ORIGINAL PAPER

Pattern recognition methodologies for pollen grain image classification: a survey

Philipp Viertel[1] · Matthias König[1]

Abstract

In a large number of scientific areas, such as immunology, forensics, paleoecology, and archeology, the study of pollen, i.e., palynology, plays an important role: from tracking climate changes, studying allergies, to forensic investigations or honey origin analysis. Since the mid-nineties of the last century, the idea for an automated solution to the problem of pollen identification and classification was formulated and since then, several attempts and proposals have been made and presented, based on different technologies, in particular in the field of Computer Vision. However, as of 2021 microscopic analyses are performed mainly manually by highly trained specialists, although the capabilities of artificial intelligence, especially Deep Neural Networks, are steadily increasing. In this work, we analyzed various state-of-the-art research work concerning pollen detection and classification and compared their methods and results. The problems, such as data accessibility, different methods of Machine Learning, and the intended applicability of the proposed solutions are explored. We also identified crucial issues that require further work and research. Our work will provide a thorough view on the current state of the art, its issues, and possibilities for the future.

Keywords Machine learning · Deep learning · Pollen classification · Palynology · Survey

1 Introduction

In the area of Computer Vision (CV) and Pattern Recognition, the classification of microscopic images is a broad topic with a large number of possible applications in various fields. The advent of Machine Learning (ML) and especially Deep Learning (DL) is major drivers in this area, combined with the steadily increasing computational power, thus leveraging microscopic pollen analysis. The advantages of automated and Artificial Intelligence (AI)-based pollen detection and classification is manifold: In all areas of applications, it can reduce costs, expenditure of time, and increase accuracy. A successful deployment of a pollen classification system is based on three important factors, of which two are mandatory: the software, i.e., the method of pollen detection and classification, the data, and the hardware. The hardware aspect, however, is only important if a dedicated system (covering all the required steps, e.g., from the specimen slide input, image acquisition, and microscope or camera controlling) is desired for an all-in-solution with an additional focus on usability. If such a software is designed to be deployed on a dedicated system that works autonomously, the question of performance and portability becomes important as well.

Furthermore, the question of how to acquire quality images of pollen grains outside of a laboratory has to be considered as well. The process of pollen acquisition requires additional steps, e.g., creating a sediment from a honey sample or capturing airborne pollen; therefore, the entire system becomes more complex.

The analysis of microscopic images is mainly done by humans due to complicated requirements, the importance of safe results, and the visual variability of small but important details. An automated solution to the steady identification of pollen is demanded in multiple areas, especially in aerobiology, such as local weather services, due to allergies in the population, or the pollen composition in honey samples, which is required for a correct and valid product label. Despite economical reasons, research work strongly indicates that the number of pollen allergies will increase in the

✉ Philipp Viertel
  philipp.viertel@fh-bielefeld.de

  Matthias König
  matthias.koenig@fh-bielefeld.de

1 Bielefeld University of Applied Sciences Campus Minden, 32427 Minden, Germany

future, especially due to climate change [3]. The Autopollen program[1] emphasizes this importance and aims at addressing this issue. It started in 2018 and will run until 2022 and is aimed at establishing a standard encompassing the entire chain—from pollen observation to pollen analysis—by working interdisciplinary with various European partners.

Microscopic images of pollen grains can be processed with AI methods, especially ML, to identify and classify pollen classes. A major approach to tackle this problem is Deep Neural Networks (DNN), specifically Convolutional Neural Networks (CNN) [31], with the original concept dating back to as early as 1998 [34]. However, early DL attempts are much older and can be traced back to 1965 [57]. DNNs in the form of CNNs are multi-layer neural networks, used to recognize visual patterns in pixel-based images. They are widely used to classify objects, understand scenes, as well as segment images in a semantic fashion. Over the years, variations and improvements have been made to these networks, such as Fully Convolutional Networks (FCN) [61], and establishing best practice network architectures, such as VGG-16 and ResNet [62] [25]. However, more traditional ML approaches are used in microscopic pollen analysis as well: Support Vector Machines (SVM), Linear Discriminant Classifiers (LDC), etc. These methods also require Feature Engineering in comparison with DL methods. All these methods have their strengths and weaknesses and their designated use cases. In the case of pollen classification, a large variety of these methods are utilized, yet it is still up for debate, which method is most suitable for the task.

In the scope of this work, we elaborate the current state of the art in the following chapters, with regard to available data sets and methods, results, as well as a comparison of existing methods and their performances. In Sect. 2, we discuss two common ways of acquiring pollen samples; via extraction from substances—in our case melissopalynology [43]—and the collection of airborne pollen, where we use the typical Burkhard trap [27] as an example, which is still the most used pollen trap for allergy-related weather forecasts. In Sect. 3, we discuss pollen as biological entities and their morphology, which forms the foundation for many Feature Engineering methods. It will be shown that pollen grains are complex and diverse in their features and why certain issues in classifying them via CV methods are rooted in their inherent biological structure. The CV methods, i.e., classifiers and descriptors, that are most commonly used for pollen recognition will be briefly discussed and summarized in Sect. 4. In Sect. 5, we explore the majority of work that exists and contributes significantly to the purpose of automatic pollen recognition and classification at this juncture. In Sect. 6 we will address the availability of data sets, their accessibility

and features, and compare them with each other. The same is applied to the reviewed methods and results. Finally, based on our reviews and findings, we summarize our results in Sect. 7. We will give a recommendation concerning specific actions that should be done in order to improve the methods, from a software, hardware, and data perspective. We believe that our work contributes as a stepping stone for further research and evaluation into the field of automated pollen classification.

2 Pollen analysis

2.1 Prepared pollen samples

Pollen analysis of honey, melissopalynology [43], is a specialized discipline in the field of palynology, aimed at determining the pollen taxa via samples of honey. As of 2021, the process that is required to identify the pollen and its geographical origin is performed manually. Beekeepers as well as large industrial honey producers encounter problems labeling their products properly, due to the fact that the honey composition varies strongly and requires a professional pollen analysis. However, such an analysis is seldom done and producers often resort to using generic names, such as summer honey or spring honey. Specialized institutes offer pollen analysis as a commercial service which allow the producer to label their product with the correct name and additional information concerning allergies and geographical origin. These procedures are costly and time-consuming since each batch of honey yield can differ in its composition. The conventional analysis process requires observation and discrimination of the specific features by a highly trained palynologist. Research work, such as [21], propose methods to improve the manual process by defining minimal requirements for a pollen analysis, to reduce time and labor, and still achieve a high identification accuracy. Currently, the manual process is still the most accurate method and the only one which fulfills official standards and norms in certain countries, such as Germany. However, ML and DL methods are steadily increasing in quality as well as the available computational power, which could in sum already support the work of palynologists.

One gram of honey contains between 2,000 and 1 million pollen grains which can stem from more than 100 different plants. In a professional laboratory analysis, these pollen grains are counted and identified proportionally. This process is standardized in Germany by the norm DIN 10760, which we use as an example. First, a pollen preparation has to be created. For this purpose, 10g of honey is being diluted with 20ml of distilled water and then centrifugalized for ten minutes at $1,000g$. The supernatant liquid is removed and 20ml of distilled water is added again, to dissolve the sugar crystals completely. This is being centrifugalized again for

---

[1] URL: https://www.eumetnet.eu/activities/miscellaneous/current-activities-mi/autopollen/. Accessed January 3, 2022.

five minutes at $1,000g$. The pollen sediment is cleaned and put on a specimen slide with a pipette where it is dried on a heating plate at about 40°C and finally enclosed with glycerol gelatin under a cover slip. When the pollen preparation is solid, the identification with a light microscope (LM) is performed next, with a required magnification strength of 320 to 1,000 times. A highly trained palynologist is counting and identifying the pollen visually by their specific characteristics. The goal is ultimately to determine the relative frequency of the pollen taxa $X_p = \frac{A \times 100}{n}$, where p is the plant, A the number of pollen of the searched plant, and n the total number of counted pollen grains. It can also be helpful to remove the number of pollen from nectarless plants via $(n - n')$, where n' is the number of nectarless plants, $X_p = \frac{A \times 100}{(n - n')}$. It is necessary to count 500 to 1,000 pollen grains per honey sample.

According to the DIN 10760, 500 pollen grains are the minimum required number per sample. This requirement could be fulfilled by beekeepers, but it is highly dependent on the quality of the sediment process and the utilized equipment. Sediment creations with non-professional equipment (e.g., honey extractors) contain 120 to 600 pollen grains per 20μ sediment. Such a manual analysis comes with two major problems: Laboratory centrifuges are generally of superior quality and offer higher numbers of pollen per 20μ sediment, ranging from 4700 up to 6000 pollen grains. The second problem is the actual classification of the pollen. Also, a palynologist is well-trained in determining the pollen classes by detecting the morphological traits. Different sources are used to support the process of classifying the pollen, such as internet databases, e.g., PalDat[2], reference books [26] [24], and pollen calendars.

This DIN can be used as an exemplary instruction on how to analyze pollen extracted from honey samples. Apart from the biological and chemical processes, which require specific instruments, the actual identification process is not specified in terms of tools or equipment, as long as they fulfill the minimum requirements mentioned above. An international norm, ISO/TC 34/SC 19[3], is currently under development with the focus of honey products and bee pollen (by relying on, e.g., EU and Chinese established norms). It is important to watch these developments, since the requirements for a valid pollen analysis, concerning pollen numbers and method, can have a strong impact on automated pollen solutions. It could deem working applications invalid, due to regulations and requirements that are not met, e.g., better image quality or larger quantities of samples to ensure a more stable result and additional safety.

2.2 Airborne pollen samples

Airborne pollen grains are usually not classified on the spot, but collected via a pollen trap for, e.g., twenty-four hours. The tape, on which the pollen grains are stuck, is removed manually and analyzed in a laboratory by a palynologist. However, if the pollen is supposed to be classified fast and without human interference, the requirements for an automated solution are manifold. The system has to incorporate the necessary hardware to take in an air stream and collect the pollen in such a way that the system can capture images from pollen on a special film, detect and classify pollen on a required scale, and finally dispose them again so that the process can start anew. Ideally, this is all done autonomously, without human interaction. Apart from the software side, the hardware requirements constitute an engineering problem on its own.

