs, and also a fusion of all three of them was generated as an "ultimate" method (Table 3). As expected, a fusion of all three models gave the best result, but the difference between ResNet and VOLO models was insignificant in most cases. Scattering-only processing was only a tiny bit worse than ResNet + VOLO. One can also notice a strong performance of the VOLO_a setup: it outperformed other two leading methods and did not lose much to any of other combinations.

Confusion matrices (Fig. 5) provide an in-depth view of the VOLO_a performance for individual pollen types and their families listed in Table 1. They can also be compared with the corresponding matrices in S19 (except for the newly added species). One can notice the substantial similarities with the S19 conclusions. As repeatedly pointed out in S19 and in other earlier works, pollen of the Betulaceae family (*Alnus*, *Betula*, *Corylus*) are all

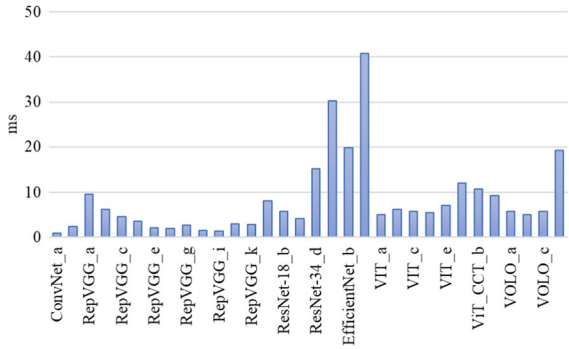


Fig. 3 Time needed for analysis of one image by a pre-trained model, [ms]. Tests were performed with a single-CPU Intel(R) Core(TM) i5-9400F CPU @ 2.90 GHz, RAM 64.0 GB

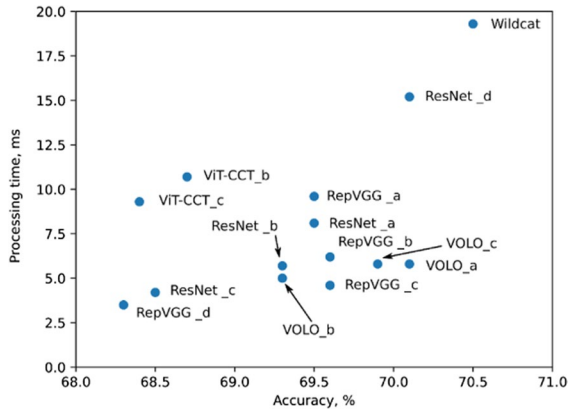


Fig. 4 Model's processing time versus its accuracy for best models

Table 3 Results of best models and their fusion

Model	Accuracy for fold-1	
	Scattering-only (%)	Scattering + fluorescence (%)
RepVGG_a	69.6	83.1
ResNet_b	69.3	82.6
VOLO_a	70.3	83.2
RepVGG_a + ResNet_b	71.3	83.6
RepVGG_a + VOLO_a	71.6	83.7
ResNet_b + VOLO_a	71.7	83.7
RepVGG_a + ResNet_b + VOLO_a	72.4	83.8