The typical way of collecting pollen, which is done by meteorological services, is to employ Burkhard pollen traps, which are modeled by the Hirst principle [27]. Such a pollen trap is powered by an electric engine and continuously draws in an air flux. The pollen (as well as dust and small debris) is collected on a slowly turning barrel which contains a plastic film. It turns with a constant speed of 2mm per hour around its own axis, which results in 48mm in 24 hours. These 48mm equate to an air volume of 14.4m^3. The pollen trap can collect particles over one week until the film needs to be replaced. A laboratory associate has to manually remove the tape from the pollen trap and analyze it in its entirety under a LM. Due to the length of the tape, the result of the analysis gives an indicator for an 24-hour arithmetic mean of pollen per m^3 air. The most important pollen in this analysis is the seven most common ones that are allergenic. For example, the German Meteorological Service operates around 40 stations[4] nationwide. The evaluated data from the pollen are processed and the results communicated with the public.

All of the steps in this process require a lot of material, work, and time, especially between collecting the pollen and producing the actual result. However, proprietary systems have been developed and work autonomously, such as the BAA 500 by Helmut Hund GmbH, which is deployed in the e-PIN system[5]. This system is the world's first of its kind; however, the costs, weight, and size of these machines are a major drawback which limits their spread and use.

---

[2] URL: https://www.paldat.org/. Accessed January 3, 2022.

[3] URL: https://www.iso.org/committee/6716626.html. Accessed January 3, 2022.

[4] URL: http://www.pollenstiftung.de/pollenvorhersage/pollenmessstationen-in-deutschland/. Accessed January 3, 2022.

[5] URL: https://epin.lgl.bayern.de/messstandorte. Accessed January 3, 2022.

3 Palynology

The dominating object of interest in palynology among the palymorphs is the pollen grain. For terminology, we follow [26] and [24]. Pollen is a flour-like mass that are produced in the stamen, i.e., the reproductive organ of a flower, of spermatophytes, i.e., seed plants. The pollen grain is the carrier of the male gametes, i.e., sperm cells. These gametophytes make up an extra generation of the seed plant and consist of the sporoderm, two or three cells, and the pollen tube. Therefore, pollen grains are the male haploid counterpart of the diploid plant body. They carry the male genetic material and are very robust and resistant, a quality which makes them especially interesting for archeological and forensic studies. Due to their resistance to hostile environments and long life span, it is possible to, e.g., reconstruct ancient vegetation in the discipline of paleoecology or to collect and preserve specific evidence from crime scenes in the area of forensic palynology. The scientific exploration of pollen is also far from complete, due to the fact, that the male gametophytes are not fully investigated yet. From around 260,000 to 422,000 plant species only 10% have been investigated and their morphology remains unknown [58].

In order to understand the scientific names of pollen and the possible degree of classification, a quick overview of biological taxonomy is necessary. Plants and organisms in general are organized in the Linnaean taxonomy system, hierarchically organized in different levels that share physical characteristics and are related by common descent. The order of the levels is: Kingdom–Phylum–Class–Order–Family–Genus–Species. As an example, the scientific name for white mustard is *Sinapis alba*. The kingdom is *Plantae*, which is the sum of all plants that are living organisms, which perform photosynthesis and cannot move. The most common and more finer categories are: order, family, genus, and species. For *Sinapis alba*, the order is *Brassicales*, the family is *Brassicaceae*, and the genus is *Sinalpis*. *Sinapis alba*, the binomial, is composed of the genus and its specific botanical name.

It is common to categorize pollen in two groups: pollen type and pollen class. However, due to the importance of using correct terminology, the difference between class and type needs further explanation. Pollen type is generally used to categorize pollen by a specific combination of characteristics and affiliating it with a taxon. The pollen class on the other hand is a method to combine pollen by one or multiple characteristics, such as shape, and aperture type. Pollen classes are helpful in identifying key characteristics but have no systematic value, as a pollen grain could belong to multiple pollen classes. Due to the use of the term class in computer science and ML in general, we refer to class the same way as it is used in non-biological use cases of object classification, such as car, airplane, dog, cat.
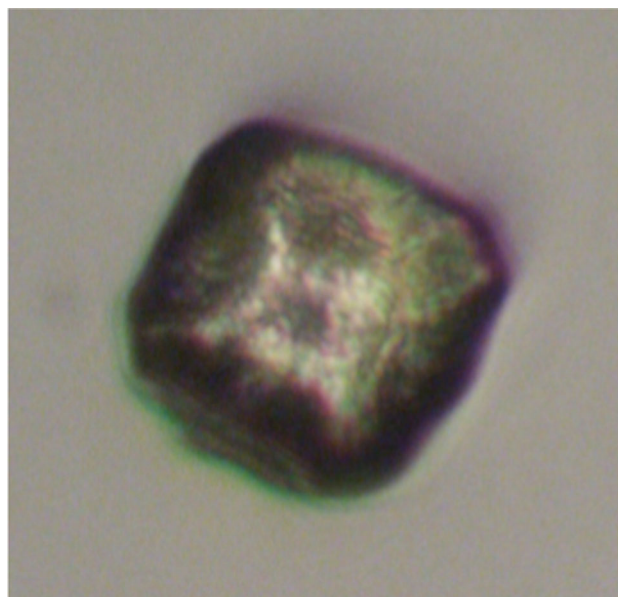

Fig. 1 *Alnus glutinosa* pollen. Medium size (26-50 μm). Pollen class: porate. Polarity: isopolar. P/E-ratio: oblate. Aperture number: 5. Aperture type: porus
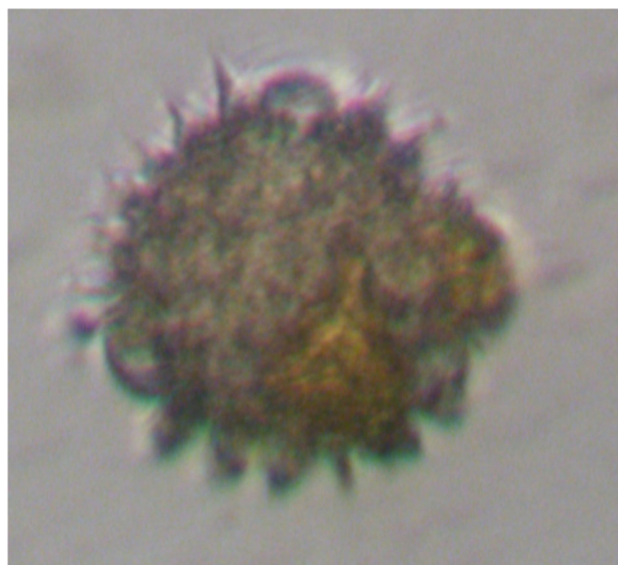

Fig. 2 *Helianthus annuus* pollen. Medium size (26-50 μm). Pollen class: colporate. Polarity: isopolar. P/E-ratio: prolate. Aperture number: 3. Aperture type: colporus

There are multiple parameters which describe the morphology of a pollen grain. These parameters can be used to model features that can be later used to detect and identify pollen types automatically via the process of Feature Engineering and Feature Matching, respectively. These parameters are: shape, size, number, position and type of apertures, and the pollen wall. These morphological features are utilized to make pollen specifiable and comparative.

3.1 Pollen morphology

Pollen has a spheroid shape, i.e., ellipsoid with two semidiameters. Therefore, the pollen shape is defined by the P/E-ratio; the ratio of the length of the polar axis P to the equatorial diameter E. If the polar axis is equal to the equatorial diameter, the pollen grain is spheroidal, i.e., isodiametric. If the polar axis is longer than the equatorial diameter, the pollen grain is described as prolate. In the last case, if the polar axis is shorter than the equatorial diameter, the pollen shape is labeled oblate.

The size of a pollen grain can vary from around $10\mu m$ to larger than $100\mu m$. To determine the size, the largest diameter of the pollen grain is used. However, the preparation method and the state of hydration have a large impact on the size. The following nomenclature is recommended: $<10\mu m$ very small, 10 to $25\mu m$ small, 26 to $50\mu m$ medium, 51 to $100\mu m$ large, and $>100\mu m$ very large.

In certain regions of the pollen wall, the outer layer of the pollen wall is missing, and instead there are openings positioned called apertures. These apertures function as the site of germination and can vary in their number. Not all pollen grains have apertures, therefore, the ones without are labeled inaperturate. The position of the apertures determines the correct terminology, e.g., a circular aperture, if positioned equatorially, is called porus. If it is positioned away from that, it is called ulcus. An extended aperture, positioned equatorially, is called colpus; if not positioned equatorially it is called sulcus. Poroid is the term for a circular aperture. Apertures can also occur in combination, such as colporus (porus and colpus) and the rare combination of colpi and colpori, called heteroaperturate. The number of apertures can vary as well, and a pollen grain with more than three apertures is called stephanoaperturate. The aperture features, number, type, and position, are fixed within a pollen species and only rarely vary, e.g., in stephanoaperturate pollen. Due to the setup of the apertures, there are a total of six different pollen views possible, depending on the pollen type: In monocots (with one equatorial aperture), there is the proximal polar, distal polar, and two different equatorial views. Dicots usually have one polar and one equatorial view.

The pollen wall, the sporoderm, is made up of two layers: the outer layer, exine, and the inner layer, intine. The exine consists of sporopollenins, which is responsible for the robustness and longevity of pollen grains due to the acetolysis and decay-resistant biopolymers. The intine is made up of cellulose and pectin. The wall, or surface of the pollen grain, can contain certain characteristics called sculpture and ornamentation.

Two examples and their descriptions are shown in Figs. 1 and 2.

An important, but for the correct classification of pollen difficult, characteristic is their ability to absorb and release
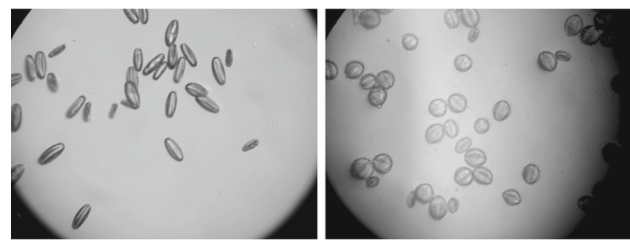
(a) Dry state.   **(b)** Hydrated state.

Fig. 3 Pollen grain of the plant *Phacelia tanacetifolia* shows completely different morphological features depending on their condition. Such aspects play an important role, especially when an automated pollen classification or counting system is deployed in real-life scenarios

water. Therefore, each pollen grain can have two different morphological states: hydrated and dry. The reason for this effect is to protect the male gametophyte against dehydration. This process is called harmomegathy. The change of the pollen shape is related to its morphological aspects, e.g., apertures. It is possible to attain the turgescent state of the pollen after it is dehydrated, by adding water again. However, this process cannot be indefinitely repeated. Pollen grains which have a thin sporoderm can be irreparably damaged in this procedure. The difference between hydrated and non-hydrated state can be seen in Fig. 3.

These morphological features of pollen grains only touch the surface of the topic. A highly trained palynologist is required to learn a lot more about the topic, in order to identify pollen correctly by visually means only. The case is especially difficult when it is required to classify different pollen, which look a lot alike in their features. A typical method to highlight the pollen characteristics is to add fuchsine, which tints the pollen grains in a pinkish tone and increases the visibility of the pollen features.

4 Computer vision

4.1 Descriptors

Methods that do not utilize DL techniques to learn features inherently require so-called descriptors which describe mathematically a feature that can help differentiate pollen types. These descriptors are formulated to detect certain characteristics in an image. The descriptors can be seen as mathematical translations of the features described in the previous section, morphological-based, or other texture or color-based features. Redondo et al. [48] (p. 15 et seq.) give an overview of the most common descriptors as well as more detailed explanations. For this purpose, we give a compact summary in Table 1. To give an example from [48], the shape can be defined as:

Table 1 Shortened version of the descriptor overview given in [48]

Category	Descriptor
Morphological	Area, Perimeter, Shape, Eccentricity, Fullness, Contour Profile
Statistical	First-Order, Second-Order Haralick
Transformed space	Local Binary Pattern
Moments	Hu
Space-frequency	Fourier, Wavelets, Gabor
Texture	Haralick

$$\text{Shape} = \frac{4 \cdot \pi \cdot \text{Area}}{\text{Perimeter}^2} \tag{1}$$

The shape indicates the elongation of the pollen grain, e.g., a value of 1 would indicate the pollen grain as a circle. This definition can also appear slightly different under the term compactness.

The list is not complete, especially due to the fact, that certain features are used with different terms by some authors, such as geometric features, which can also be morphological traits, e.g., region moments or boundary. Color, especially gray-level features (brightness), can be statistical features (first and second order). Implementation of features can also vary by their respective authors. [52] describes also a number of features, such as boundary moments and Fourier descriptors, as well as geometric ones.

4.2 Classifiers

4.2.1 Linear discriminant analysis

Linear discriminant analysis (LDA) is one of the most common methods used in pollen classification. It is used as a method for dimensionality reduction, which is especially important for bioinformatics, and/or as a linear classifier. Unlike Principal Component Analysis which is an unsupervised method, LDA requires labels to compute the linear discriminants that will maximize the separation between a set of classes. The separability between each class has to been calculated (between-class variance), i.e., the distance between the mean of the classes. This is followed by the within-class variance. Finally, Fisher's criterion [18] aims at maximizing the between-class variance and minimizing the within-class variance. To make predictions, the LDA estimates the probability that a new data sample belongs to a certain class by using Bayes' Theorem or Maximum likelihood.

LDAs can be seen as an improvement to Logistic Regressions (which can only handle two-class classification problems) and has a number of benefits: It is a simple prototype classifier, the decision boundary is linear, making them robust and fast, and the advantage of dimension reduction.

4.2.2 Support vector machines

A Support Vector Machine (SVM) algorithm tries to find a hyperplane in a N-dimensional space which classifies the data points. Although there are many possibilities for the hyperplane placement, the method aims at finding the optimal hyperplane (line) that separates the data points by maximizing the margin, which is the distance between the hyperplane and the support vectors (the points closest to the line from both classes). This allows the classification of future data points with more confidence. If the data are not linearly separable, certain kernel functions are used to transform, i.e., map, the data into a new space. These kernels are chosen depending on the specific problem.

4.3 Neural networks

4.3.1 Multilayer perceptron

Feedforward neural networks, or Multilayer Perceptron (MLP), are simple neural networks with at least one hidden layer. Such a network is made up of a large number of neurons which are organized into layers. The minimum number of required layers is three; one input layer[6], a hidden layer, and an output layer. Such a network is not really deep, and only the addition of further hidden layers, which can vary largely in number, is responsible for coining the term deep in regard to learning. For a classification task, such a network can be displayed as a function $y = f^*(x; \theta)$, where the input $\mathbf{x}$ is mapped to a category $\mathbf{y}$. θ is the parameter which is learned through the process. Since every neuron is basically a function, the network is made up of multiple connected functions. This can also be written as $f^{(1)}$ for indicating the major function of the first layer. All of the layers can be written in a function chain $f(x) = f^{(3)}(f^{(2)}(f^{(1)}(x)))$. The length of this chain indicates the depth of the model. In the actual training process, the function $f^*(x)$ approximates $f(x; \theta)$ with the training data as the input, adapting θ to get f as close as possible to f^*. Each training example $\mathbf{x}$ comes also with a

---

[6] The input layer can be excluded from being counted as a layer and simply be referred to as the input.

label **y**, which indicates what it resembles (through a numerical value). Therefore, the network must produce an output value that is close to **y**.

The neurons from each layer, input, hidden, as well as output layer, are each connected to the neurons of their neighboring layer. If a specific neuron in the layer B should be activated, the weights of the connections are crucial. The weights from each connection from layer A to the neuron in layer B must be set in such a way, that the desired neuron is actually activated. These weights $\theta_1, \ldots, \theta_n$ have numerical values. All the activated neurons from layer A are computed with their according weights $\theta_1 a_1 + \ldots + \theta_n a_n$. The resulting value of this weighted sum gets processed by another function, the activation function, e.g., sigmoid, or linear activation (no transformation) or typically Rectified Linear Units (ReLUs), which are most common in CNNs (due to minimizing the vanishing gradient problem). Activation functions are an abstraction of the rate of action potential, i.e., the firing of a neuron.

4.3.2 Convolutional neural networks

Convolutional Neural Networks (CNN), first introduced in [34], are widely used in DL applications. Since then, research increased the depth of neural networks [31] and developed architectures to fit their specific scenarios [32] [10]. Most state-of-the-art applications and solutions with the focus on RGB image object classification are based upon CNNs, due to their property to be most suited for handling tasks involving visual information. A CNN consists of several blocks, i.e., layer types, which will be described in short to give an understanding of how CNNs work and how they differ from MLPs. For more details, see also [23].

The network takes a RGB image containing a labeled object that the network aims at learning by generalizing its features, to predict further, unseen instances of the same class. The main part of the CNN is the name-giving convolutions. In this layer, specific filters (or kernels) are performing the convolution operation on the input. These kernels come with two hyperparameters, size and stride. Depending on the stride, these filters are applied on the input images and basically scan the image for structures, such as edges. This is done multiple times to detect more complex structures such as contours and object parts. This creates a set of feature images or feature maps (convolved features). The number of filters determines the number of filtered images, which can get very numerous and makes the following pooling layer necessary.

Pooling requires a window size (similar to the kernel) which is usually set up as 2×2 or 3×3 and a stride of 2 pixels usually. The pooling window moves in the according stride over every single convoluted image and chooses the maximum value (max pooling). This results in an image, that is smaller but contains still the same important information in a more concentrated and dense form by discarding unnecessary information.

The final layer in a CNN is the fully connected layer. This layer decides the actual output of the network, depending on the values of the filtered images, resulting from the last layer before the fully connected one, and produces the classification probabilities by activating the respective output neuron. The number of neurons is identical to the number of output classes in this layer.

An important feature in CNNs is Backpropagation to compute the gradient of the loss function. The weights are updated to minimize the loss (i.e., the deviation from the target output). Typical methods are gradient descent or stochastic gradient descent.

4.4 Metrics

Any pollen classification method or system is put to the test, i.e., it is evaluated preferably on a set of pollen images that is unknown to the model. In order to give an indicator of its performance, most methods usually use at least one of the following four metrics: Accuracy, Precision, Recall, and F_1.

Accuracy is the most common one and describes the fraction of predictions that the model classified correctly:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total numbers of predictions}} \quad (2)$$

This metric can also be called Correct Classification Rate (CCR) and is defined in the following form, with the number of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

Precision, however, indicates the proportion of correct positive identifications:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

Recall indicates the proportion of true positives that were identified correctly:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

F_1 is the combination of Precision and Recall via the harmonic mean:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

Although Recall, Precision, and F1 are intended for binary classification problems, they can be used for multi-class classification problems when averaged. The sum is calculated over all rows/columns of the confusion matrix, e.g., the average precision is the fraction of instances where the prediction i is correct out of all instances where i is predicted. If not noted elsewhere, results that are using any of these three metrics, are averaged and refer to multi-class classification.

5 Methods

The need and the benefits of an automated pollen recognition system were described as early as in 1996 [63]. Stillman and Flenley mention the time-consuming process of extraction and identification by manual preparation and analysis. The time of such an analysis is estimated with 2 to 10 hours. The process becomes more difficult if the identification is supposed to go below genus (e.g., family). An increase in speed would mean a decrease in costs and help palynologists in their work. Summarized, the authors name six requirements; the need for more sites, for fine resolution, for larger counts, for speed, for objectivity, and finer determination. The most important ones; speed, objectivity, and better determination, are the objectives that can be tackled effectively with the help of ML methods. The need for speed is obvious; however, the need for objectivity is necessary, due to the fact that experts can err and even groups of experts are not always of the same opinion in their examination. This problem can possibly be tackled by training a ML model on an adequate amount of data, so that even the finest distinctions in pollen grains can be learned. In particular cases, however, additional methods could be useful, such as research into synthetical generation of (more rare) pollen grain images [64] or an additional knowledge system (based on geography, season, and likelihood). The need for finer determination is related to the previous factor; in terms of taxonomy, it is useful to determine a pollen grain (i.e., the flower) beyond the level of family. All of these factors weigh more or less depending on the desired use case.