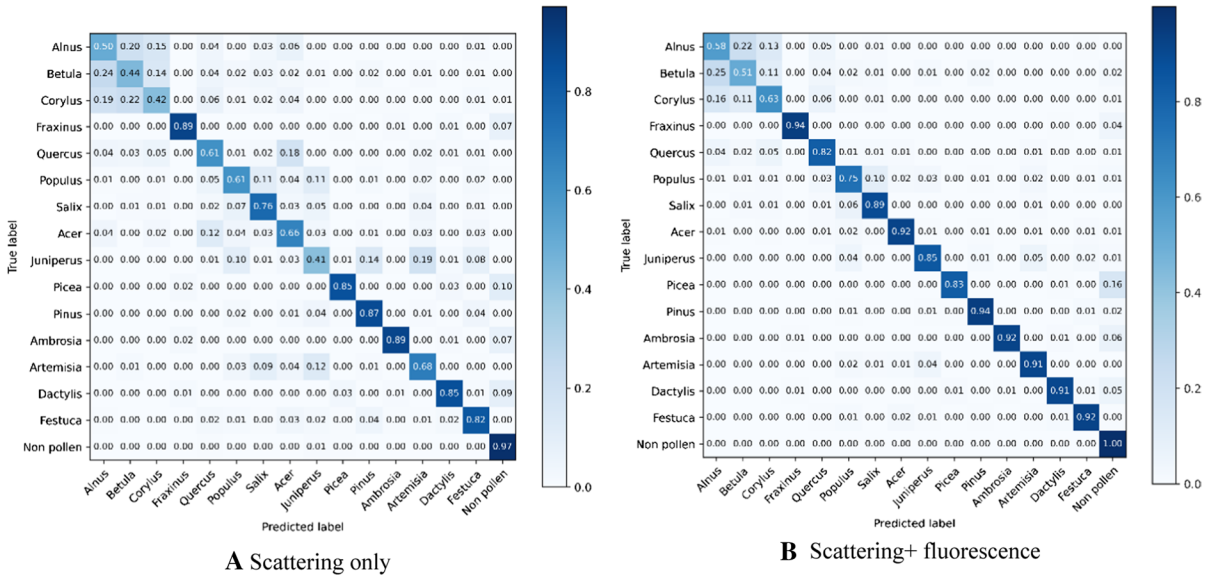
but indistinguishable from each other. In the current case, the scattering skills inside this group are about 40% (a pure random attribution would evidently give 33%). For the dual model scattering + fluorescence, the results are somewhat better-accuracy reaches 50%.

The confusion matrix (Fig. 5) also explains an otherwise unexpected fact that the skills of the reference method ConvNet_a, in the current experiment (62%) are noticeably higher than those reported by S19 (44%). This is because the set of pollens was enriched with 5 species, which turned out to be well recognized (almost 90% of correct recognition of *Ambrosia*, *Picea*, and *Fraxinus*, 85% of *Dactylis*, and 97% of non-pollen particles). The example of Fig. 5 is for VOLO_a, but the same tendency exists in other methods.

4 Discussion

4.1 Sampling uncertainty and the device "memory"

Comparing the panels of Fig. 2, one can see about-5% lower recognition skills for the fold-2 dataset, which used first 1000 particles as the test subset. It confirms that at the beginning of each experiment a few % of extra contamination is routinely recorded in comparison with the rest of the dataset, despite the precautions taken against it. For this study, such tendency was rather beneficial as it gave a possibility to test the methods for different levels of contamination of the sample. For practical applications, this feature can cause problems in daily operations. Indeed, with an air flow rate of 2 l min^{-1}, the device sucks about 2.6 m^3 day^{-1}. Pollen concentrations in the middle of, e.g., birch season, can exceed 1000 grains m^{-3} as a daily average. As a result, within a day or two, the device will deal with about the same number of particles as in the samples of Table 1. The device "memory" can then noticeably affect the following days and suggest a longer season than it is in reality. This real-life problem can be even more significant than in the current experiment because in this study the device was cleaned and flushed between the samples, whereas on real-life applications it will be operating on a continuous basis. This issue deserves further investigation in follow-up studies.



A Scattering only **B** Scattering+ fluorescence

Fig. 5 Confusion matrices of the VOLO_a method. Noise in the matrices (the last row and the right-most column) indicate the non-pollen particles

4.2 Best models for scattering-based particle classification

Considering the results of the tested methods and set-ups, one can see that the best methods are quite close to each other (about 70% and 65% for the folds 1 and 2, respectively). However, the poorly scoring methods lost more in the fold-2. ConvNet_a had the worst performance (less by almost 10%); however, it is also the simplest, fastest, and arguably the most-classical among all tested algorithms. It looks like the simpler methods are more sensitive to the cleanness of the datasets. Since pollen in real atmosphere can be noticeably different from those supplied in laboratory conditions, one has to pay particular attention to choose the robust approaches despite their relatively high costs.