That an automated pollen analysis is feasible and can produce real benefits was already shown in [12]. With simple image processing techniques, the authors compared the pollen counting speed with human analysts. The method produced results per image in 60 seconds, while the human eye required 5 up to 68 minutes.

Apart from methods, such as fluorescence spectrometers, i.e., UV lasers, ([47] [29]) which can be found in proprietary solutions[7] to identify pollen, the available methods in CV can be placed in two categories: manual feature engineering,

including features based on texture, color, morphology of pollen grains as well as feature extraction and classification via DL, where points of interest are extracted by a neural network. Since the focus of this work is on the CV approach, techniques such as UV lasers are ignored.

Feature-Engineering-based methods can be split further up. Morphological ones, which use features, such as shape, geometry, as described in Sect. 3.1 or texture-based methods, which utilize the specific textural characteristics of pollen types. Hybrid methods use a mixture of different features (morphological, textural, statistical features, etc.), while DL methods learn features from a set of training data on their own. However, the process of how features in DL were generated is not comprehensible by humans in their entirety anymore.

The task of classifying pollen can require an additional step. Depending on the method that is used, the problem of acquiring an adequate image of a pollen grain is the first issue to begin with it. This defines especially the framework of how the experiments and tests are performed and how the proposed solution is intended to work in real-life scenarios. The scenarios can vary, e.g., the data set of pollen images can come already segmented, i.e., that each image contains only one pollen, while others require segmentation of the pollen grains from other particles (e.g., dust) [56]. Therefore, it is important to keep the applicability in mind and the intended use case.

A summary of all evaluated methods is given in Table 2. In the following sections, we will discuss all proposed methods in detail.

5.1 Texture-based methods

An early example of texture-based methods is [33], where the authors worked with 192 scanning electron microscope (SEM) images of six pollen taxa. The texture analysis was performed specifically on the exine of the pollen, and for that purpose a gray-tone spatial dependence analysis was used. The resulting co-occurrence matrix contains information regarding the gray levels of the images. From this, it is possible to deduce certain features (angular second moment and contrast). Fisher linear discriminant functions were used as a classifier, achieving an accuracy of over 94%.

The authors of [8] used a method to discriminate the genus of pollen based on their texture. The proposed method works with five different pollen types accompanying a data set of 20 pollen loads. Regions Of Interest (ROI) are extracted from the original images and with the combination of texture filtering methods and feature selector FSM (floating search method) a classification accuracy of about 87.4% was achieved. The proposed solution has one drawback: If the number of plant species is increased, the system performs in such a way, that it is not applicable anymore.

---

[7] URL: https://swisens.ch/en_uk/swisens-poleno/. Accessed January 3, 2022.

Table 2 Overview of current state-of-the-art methods of classifying pollen grains by descriptors (features), classifiers (classification methods), and accuracy. Notice that only the first two publications used the same data set and allow a decent comparison. The other publications are ordered chronologically, from oldest to newest. The acronyms are, if not stated elsewhere, the following: Logistic Linear Classifier (LLC), Local Linear Transforms (LLT), Multi-Layer Perceptron (MLP), Minimum Distance Classifier (MDC), Support Vector Data Description (SVDD), k-nearest neighbors (kNN), Color, Shape, and Texture (CST), Bag of Visual Words (BOW), C-Support Vector Classification (C-SVC), Radial Basis Function SVM (RBF SVM), Local Binary Pattern (LBP), and Histogram of Oriented Gradient (HOG)

Work	Pollen taxa	Descriptor	Classifier	Accuracy
[22]	23	BOW, color, shape and texture	C-SVC (SVM), CST+BOW	64%
[59]	23	Automatic feature extraction	Linear discriminant classifier, CNN	97%
[33]	6	Texture analysis	Multivariate statistical classification	94%
[35]	4	Texture features	NN, leave-one-out classifier	100%
[7]	30	Color, shape features	Mahalanobis distance	77%
[5]	30	Color, shape, geometric features	Mahalanobis distance	88.25%
[39]	3	Morphometric features	Regression trees	n.a.
[53]	26	Invariant gray-scale features (3D)	SVM	92%
[36]	13, 4	Texture/shape	Linear discriminant analysis	100%
[52]	3	Fourier descriptors	MDC	90%
[8]	5	Texture features (FSM)	MDC (majority voting)	85%, 87.4%
[49]	26, 7	Group integration	Nearest Neighbor, SVM	96.9%, 99.7% (SVM)
[9]	3	Shape, stat. gray-level, pore/colpus	Linear normal classifier	97.2%
[51]	3	Texture and shape features	SVM/MDC/MLP	89%
[46]	3	Invariant features (local jets)	Bayesian classifier	83%
[67]	7	Texture features, LLT	Adaptive Bayesian Combination, LLC	90.58%
[1]	3	Geometric, shape, texture features	MLP	90%
[66]	26, 33	3D discrete spherical features	SVM	96.3%, 91.8%
[28]	3, 40	Color, texture, optical spatial frequency	Mahalanobis distance	77%
[11]	5	Morphological, shape, textural, color	KNN, Gaussian, SVDD	92.3%
[48]	15	Morphological, statistical, three space-frequency	Linear discriminant analysis	99.4%
[14]	30	Automatic feature extraction	CNN	90%
[2]	23	Color, texture	SVM, Random Forest, Logistic regression	79%
[15]	10	Automatic feature extraction	CNN+RNN	100%
[30]	5, 11	Automatic feature extraction	CNN	99.8%, 95.9%
[20]	11	Automatic feature extraction	CNN	99.75%
[65]	27, 33, 28	Local key points (3D)	SIFT	88.25%
[41]	6	Geometric, textural	Random Forest	88.24%
[60]	46	Automatic feature extraction	CNN + LDC	97.86%
[4]	4	LBP, HOG, Automatic feature extraction	RBF SVM, CNN	87%, 90%

[36] conducted three experiments, all based on textural information: texture variables with shape analysis to classify four pollen taxa (originating from the New Zealand flora) with LDA, texture and shape features via a neural network classifying 13 pollen taxa, and texture alone on the same 13 pollen taxa. The authors used 18 samples for each 13 pollen taxon, and the average number of images for training was 54%. In all three experiments, a classification rate of 100% was achieved.

[67] apply co-training, which is used to train two separate classifiers, in which unlabeled data are being labeled and combined with the other classifier. This iterative algorithm can be a suitable method, if the data lack labels, e.g., due to complexity or nescience. To classify pollen images, the authors utilize two major features: Haralick's texture features and Local Linear Transforms. The classifier for the co-training is a Logistic Linear Classifier. The data set used consists of seven pollen types with 196 images on average per class (50% split into training and testing). With the adaptive Bayesian combination method an accuracy of 90.58% at best with 686 training samples was achieved.

5.2 3D-feature-based methods

Ronneberger et al. [53] propose a general-purpose object recognition system that uses a 3D volume data set created with a confocal laser scanning microscope and is used to classify airborne pollen. The system is trained on a data set containing 26 pollen taxa with 385 samples and was evaluated on 15 samples per class. With the confocal laser scanning method a 3D image can be obtained by stacking up multiple 2D image recorded from different focus planes. Utilizing 14 gray-scale invariant features and SVMs for the classification process, the authors achieved an accuracy of 92% and for exclusively allergy-related pollen 97.4%.

[66] proposes Discrete Spherical Fourier Descriptors (DSFD). The surface curvature voxels of pollen are extracted and decomposed into radial and angular components (via Spherical Harmonic Transform). A discrete Fourier transformation is applied to obtain the 3D descriptors used for pollen recognition. A classification is performed by a SVM. The authors used two data sets for the evaluation. One with 389 pollen grains of 25 different taxa (created under ideal conditions) and another one with 22750 pollen grains from 33 pollen taxa (Pollenmonitor). On the first data set a CCR of 96.3% was achieved and on the second data set a CRR rate of 91.8%.

A different approach is done by the authors of [65]. Instead of relying on descriptors for specific features the authors utilize the SIFT method [40] on 3D pollen images. The approach consists of four steps; key points (which are scale-invariant) are found via local differential vectors, which yield the local maximum points of the gradients. The image is divided into blocks for each layer of the 3D Gaussian image pyramid to extract the positive and negative differential vectors. In the next step, local key points are obtained. Lastly, to solve the problem of rotation the authors propose a rotation invariant feature transform method which uses a 3D rotation matrix, based on alterations made to the original SIFT method. This produces a vector histogram descriptor, describing the statistical distribution of the gradient vectors for the key points. The experiments were performed on three different data sets: Confocal [54], Pollenmonitor [46], and CHMonitor [65] (taking 25% random images per category as training images) achieving an average precision of 88.25% (over all data sets).

[49] uses Group Integration, which is a method to generate invariant features. The authors added two features to the Group Integration method: Local directional information and Spherical Harmonic Expansion for more descriptive features, as well as an algorithm using importance sampling. The 3D volume data consist of 26 German pollen taxa with 385 samples. A compressed data set of 7 classes was created by maintaining the allergologically relevant classes and merging the irrelevant ones. Using a Nearest-Neighbor Classifier and a SVM with Histogram Intersection Kernel an accuracy of 94.5% and 96.9% was achieved on the entire 26 classes set and 97.4% and 99.7% on the combined 7 classes set.

In [20], the authors use a CNN to identify pollen on airborne pollen slides. The proposed system does not require any pre-processing and is trained on a set of 251 videos (i.e., 386 samples with 3375 fully visible grains, as the system was only trained with such, any partially visible grains were ignored); therefore, the network learns 3D information of the pollen due to capturing the various focal planes. A second set of 135 videos (1234 pollen grains from 11 pollen types) are used to evaluate the method. The training and test data are split 60%-40% (251 samples with 2037 grains and 135 samples with 1234 grains, respectively). The authors manage to achieve a recall value of 98.54% and 99.75% precision. The model which was deployed in this work is Faster R-CNN [50] with a feature pyramid network for training and RetinaNet [37] for evaluation with only slight alterations.

5.3 Full-solution methods

[1] is one of the few examples, where the authors attempt to provide a full solution, i.e., develop a method for pollen classification embedded in a larger system, that performs the entire process and takes the image acquisition and the hardware aspect into consideration. Therefore, the system contains multiple parts: two microscopes (one to obtain a wide view for pollen identification and a second one to inspect the location candidates), lighting, and movements. The image processing consists of auto-focus algorithms, segmentation algorithms, and classification algorithms. To perform the classification a MLP is used. In total, 43 image

features (size, shape, and texture) are extracted from each image and used as an input to the network. Utilizing three pollen types, a classification rate of 90% was achieved. The system was presented in its entirety again in [28]. The image processing part utilized here is based upon [68]. The method uses Z-stacks, i.e., multiple images are taken at different focus levels. That allows a high visibility of vertical details of the pollen grain. The best portions of the images are combined into one single image. The feature set of 43 different image features is based on texture, shape, and spatial frequency. A neural network is utilized for classification, and when tested on conventional microscope images, an accuracy of 94% was achieved.[8]

M. Chica [11] developed an image processing and classification system to detect fraudulent pollen. LM images from five different pollen types were selected: *Echium*, *Cistus*, *Rubus*, *Olea*, and *Quercus* (five of the most common pollen types in Spain), as well as one class for outliers. A digital camera (USB DS-Fi1) is used to acquire the images from the microscope, a Nikon E200 (40x), at a resolution of 2560×1920 pixels. Before the actual classification, the pollen grain is segmented from the background via a group of image processing methods: histogram equalization to enhance the contrast and a median filter to remove unwanted noise. It is important to notice that this entire process does not take non-pollen objects into consideration, which could be in, e.g., airborne pollen or other polluted pollen samples. After the segmentation, a set of discriminative features is calculated for each pollen type. The result is 28 features per class (i.e., pollen type) which can be split in three categories: shape-related features, textural and color information, and exine descriptors. For each individual pollen type, a one-class classifier is trained and formed into a multi-classifier. The number of samples for each pollen type varies from 101 to 446 (1063 in total) and is split into a training and a test set of 80% and 20%, respectively. The best result of 92.3% accuracy was achieved with the one-class k nearest neighbor (kNN) method.

The work of [5] and specifically [6] and [7] proposes a semiautomatic pollen recognition system. This system is composed of two modules: pollen slide analysis and recognition. It takes aerobiological slides that are colored with fuchsine[9]. To capture images, the system is equipped with a LM with a 60x lens, a color camera with a framegrabber card, and a micro-positioning device for placing the slide properly under the microscope. The authors identified two major issues: First, if the system operates autonomously, it

has to adjust the focus to acquire a proper image. The authors solved this issue by developing an algorithm that adjusts the focus on a sharpness criterion, which finds the best focusing position for the current sample. The second issue describes the problem of pollen grain detection in the scene. Fuchsine is helpful here, since pollen is sensitive to the colorant, e.g., dust particles can be separated. However, there are other particles that can be sensitive to it as well. Therefore, an additional algorithm based on Markovian relaxation to localize the pollen was conceived. The authors achieved a localization rate of 90% during the image acquisition. The localized pollen is then captured from multiple foci to identify the pollen by its characteristics that become visible via its constructed 3D shape. This process includes general and specific pollen knowledge, derived from pollen morphology, to create distinct features (e.g., compactness or convex hull area) for classification. A recognition rate of 73% to 77% was achieved by evaluating on a reference database of 350 pollen grains from 30 different types (via leave-one-out validation).

5.4 Hybrid methods

Hybrid methods utilize a mixture of different features, including more abstract features such as statistical gray-level features. Therefore, these approaches are more generalized and can utilize atypical classifiers, e.g., elaborated in [46], with rare examples such as [39], utilizing regression trees. The authors of [46] use a system to classify biological particles with a focus on classifying particles found in human urine. However, the method was also trained and tested on a set of airborne pollen. In the first step, patches that include most likely particles are detected by a set of filtering methods resulting in a bounding box around a particle. Next, an invariant feature vector is computed from the image. A Bayesian classifier is used to learn and distinguish the features. The pollen data set contains 1429 images containing 3686 pollen grains. Three categories were used, and a correct classification rate of 83% was achieved.

The authors of [9] used a linear normal classifier and three different features: shape, statistical gray-level, and pore and colpus features. The data used were categorized into three groups: grass, birch, and mugwort pollen for allergy-related purposes. The data consist of 245 isolated pollen grain images (79 grass, 79 birch, and 96 mugwort). Employing cross-validation for evaluation the linear normal classifier (with 12 features) performed best among all other tested classifiers (e.g., nearest mean classifier, quadratic normal classifier). An average accuracy of 97.2% was achieved.

[51] combines morphological features (shape) and texture features (sculpture, i.e., texture of the grain) to detect, count, and classify pollen grains. The system was evaluated on three different pollen taxa of the *Urticaceae* family, with 9, 10, and 6 images each. Each image contains around two to 16 pollen

---

[8] A dedicated data set by [19] was used; however, this data set is as of 2021 not accessible anymore.

[9] *Rosaniline hydrochloride* is a magenta dye that is often used to color pollen grain samples to highlight morphological and textural features, e.g., for preparations in combination with glycerin gelatin.

grains. Since the pollen used here belongs to a family that has a specific circular morphological characteristic, a Hough transformation (HT) is used to detect circles in the images, followed by the extraction of the pollen silhouettes by applying the Snake-Contour approach and calculating shape and texture features. As a classifier a MDC (Minimum Distance Classifier), MLP (Multi-Layer Perceptron) and a SVM were evaluated, achieving a best average accuracy of 89%.

[52] uses brightness and shape descriptors (e.g., geometric features, region moments) based on morphological attributes to classify three pollen taxa (98 *Parietaria Judaica*, 100 *Urtica membranacea*, and 93 *Urtica Urens* pollen grains from a total of 77 images). After pollen extraction the feature sets are computed for each grain which the classifier uses to label the pollen via a minimum distance classifier (and majority voting). The Fourier descriptors achieved the best result of 90% correct classification rate.

The authors of [41] use geometric and textural features to describe pollen classes. A self-created data set with 6 classes and 584 images in total (90 to 100 images per class) was used to evaluate the proposed method. For each pollen grain image, a set of 98 features were generated. To reduce the number of features the feature importance was calculated using the Gini Index. Using the Random Forest method with a nested cross-validation (for feature selection) a mean accuracy of 88.24% was achieved.

In [45] a group of semiautomatic pollen extraction methods are evaluated. After extraction and pre-processing the authors present a set of 24 geometric- and 26 texture-based parameters which they used to train a MLP for feature detection. The pollen classificator actually implements 11 different networks to complement the individual results and to balance produced errors. The data set used consists of 345 images of 17 different pollen classes. The samples per class are uneven, ranging from 15 to 47. The best results achieved were between 90.54 and 92.81%. However, the scope is limited to pollen extraction and the problem of automatically detecting pollen grains from samples is not addressed.

The authors of [48] also use a variety of features and an additional descriptor for contour inner segmentation. To test the proposed method, a data set of 15 pollen types with 120 samples per type was captured. A large number of different feature descriptors were used in various combinations with the added contour descriptor (in total the number of descriptors is 6320) to experiment with the data. The large number of experiments showed the efficiency of the various descriptors and classifier methods as well as the improvements that the new contour descriptor can yield. The descriptors can be categorized as such: 6 Morphological (Area, Perimeter, Shape, Eccentricity, Fullness, Contour Profile), Statistical (13 first-order and 241 second-order Haralick (co-occurrence matrix)), 4 Local Binary Patterns (mean, variance, asymmetry, kurtosis), 7 Hu Moments and Space-

frequency descriptors (Fourier, Wavelets, Gabor, each 964 (241x4)). Three different classifiers were used: Fisher classifier, Support Vector Machines (SVM), and Random Forest. This requires a large quantity of experiments, whereas the authors admit that the results, although often very high, cannot be compared to other studies, due to, e.g., feature vector discrepancies. The problem of comparability will be addressed in detail in Sect. 6. However, the work introduces a novel descriptor, the Contour Profile, which describes microstructures in the pollen grain. Due to the morphological structure of the exine of certain pollen, the variance of gray levels at the contour of these pollen is higher than those pollen without a reticular surface. On its own, this descriptor is not very effective, but in addition to other descriptors it can raise the classification accuracy by 50%, with the best result being 99.4% by using a combination of morphological, statistical, and three space-frequency descriptors.

In [22], the authors created a data set of 805 pollen images in total of 23 different pollen types from the Brazilian Savannah called POLEN23E. For each pollen type, there are 35 images captured at different angles. Three feature extractors (Bag of Visual Words (BOW), Color, Shape, and Texture (CST), as well as a combination of both) and ML techniques (two types of Support Vector Machines (SVM), decision tree, and k-nearest neighbors (kNN)) were tested on the data set. For the evaluation, a threefold randomized cross-validation method was applied. The best result of 64% Correct Classification Rate (CCR) was achieved with C-SVC (SVM) and CST+BOW.

[2] utilizes low-level features, such as color and texture. A set of descriptors is used: Gray-level Co-Occurrence Matrices (GLCM), Local Binary Patterns (LBP), Auto Color Correlograms (ACC), and Weber's Local Descriptor (WLD). The data sets (Duller's data set [16] and POLEN23E [22]) were split 80% and 20%, for training and testing, respectively, with a fivefold cross-validation. The authors used three different classifiers: SVM, Random Forests, and Logistic Regression. On Duller's data set, the authors achieved 96% accuracy using GLCM and Logistic Regression. The ensemble of classifiers also yielded 96% and did not perform better, which is an interesting fact. On the POLEN23E data set, the highest achieved accuracy is 74%, with the following combinations; LBP + SVM, ACC + Logistic Regression, and WLD + SVM, with Random Forest performing the worst. The combination of all descriptors (ensemble rule, i.e., majority rule) manages to achieve an accuracy of 79%.

5.5 Deep learning methods

In one of the earliest attempts, as early as 1999, the authors of [35] use a neural network to determine pollen grains. The authors used four New Zealand pollen types with 18 sample images each (total of 72 texture images). The authors picked

pollen taxa that can be easily differentiated by their shapes and utilized a 5-4-4-3 feedforward Multi-Layer Perceptron (MLP). An accuracy of 100% was achieved.