For both datasets, the best results were shown by the Wildcat model with the ResNet backbone. But the gain in accuracy is not significant compared to the VOLO model, which is about 4 times faster (Fig. 4). The reason for high costs is that the algorithm Wildcat, in the full image, tries to detect a true place of the pollen fingerprint. Wildcat also had a larger fraction of false positives. The result of VOLO corroborates with the results obtained on computer vision datasets (Yuan et al., 2021b), thus confirming that the attention-based DNN architectures are promising for scattering image processing.

Most of the tested computer vision models show better performance than simple ConvNet used in the S19 study or recently by other researchers (Boldeanu et al., 2021). Combination of the best models resulted in only slightly higher recognition accuracy.

4.3 Scattering and fluorescence recognition channels

One of the practical motivations for the current study was high costs associated with the fluorescence signal. We tried to find a method that might reach an acceptable level of recognition skills using only scattering signal of Rapid-E, thus extending the interval between the major device maintenance and recalibration. That idea was in line with earlier attempts to recognize pollen with two scattered-light sensors (Kawashima, 2007). Unfortunately, that goal of the study has been reached only partly. From one side, the added value of fluorescence channel has been reduced from ~ 20% of the skills gain in S19 to < 15% of this study. From the other side, these 15% extra accuracy is still far too large to be ignored. We conclude that usage of both channels is necessary to reach a minimally acceptable level of skills of 80%.

Such a conclusion, albeit in line with previous studies, is still somewhat surprising. The angular distribution of elastically scattered light is a more comprehensive signal than the fluorescence spectral records made with just one excitation wavelength and, in theory, contains more information. Images of the grains (both photographs and holograms) are used by other devices with quite decent results (Oteros et al., 2015; Schaefer et al., 2021). On the other hand, scattering image depends on factors not directly associated with the particle type: it might rotate while passing the laser beam, can be swollen or dehydrated, damaged in various ways, etc. All these factors may substantially complicate the analysis and, in the end, make this signal noisy. Updated with "In addition, devices are already available on the market that use only morphological characteristics of pollen to identify bioaerosol particles in near real-time. The BAA500 device take an image of each detected particle and provide results after image processing. The noninvasive measurement method based on pollen shape and size is implemented at Swisens Poleno, where the structure of airborne particles is recognized from holographic images." Measuring polystyrene spheres showed that the image-based approach works better than the scattering signal one; however, it can only be applied to larger particles (Lieberherr et al., 2021).

Fluorescence addresses a different dimension-chemical composition of the particle skin, which is independent from the particle motion (but in ambient-air conditions can be affected by its water content or physical damage). As a result, the comparatively limited signal appears more informative than the complicated scattering image. It is worth mentioning that in ambient air, pollen can have smaller particles stuck to their surface, which can substantially alter both scattering and fluorescence signals. Such particles will, most-probably, not classified as pollen by this technology.

To save the resource of the expensive and short-lived fluorescence-inducing laser, one might consider some hybrid approaches. For instance, the fluorescence channel might be activated for a fraction of time to provide a reference profile of pollens in the air, which is used as a "default" template between the full-signal intervals. Feasibility and quality of such approaches need to be clarified in follow-up studies.

4.4 Classification inside families of similar pollens: still problematic

As seen from the confusion matrix, a majority of incorrect recognitions is related to the miss-classification inside the same plant family. It is particularly important for Betulaceae. Three tested genera from Betulaceae, *Alnus*, *Betula*, *Corylus,* are heavily misclassified inside the family even by the best setup of the study (Fig. 5). Significant misclassification was found in the Salicaceae family (tested pollen of genera *Populus*, *Salix*, 10% of a mix-up with each other). Exceptional behavior was found only for two representatives of the Asteraceae family: *Ambrosia* and *Artemisia* are practically never mixed with each other. These conclusions are common also for other studies that include the corresponding pollens (Boldeanu et al., 2021; Šaulienė et al., 2019).

5 Conclusions

A set of 32 different computer vision models and model setups has been applied to the problem of classification of 15 pollen types based on their scattering images generated by the Rapid-E flow cytometer.

Several models showed similar performance, but clear differences in the skills were observed for many approaches and reproduced through the twofold cross-evaluation. The best recognition accuracy from the scattering image was achieved with the Wildcat model, which uses the ResNet model as the backbone. However, this model is three times more expensive computationally than several methods that took the second place with a small margin: RepVGG_a, ResNet_b, and VOLO_a.

In real-life recognition systems, it would be appropriate to use the VOLO model setup "a" (Table 2) because it showed fewer false-positive recognitions than other methods (albeit still too many for a standalone real-life applications). No combination of the three winning methods was found to be substantially superior to a single-model setup.

Fusing the scattering model and fluorescence-based recognition approach described in our previous publication significantly improved the overall skills and reduced the number of false positive recognitions, finally approaching a minimally acceptable level of recognition skills of 80%. The gain compared

to the scattering-only algorithm has been reduced from over 20% in the previous studies to ~ 15% here, which is still a very significant contribution.

The resulting overall accuracy of the combined new scattering + fluorescence recognition models reached 83%, more than 10% up from the previous-study recognition algorithm. However, the bulk of the improvement was due to addition of 5 new pollen types, which appeared well-recognizable.

The experiment showed that methods based on scattering alone cannot be considered for the real-life monitoring, i.e., the expensive fluorescence data are necessary to obtain acceptable identification skills.

Acknowledgements This research has been supported by the European Social Fund (project no. 09.3.3-LMT-K-712-01-0066) under grant agreement with the Research Council of Lithuania (LMTLT) and (1.57) 15600-INS-138 Agreement for the Assessment of Bioaerosol Concentrations in Real time. Support of Academy of Finland project PS4A (grant nbr 318194) is kindly appreciated.

Author contributions GD ran simulations, processed model outputs, analyzed model performance, and prepared the draft of the manuscript. IS, LS and GV performed measurements, LV took responsibility for data curation, and IS, LS, LV and MS reviewed and edited the manuscript. IS took care of project administration.

Data availability The calibration dataset is available in the Zenodo repository (Valiulis et al., 2020, 2021).

References

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Boldeanu, M., Cucu, H., Burileanu, C., & Mărmureanu, L. (2021). Multi-input convolutional neural networks for automatic pollen classification. *Applied Sciences, 11*(24), 11707. https://doi.org/10.3390/app112411707

CEN/EN 16868:2019 (2019). Ambient air-Sampling and analysis of airborne pollen grains and fungal spores for networks related to allergy networks–Volumetric Hirst method. *European Standard*, European Committee for Standardisation, Brussels, Belgium (p. 38).

Crouzy, B., Stella, M., Konzelmann, T., Calpini, B., & Clot, B. (2016). All-optical automatic pollen identification: Towards an operational system. *Atmospheric Environment, 140*, 202–212. https://doi.org/10.1016/j.atmosenv.2016.05.062

Daunys, G., Šukienė, L., Vaitkevičius, L., Valiulis, G., Sofiev, M., & Šaulienė, I. (2021). Clustering approach for the analysis of the fluorescent bioaerosol collected by an automatic detector. *PLoS ONE, 16*, e0247284. https://doi.org/10.1371/journal.pone.0247284

Després, VivianeR., Huffman, J. A., Burrows, S. M., Hoose, C., Safatov, AleksandrS., Buryak, G., Fröhlich-Nowoisky, J., Elbert, W., Andreae, MeinratO., Pöschl, U., & Jaenicke, R. (2012). Primary biological aerosol particles in the atmosphere: A review. *Tellus b Chemical and Physical Meteorology, 64*, 15598. https://doi.org/10.3402/tellusb.v64i0.15598

Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., & Sun, J. (2021). Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13733–13742).

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., & Houlsby, N. (2020). An image is worth 16 × 16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. https://doi.org/10.48550/arXiv.2010.11929.

Durand, T., Mordan, T., Thome, N., & Cord, M. (2017). Wildcat: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 642–651).