The authors of [59] work on the same data set as [22] and improve the results by utilizing Convolutional Neural Networks (CNN). To evaluate their model accuracy, the authors used a 10-fold cross-validation and achieved a CCR rate of 97%.

The authors of [14] used a CNN with seven layers and used two different data sets: one LM and one scanning electron microscope (SEM). The LM set has 1,000 images but was artificially increased via image transformation methods to 14,000 samples. A second architecture was used based on the ImageNet model [31] and trained via transfer learning. A classification rate of 90% was achieved on the LM set (and 94% on the SEM set). An average precision and F score of 92% and 89% were achieved. Two years later, the same authors worked with sequential images and utilized z-stacking (i.e., multifocal images) [15]. Two different networks were used: a CNN and a Recurrent Neural Network (RNN). The CNN learns the visual characteristics, while the RNN (with 512 units of long-short-term memory (LSTM)) is necessary to establish the sequential information from the multifocal z-stack, due to the specific feature of RNNs of taking temporal information into consideration. The CNN is based on the VGG-16 architecture and was fine-tuned via transfer learning with 392 z-stacked sequences of 10 different pollen types, containing 10 images per sequence. In total 2940 images for training and 980 images for evaluation were used. The authors achieved a classification accuracy of 100%.

[30] provides another attempt at utilizing DL methods to classify pollen. The authors manually created a data set of 11 plant species pollen with 1774 images in total. Instead of utilizing an existing network (via transfer learning), the authors created their own network and gave an overview of its architecture, enabling others to recreate their work in theory. The data were also augmented by shifting and rotating the images to achieve an average number of 200 images per class. The data are organized in three sets: one containing five, one containing nine, and another one containing 11 classes. On the 5 class set an accuracy of 99.8% was achieved; however, on the 11 class set an accuracy of 95.9% was achieved, a drop by 3.9% in accuracy by adding 6 more classes. As a reason, the authors mention the morphological similarities of a large number of pollen.

[4] introduce their own data set, POLLEN13K, as well as a benchmark on the data with various ML methods. The data set contains around 13,000 images of four different pollen taxa and an additional class for debris, bubbles, and other non-pollen objects. The authors evaluate a number of algorithms on the data set, based on two different feature sets: HOG [13] and LBP [44]. A multitude of methods was evaluated—linear SVM, RBF SVM, Random Forest, Adaboost, MLP—each

with LBP and HOG features. The best accuracy of 87% was achieved with HOG and RBF SVM. When using CNNs, the best accuracy was achieved with AlexNet [31] (30 epochs training) of 90% and a smaller VGG-net (20 epochs training) of 90% as well.

In one of the newest works [60], the authors use the microscope system (now known as Classifynder) as described in [28]. However, the authors used the system only to create the image set used in their work, which consists of 19,500 images from 46 different pollen types (with varying numbers of images per taxon by 40 to 1700)[10]. The DL model AlexNet [31] was used to automatically extract features. After the transfer learning process, the features are extracted and fed into a LDC. For evaluation the data set was split into 90% training and 10% validation and utilizing a 10-fold cross-validation. On the validation set, a CCR of 97.86% was achieved and a precision of 0.979.

6 Evaluation

6.1 Data sets

The research of the current state of the art makes one strong deficit clear: the lack of a high-quality benchmark data set. The large majority of works evaluate their method either on unpublished data or on data sets that are as of 2021 not available anymore. Only a few works evaluate their method on the same data set (e.g., POLEN23E [22] or Pollenmonitor [46][55]), which allows a meaningful comparison of the proposed methods. Under the current circumstances it is difficult to compare the methods and make general assumptions. For object classification in real-life scenarios, e.g., RGB images with labeled objects such as chairs, planes, etc., benchmark data sets are a longtime standard, e.g., the PASCAL VOC data set [17] or MS COCO [38]. The rapid success in these areas can be attributed, among other aspects, to the availability of such large data sets that are used in various works for training and evaluation. Therefore, proposed pollen classification solutions should be evaluated on uniform data sets that are accessible to the scientific community. The specifics of a data set (e.g., the geographical origin) could be overcome as well, by adding a large variety of pollen covering all major geographical habitats. If the pollen taxa get reduced to, e.g., the number of allergy relevant pollen, a worldwide data set could be established and be helpful not just for automatic pollen classification. The recent data sets POLLEN13K [4] and the New Zealand pollen set [60] are steps in the right direction. Especially the latter, due to its large number of pollen taxa,

---

[10] The data set is available at: https://figshare.com/articles/New_Zealand_pollen/12370307. Accessed January 3, 2022.

Table 3 Comparison between the most common used pollen data sets. Although most works use their own individual data sets, these data sets go by a specific name. As of 2020, the POLLEN13K as well as POLEN23E are easily accessible, whereas Pollenmonitor, Confocal, and especially CHMonitor remain not readily accessible

Data set	No. of images	Classes	Availability	Info
POLEN23E [22]	805	23	Yes	Brazilian Savannah
POLLEN13K [4]	>12,000	4 (+ debris)	Yes	Corylus avellana (well-developed + anomalous), Alnus, Cuprissaceae
Duller et al. [16]	630	7	n/a	*Plantago lanceolata, Quercus robor, Alnus glutenosa, Polypodium vulgare, Rumex acetosella, Conopodium majus, Dactylis glomerata*
Confocal [54]	385	26	n/a	26 most relevant German pollen taxa
Pollenmonitor [46]	22,750	33	n/a	Online Pollenmonitor
CHMonitor [65]	n/a	28	n/a	Real-time sampling image data set from China
New Zealand set (NZ) [60]	19,500	46	Yes	New Zealand and Pacific region
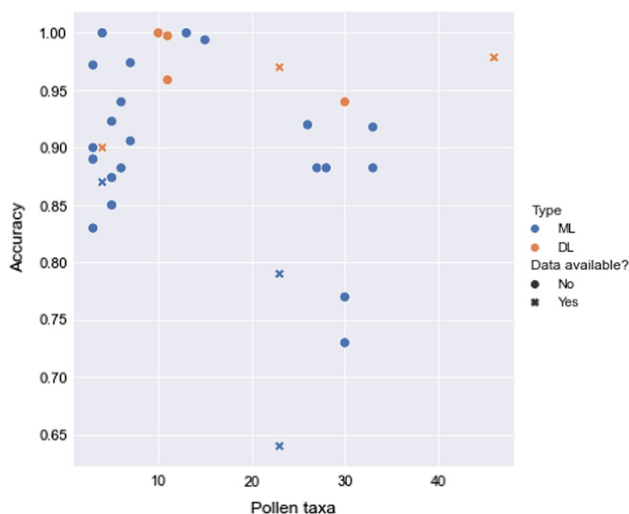
Fig. 4 Distribution of the evaluated methods by the number of pollen taxa (trained and tested on) and the best achieved accuracy. It is differentiated by the method type, ML with Feature Engineering and DL with automatic feature extraction. A cross indicates that the data set is available online

Fig. 5 Distribution of achieved results by data sets. Orange points indicate that the data set is available online. Multiple points do not have to resemble differing authors, but can also indicate multiple methods by the same authors

could be used as a benchmark data set. However, its applicability is of course limited by its regional focus.

An overview of the data sets is shown in Table 3.

6.2 Methods and verifiability

The question of verifiability is a difficult one and depends on multiple factors, which we will elaborate one by one. Deep Neural Networks in general have a reputation of being nontransparent due to their intrinsic feature extraction ability. Additionally, the training phase of a neural network (together with the hyperparameter tuning) is critical in the accuracy of a net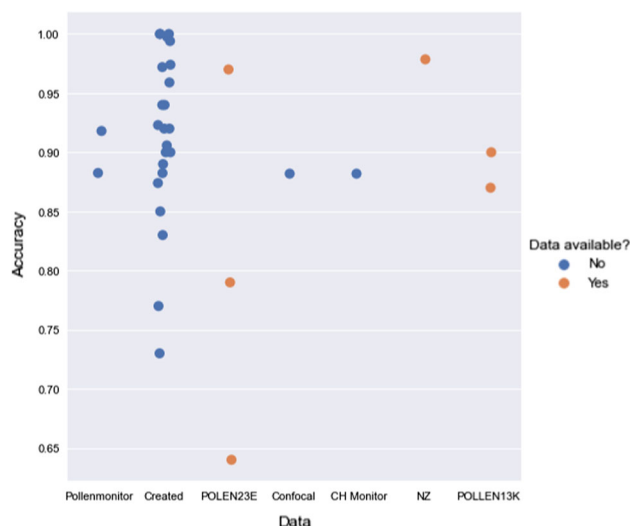work and can decide the final results drastically. The training of, e.g., 20 epochs is never the same due to the stochastic nature of most ML algorithms. Hardware, e.g., the CPU type, and the software version have an impact due to rounding differences. This makes it difficult, if a proposed and trained network is not available, to recreate the exact results. The lack of important hyperparameter information is contributing to that as well.

Most research work is trained and evaluated on the authors exclusive data, which are often not publicly available. This is a large problem when one has to decide which method works most efficiently. It is not certain that a specific method can be transferred from the authors scenario into a different use case scenario. The method could be too data-dependent or

specifically fine-tuned to that very issue. This is a problem of inner-work comparability.

Of all the reviewed works, only six work on data sets that have a distinct name and are, or used to be, available to verify. However, of these six data sets, only three are accessible via download, POLEN23E, POLLEN13K, and the New Zealand Pollen data set. Since POLLEN13K and the New Zealand data have been released 2020, new research work utilizing these images has not been produced yet, thus making the POLEN23E data set the only one where more than one author evaluated their proposed method.

The summary of our findings is visualized in Figs. 4 and 5. As mentioned, most research work is done on data sets that cannot be validated. A few data sets, such as POLEN23E, POLLEN13K, and the New Zealand pollen set, are available online and it is probably only a question of time, until further work is performed on these data sets. When the achieved results are compared with each other in Fig. 4, we notice a large number of results by ML methods were achieved with less than 10 pollen taxa, and a varying accuracy between 82% and 100%. However, an accuracy of more than 97% was achieved with 46 pollen taxa by utilizing DL methods on an online available data set, while the large majority of the ML methods utilize proprietary data sets.

When we look at two results more closely, that is the work of [22] and [59], which both utilized the POLEN23E data set, we can compare individual results. We will refer to [22] as the ML method with Feature Engineering and [59] as the DL method for better readability.

For all 23 classes, the DL method evaluates 30 samples. Six classes are 100% correctly classified with the remainder ranging from 24 to 29 correctly classified samples. The two worst predictions were *Matayba*, of which four were falsely classified as *Eucalyptus* and two as *Arrabidaea*, as well as *Myrcia*, which was three times falsely classified as *Faramea* and two times as *Protium*.

The ML method used 33 samples per class. The five worst results were 17 (out of 33) for *Hyptis*, 13 for *Faramea*, 15 for *Myrcia*, 13 for *Qualea*, and 17 for *Urochloa*.

Although the ML method achieved 20 correct classifications out of 33 for *Matayba*, and therefore performing just slightly worse than the DL method, the issue with correctly classifying *Myrcia* is overlapping in both methods. *Myrcia* is a genus that consists of around 770 species. The images for this class contain a large variety of optical characteristics, which makes it difficult to establish a set of features, that are valid for each perspective in every case. A few examples of its variety are shown in Fig. 6. However, due to the small amount of data, the DL method has problems and probably cannot establish a proper generalization. Especially when compared to *Eucalyptus*, we can see significant similarities that make a distinct classification problematic. An example is shown in Fig. 7. Static Feature Engineering can reach its limits here,
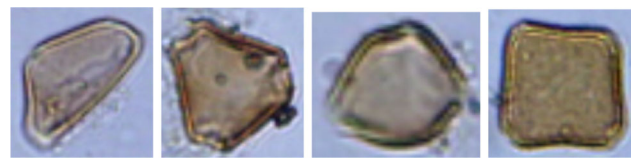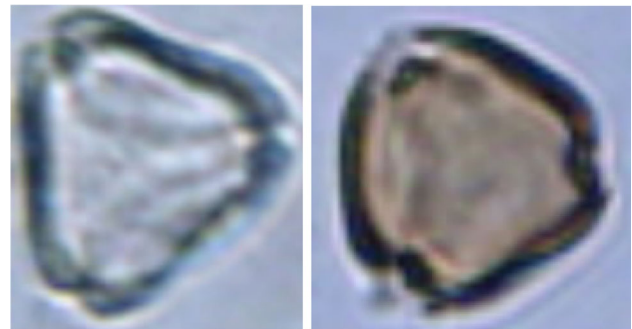

Fig. 6 Large variety of *Myrcia* pollen grains. Images taken from the POLEN23E [22] data set. *Myrica gale*, e.g., is described as monad, porate, isopolar, and aperture number 3

(a) *Eucalyptus*   **(b)** *Myrcia*

Fig. 7 Comparison between an *Eucalyptus* and a *Myrcia* pollen grain. Apart from color and texture, the exine and the apertures show strong similarities. *Eucalyptus globulus*, e.g., is described as monad, synaperturate, and aperture number 3. Both shapes are described as triangular (polar view) and the dominant orientation as oblique. Images taken from the POLEN23E [22] data set

due to the large varieties of position, perspective, and foci of pollen grain when observed under a LM.

This issue could be solved in two ways: More work has to be put into Feature Engineering, especially hybrid solutions that utilize a large spectrum of features, such as innovative ones as Group Integration. The authors of [42] show that textural features (when multiple ones combined) can reach a high classification accuracy around 95%, ignoring other feature classes such as morphological ones. However, the data set lacks validation by other methods, but the possibility of handling difficult pollen classes via Feature Engineering does exist. The other path is the DL approach[11]; in order to generalize and learn features, it requires a larger number of training samples to cover a wide range of positions and perspectives. In total, 35 images per class are not enough to solve this problem. Increasing the number of training samples can solve such issues. The addition of a knowledge system (e.g., a pollen calendar) can also help in the classification process.

The harmomegathic effect, described in Sect. 3.1, is hardly ever mentioned in the methods; therefore, we assume the preferred state of the pollen is hydrated. We assume this,

---

[11] It is important to notice that although CNNs work like a black-box, it is assumed that they have a texture bias instead of a shape bias, at least in the case of AlexNet[31].

due to the effects involved in analyzing pollen samples, e.g., from honey. In a laboratory pollen can also easily be dried. However, when pollen classification is supposed to be performed on the spot (airborne samples), harmomeghaty can be an additional problem, due to rain or moisture in the air.

7 Conclusion and discussion

The data we cited point strongly at an ever increasing importance of pollen analysis in various disciplines, from medicine and forensics to climate change research. Therefore, it is important to research and introduce automated and smart solutions for pollen analysis in all possible fields of application. In this work, we analyzed a large number of CV-based pollen classification methods in the area of ML and DL. To illustrate the problem in detail, we selected two specific use cases: the analysis of pollen samples in a laboratory for food analysis and safety (i.e., pollen samples from honey), and the analysis of airborne pollen samples for allergy-related weather forecasts. The actual use of AI-based pollen classification systems by countries or institutes is limited. The participation, results, and proposals that are shown by the partners of the AutoPollen project (see Sect. 1) show that. We stressed the importance for national and international standards that can help to standardize the process via fixed requirements.

As late as 2020, data sets such as POLLEN13K and the New Zealand data set emerged and offer benchmark-like quality. Most of the research work is done on proprietary data that cannot be evaluated properly. Many proposed techniques should be re-evaluated on one of the mentioned data sets. However, the data sets are limited in their actual applicability, since they are limited to a specific geographical flora. A possible solution would be to combine data sets together, if certain quality criteria are met, to offer a world wide data set, or at least a world wide data set that includes most allergy-related pollen, as it was already done for Germany (but not publicly available anymore). Such an international data set would make it easier for researchers to validate their methods with their own test samples and the specific data from the benchmark set. However, differences in image quality, resolution, and pollen coloring (e.g., fuchsine) can be problematic. The best solution would be to create one large benchmark set with uniform image standards and quality. Different data augmentation methods, such as synthetic image generation, require further investigation as well.

The comparison of the achieved results indicates that DL methods are favored and produce better quality results, especially when it comes to larger data sets with a higher number of pollen taxa. Promising results are also performed on data sets that are online available and can be evaluated. However, most works achieve an accuracy above 80%, which indicates

that the issues could be in the details and specifics of certain pollen types and data limitation.

For the purpose of research, a group of interesting possibilities are left untouched and require work such as a low-power pollen classification system. From the reviewed works, only a very few offer an entire system that incorporates all necessary steps (e.g., image acquisition, pre-processing, hardware-related components) for a real-world deployment. However, these system are usually powered by a strong PC, not a low-power system, and also the applied neural networks require a large amount of computational power and energy, for (continuous) training and inference. For stationary use, such systems work, but when viewed under the premise of low-energy, energy-harvesting, low-cost, or portability, a different approach might be necessary.

References

1. Allen, G.P., Hodgson, R.M., Marsland, S.R., Flenley, J.R.: Machine vision for automated optical recognition and classification of pollen grains or other singulated microscopic objects. In: 2008 15th International Conference on Mechatronics and Machine Vision in Practice, pp. 221–226 (2008). https://doi.org/10.1109/MMVIP.2008.4749537

2. Arias, D.G., Cirne, M.V.M., Chire, J.E., Pedrini, H.: Classification of pollen grain images based on an ensemble of classifiers. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 234–240 (2017). https://doi.org/10.1109/ICMLA.2017.0-153

3. Barnes, C.S.: Impact of climate change on pollen and respiratory disease. Curr. Allergy Asthma Rep. **18**(11), 59 (2018). https://doi.org/10.1007/s11882-018-0813-7

4. Battiato, S., Ortis, A., Trenta, F., Ascari, L., Politi, M., Siniscalco, C.: Detection and classification of pollen grain microscope images. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 4220–4227 (2020). https://doi.org/10.1109/CVPRW50498.2020.00498

5. Bonton, P., Boucher, A., Thonnat, M., Tomczak, R., Hidalgo, P.J., Belmonte, J., Galan, C.: Colour image in 2D and 3D microscopy for the automation of pollen rate measurement. Image Anal. Stereol. **21**(1), 25–30 (2002). https://doi.org/10.5566/ias.v21.p25-30

6. Boucher, A., Hidalgo, P.J., Thonnat, M., Belmonte, J., Galan, C., Bonton, P., Tomczak, R.: Development of a semi-automatic sys-

tem for pollen recognition. Aerobiologia **18**(3), 195–201 (2002). https://doi.org/10.1023/A:1021322813565

7. Boucher, A., Thonnat, M.: Object recognition from 3D blurred images. In: Object recognition supported by user interaction for service robots, 1: 800–803 (2002). https://doi.org/10.1109/ICPR.2002.1044880

8. Carrión, P., Cernadas, E., Gálvez, J.F., Damián, M., de Sá-Otero, P.: Classification of honeybee pollen using a multiscale texture filtering scheme. Mach. Vision Appl. **15**(4), 186–193 (2004). https://doi.org/10.1007/s00138-004-0150-9

9. Chen, C., Hendriks, E.A., Duin, R.P.W., Reiber, J.H.C., Hiemstra, P.S., de Weger, L.A., Stoel, B.C.: Feasibility study on automated recognition of allergenic pollen: grass, birch and mugwort. Aerobiologia **22**(4), 275–284 (2006). https://doi.org/10.1007/s10453-006-9040-0

10. Chen, W., Qu, T., Zhou, Y., Weng, K., Wang, G., Fu, G.: Door recognition and deep learning algorithm for visual based robot navigation. In: 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014) (2014). https://doi.org/10.1109/robio.2014.7090595

11. Chica, M.: Authentication of bee pollen grains in bright-field microscopy by combining one-class classification techniques and image processing. Microsc. Res. Tech. **75**, 1475–85 (2012). https://doi.org/10.1002/jemt.22091

12. Costa, C.M., Yang, S.: Counting pollen grains using readily available, free image processing and analysis software. Ann Bot **104**(5), 1005–1010 (2009). https://doi.org/10.1093/aob/mcp186

13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, pp. 886–893 (2005). https://doi.org/10.1109/CVPR.2005.177

14. Daood, A., Ribeiro, E., Bush, M.: Pollen grain recognition using deep learning. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., Isenberg, T. (eds.) Advances in visual computing lecture notes in computer science, pp. 321–330. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-50835-1_30