Hassani, A., Walton, S., Shah, N., Abuduweili, A., Li, J., & Shi, H. (2021). Escaping the big data paradigm with compact transformers. arXiv preprint arXiv:2104.05704.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. arXiv 2015. arXiv preprint arXiv:1512.03385.

Hirst, J. (1952). An automatic volumetric spore trap. *Annals of Applied Biology, 39*(2), 257–265. https://doi.org/10.1111/j.1744-7348.1952.tb00904.x

Huffman, J. A., Perring, A. E., Savage, N. J., Clot, B., Crouzy, B., Tummon, F., & Pan, Y. (2020). Real-time sensing of bioaerosols: Review and current perspectives. *Aerosol Science and Technology, 54*(5), 465–495. https://doi.org/10.1080/02786826.2019.1664724

Kawashima, S., Clot, B., Fujita, T., Takahashi, Y., & Nakamura, K. (2007). An algorithm and a device for counting airborne pollen automatically using laser optics. *Atmospheric Environment, 41*(36), 7987–7993. https://doi.org/10.1016/j.atmosenv.2007.09.019

Kawashima, S., Thibaudon, M., Matsuda, S., Fujita, T., Lemonis, N., Clot, B., & Oliver, G. (2017). Automated pollen monitoring system using laser optics for observing seasonal changes in the concentration of total airborne pollen. *Aerobiologia, 33*(3), 351–362. https://doi.org/10.1007/s10453-017-9474-6

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2021). Transformers in vision: A survey. arXiv preprint arXiv:2101.01169.

Kiselev, D., Bonacina, L., & Wolf, J. P. (2011). Individual bioaerosol particle discrimination by multi-photon excited fluorescence. *Optics Express, 19*(24), 24516–24521. https://doi.org/10.1364/OE.19.024516

Kiselev, D., Bonacina, L., & Wolf, J. P. (2013). A flash-lamp based device for fluorescence detection and identification of individual pollen grains. *Review of Scientific Instruments, 84*(3), 033302. https://doi.org/10.1063/1.4793792

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 3*, 25.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324. https://doi.org/10.1109/5.726791

GLieberherrKAudersetBCalpiniBClotBCrouzyMGysel-BeerT-KonzelmannJManzanoAMihajlovicAMoallemiDO'ConnorBSikoparijaESauvageatFTummonKVasilatou2021Assessment of real-time bioaerosol particle counters using reference chamber experimentsAtmospheric Measurement Techniques10.5194/amt-14-7693-2021Lieberherr, G., Auderset, K., Calpini, B., Clot, B., Crouzy, B., Gysel-Beer, M., Konzelmann, T., Manzano, J., Mihajlovic, A., Moallemi, A., O'Connor, D., Sikoparija, B., Sauvageat, E., Tummon, F., & Vasilatou, K. (2021). Assessment of real-time bioaerosol particle counters using reference chamber experiments. Atmospheric Measurement Techniques. https://doi.org/10.5194/amt-14-7693-2021

Maya-Manzano, J. M., Smith, M., Markey, E., Hourihane Clancy, J., Sodeau, J., & O'Connor, D. J. (2021). Recent developments in monitoring and modelling airborne pollen, a review. *Grana, 60*(1), 1–19. https://doi.org/10.1080/00173134.2020.1769176

Miki, K., Fujita, T., & Sahashi, N. (2021). Development and application of a method to classify airborne pollen taxa concentration using light scattering data. *Scientific Reports, 11*(1), 1–12. https://doi.org/10.1038/s41598-021-01919-7

O'Connor, D. J., et al. (2011). The intrinsic fluorescence spectra of selected pollen and fungal spores. *Atmospheric Environment, 45*(35), 6451–6458.

Oconnor, D. J., et al. (2014). Using spectral analysis and fluorescence lifetimes to discriminate between grass and tree pollen for aerobiological applications. *Analytical Methods, 6*(6), 1633–1639.

Oteros, J., Buters, J., Laven, G., Röseler, S., Wachter, R., Schmidt-Weber, C., & Hofmann, F. (2017). Errors in determining the flow rate of Hirst-type pollen traps. *Aerobiologia, 33*, 201–210. https://doi.org/10.1007/s10453-016-9467-x

Oteros, J., Pusch, G., Weichenmeier, I., Heimann, U., Möller, R., Röseler, S., & Buters, J. T. (2015). Automatic and online pollen monitoring. *International Archives of Allergy and Immunology, 167*(3), 158–166. https://doi.org/10.1159/000436968

Pawankar, R. (2014). Allergic diseases and asthma: A global public health concern and a call to action. *World Allergy Organization Journal, 7*(1), 1–3. https://doi.org/10.1186/1939-4551-7-12

Pöhlker, C., Huffman, J. A., Förster, J.-D., & Pöschl, U. (2013). Autofluorescence of atmospheric bioaerosols: Spectral fingerprints and taxonomic trends of pollen. *Atmospheric Measurement Techniques, 6*, 3369–3392. https://doi.org/10.5194/amt-6-3369-2013

Pöhlker, C., Huffman, J. A., & Pöschl, U. (2012). Autofluorescence of atmospheric bioaerosols-fluorescent biomolecules and potential interferences. *Atmospheric Measurement Techniques, 5*, 37–71. https://doi.org/10.5194/amt-5-37-2012

Šaulienė, I., Šukienė, L., Daunys, G., Valiulis, G., Vaitkevičius, L., Matavulj, P., & Sofiev, M. (2019). Automatic pollen recognition with the Rapid-E particle counter: The first-level procedure, experience and next steps. *Atmospheric Measurement Techniques, 12*, 3435–3452. https://doi.org/10.5194/amt-12-3435-2019

Sauvageat, E., Zeder, Y., Auderset, K., Calpini, B., Clot, B., Crouzy, B., & Vasilatou, K. (2020). Real-time pollen monitoring using digital holography. *Atmospheric Measurement Techniques, 13*(3), 1539–1550. https://doi.org/10.5194/amt-13-1539-2020

Schaefer, J., Milling, M., Schuller, B. W., Bauer, B., Brunner, J. O., Traidl-Hoffmann, C., & Damialis, A. (2021). Towards automatic airborne pollen monitoring: From commercial devices to operational by mitigating class-imbalance in a deep learning approach. *Science of the Total Environment, 796*, 148932. https://doi.org/10.1016/j.scitotenv.2021.148932

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105–6114). PMLR.

Tan, M., & Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. arXiv preprint arXiv:2104.00298.

Tešendić, D., Boberić Krstićev, D., Matavulj, P., Brdar, S., Panić, M., Minić, V., & Šikoparija, B. (2020). RealForAll: Real-time system for automatic detection of airborne pollen. *Enterprise Information Systems*. https://doi.org/10.1080/17517575.2020.1793391

Tummon, F., Arboledas, L. A., Bonini, M., Guinot, B., Hicke, M., Jacob, C., & Clot, B. (2021). The need for Pan-European automatic pollen and fungal spore monitoring: A stakeholder workshop position paper. *Clinical and Translational Allergy, 11*(3), e12015. https://doi.org/10.1002/clt2.12015

Valiulis, G., Šukienė, L., Vaitkevičius, L., Daunys, G., Sofiev, M., & Šaulienė, I. (2020). 2019–2020 woody plants pollen dataset from automatic particle detector in Šiauliai (1.2.0). Zenodo. https://doi.org/10.5281/zenodo.5576824

Valiulis, G., Šukienė, L., Vaitkevičius, L., Daunys, G., Sofiev, M., & Šaulienė, I. (2021). 2019–2020 herbaceous plants pollen dataset from automatic particle detector in Šiauliai (1.2.0). Zenodo. https://doi.org/10.5281/zenodo.5576879

Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Jiang, Z., & Yan, S. (2021a). Tokens-to-token vit: Training vision transformers from scratch on imagenet. arXiv preprint arXiv:2101.11986.

Yuan, L., Hou, Q., Jiang, Z., Feng, J., Yan, S. (2021b). Volo: Vision outlooker for visual recognition. arXiv preprint arXiv:2106.13112.