15. Daood, A., Ribeiro, E., Bush, M.: Sequential recognition of pollen grain Z-Stacks by combining CNN and RNN. In: Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference (FLAIRS 2018), pp. 8–13 (2018)

16. Duller, A., Guller, G., France, I., Lamb, H.: A pollen image database for evaluation of automated identification systems. Quat. Newsl. **89**, 4–9 (1999)

17. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The pascal visual object classes (VOC) challenge. Int. J. Comput. Vision **88**(2), 303–338 (2010)

18. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugen. **7**(2), 179–188 (1936). https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

19. France, I., Duller, A.W.G., Duller, G.A.T., Lamb, H.F.: A new approach to automated pollen analysis. Quat. Sci. Rev. **19**(6), 537–546 (2000). https://doi.org/10.1016/S0277-3791(99)00021-9

20. Gallardo-Caballero, R., García-Orellana, C.J., García-Manso, A., González-Velasco, H.M., Tormo-Molina, R., Macías-Macías, M.: Precise pollen grain detection in bright field microscopy using deep learning techniques. Sensors **19**(16), 3583 (2019). https://doi.org/10.3390/s19163583

21. Galán, C., Smith, M., Thibaudon, M., Frenguelli, G., Oteros, J., Gehrig, R., Berger, U., Clot, B., Brandao, R.: EAS QC working group: pollen monitoring: minimum requirements and reproducibility of analysis. Aerobiologia **30**(4), 385–395 (2014). https://doi.org/10.1007/s10453-014-9335-5

22. Gonçalves, A.B., Souza, J.S., Silva, G.G.D., Cereda, M.P., Pott, A., Naka, M.H., Pistori, H.: Feature extraction and machine learning for the classification of Brazilian savannah pollen grains. PLOS ONE (2016). https://doi.org/10.1371/journal.pone.0157044

23. Goodfellow, I., Bengio, Y., Courville, A.: Deep learning. MIT Press, USA (2016)

24. Halbritter, H., Ulrich, S., Grímsson, F., Weber, M., Zetter, R., Hesse, M., Buchner, R., Svojtka, M., Frosch-Radivo, A.: Illustrated pollen terminology. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-71365-6

25. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). https://doi.org/10.1109/CVPR.2016.90

26. Hesse, M., Halbritter, H., Weber, M., Buchner, R., Frosch-Radivo, A., Ulrich, S., Zetter, R.: Pollen terminology: an illustrated handbook. Springer-Verlag, Wien (2009). https://doi.org/10.1007/978-3-211-79894-2

27. Hirst, J.M.: An automatic volumetric spore trap. Ann. Appl. Biol. **39**(2), 257–265 (1952). https://doi.org/10.1111/j.1744-7348.1952.tb00904.x

28. 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. Rev. Palaeobot. Palynol. **167**(3), 175–183 (2011). https://doi.org/10.1016/j.revpalbo.2011.08.006

29. Huston, A., Sivaprakasam, V., Scotto, C., Lin, H.B., Eversole, J., Schultz, A., Willey, J.: Optical classification of bioaerosols using UV fluorescence and IR absorption spectroscopy. Proceedings of SPIE - The International Society for Optical Engineering **5617** (2004). https://doi.org/10.1117/12.578062

30. Khanzhina, N., Putin, E., Filchenkov, A., Zamyatina, E.: Pollen grain recognition using convolutional neural network. In: Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2018), pp. 409–414. Bruges, Belgium (2018)

31. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) Advances in neural information processing systems, pp. 1097–1105. Curran Associates Inc, USA (2012). https://doi.org/10.1145/3065386

32. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian Edge potentials advances in neural information processing systems, pp. 109–117. Curran Associates Inc, USA (2011). https://doi.org/10.5555/2986459.2986472

33. Langford, M., Taylor, G.E., Flenley, J.R.: Computerized identification of pollen grains by texture analysis. Rev. Palaeobot. Palynol. **64**(1), 197–203 (1990). https://doi.org/10.1016/0034-6667(90)90133-4

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

35. Li, P., Flenley, J.: Pollen texture identification using neural networks. Grana **38**(1), 59–64 (1999). https://doi.org/10.1080/001731300750044717

36. Li, P., Treloar, W.J., Flenley, J.R., Empson, L.: Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. J. Quat. Sci. **19**(8), 755–762 (2004). https://doi.org/10.1002/jqs.874

37. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal Loss for Dense Object Detection. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2999–3007 (2017). https://doi.org/10.1109/ICCV.2017.324

38. Lin, T.Y., Maire, M., Belongie, S.J., Bourdev, L.D., Girshick, R.B., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft COCO: common objects in context computer vision

- ECCV 2014, pp. 740–755. Springer International Publishing, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48

39. Lindbladh, M., O'Connor, R., Jacobson, G.L.: Morphometric analysis of pollen grains for paleoecological studies: classification of Picea from eastern North America. Am. J. Bot. **89**(9), 1459–1467 (2002). https://doi.org/10.3732/ajb.89.9.1459

40. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94

41. Manikis, G.C., Marias, K., Alissandrakis, E., Perrotto, L., Savvidaki, E., Vidakis, N.: Pollen grain classification using geometrical and textural features. In: 2019 IEEE International Conference on Imaging Systems and Techniques (IST), pp. 1–6 (2019). https://doi.org/10.1109/IST48021.2019.9010563

42. Marcos, J.V., Nava, R., Cristóbal, G., Redondo, R., Escalante-Ramírez, B., Bueno, G., Déniz, O., González-Porto, A., Pardo, C., Chung, F., Rodríguez, T.: Automated pollen identification using microscopic imaging and texture analysis. Micron **68**, 36–46 (2015). https://doi.org/10.1016/j.micron.2014.09.002

43. Ohe, W.V.D., Oddo, L.P., Piana, M.L., Morlot, M., Martin, P.: Harmonized methods of melissopalynology. Apidologie **35**(Suppl. 1), 18–25 (2004). https://doi.org/10.1051/apido:2004050

44. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002). https://doi.org/10.1109/TPAMI.2002.1017623

45. Pozo-Baños, M.D., Ticay-Rivas, J.R., Cabrera-Falcón, J., Arroyo, J., Travieso-González, C.M., Sánchez-Chavez, L., Pérez, S.T., Alonso, J.B., Ramírez-Bogantes, M.: Image processing for pollen classification. Biodivers. Enrich. Diverse World (2012). https://doi.org/10.5772/48456

46. Ranzato, M., Taylor, P., House, J., Flagan, R., LeCun, Y., Perona, P.: Automatic recognition of biological particles in microscopic images. Pattern Recognit. Lett. **28**(1), 31–39 (2007). https://doi.org/10.1016/j.patrec.2006.06.010

47. Rao, Z., Hua, D., He, T., Wang, Q., Le, J.: Ultraviolet laser-induced fluorescence lidar for pollen detection. Optik - Int. J. Light Electron Opt. (2017). https://doi.org/10.1016/j.ijleo.2017.02.075

48. Redondo, R., García, G.B., Chung, F., Nava, R., Marcos, J.V., Cristóbal, G., Rodríguez, T., González-Porto, A.V., Pardo, C., Déniz-Suárez, O., Escalante-Ramírez, B.: Pollen segmentation and feature evaluation for automatic classification in bright-field microscopy. Comput. Electron. Agric. **110**, 56–69 (2015). https://doi.org/10.1016/j.compag.2014.09.020

49. Reisert, M., Burkhardt, H.: Invariant features for 3D-Data based on group integration using directional information and spherical harmonic expansion. In: 18th International Conference on Pattern Recognition (ICPR'06), vol. 4, pp. 206–209 (2006). https://doi.org/10.1109/ICPR.2006.721

50. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017). https://doi.org/10.1109/TPAMI.2016.2577031

51. Rodriguez-Damian, M., Cernadas, E., Formella, A., Fernandez-Delgado, M., Sa-Otero, P.D.: Automatic detection and classification of grains of pollen based on shape and texture. IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.) **36**(4), 531–542 (2006). https://doi.org/10.1109/TSMCC.2005.855426

52. Rodriguez-Damian, M., Cernadas, E., Formella, A., Sa-Otero, R.: Pollen classification using brightness-based and shape-based descriptors. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., pp. 212–215 (2004). https://doi.org/10.1109/ICPR.2004.1334098

53. Ronneberger, O., Burkhardt, H., Schultz, E.: General-purpose object recognition in 3D volume data sets using gray-scale invariants - classification of airborne pollen-grains recorded with a confocal laser scanning microscope. In: Object recognition supported by user interaction for service robots, pp. 290–295 (2002). https://doi.org/10.1109/ICPR.2002.1048297

54. Ronneberger, O., Schultz, E., Burkhardt, H.: Automated pollen recognition using 3D volume images from fluorescence microscopy. Aerobiologia **18**(2), 107–115 (2002). https://doi.org/10.1023/A:1020623724584

55. Ronneberger, O., Wang, Q., Burkhardt, H.: Fast and robust segmentation of spherical particles in volumetric data sets from brightfield microscopy. In: 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 372–375 (2008). https://doi.org/10.1109/ISBI.2008.4541010

56. Santos, E.M.D.S., Marcai, A.R.S.: Segmentation of microscopic images for pollen grains detection. In: 8th International Conference of Pattern Recognition Systems (ICPRS 2017), pp. 1–6 (2017). https://doi.org/10.1049/cp.2017.0169

57. Schmidhuber, J.: Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015). https://doi.org/10.1016/j.neunet.2014.09.003

58. Scotland, R.W., Wortley, A.H.: How many species of seed plants are there? Taxon **52**(1), 101–104 (2003). https://doi.org/10.2307/3647306

59. Sevillano, V., Aznarte, J.L.: Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. PLOS ONE **13**(9), e0201807 (2018). https://doi.org/10.1371/journal.pone.0201807

60. Sevillano, V., Holt, K., Aznarte, J.L.: Precise automatic classification of 46 different pollen types with convolutional neural networks. PLOS ONE **15**(6), e0229751 (2020). https://doi.org/10.1371/journal.pone.0229751

61. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 39: 640–651 (2017). https://doi.org/10.1109/TPAMI.2016.2572683

62. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations (2015)

63. Stillman, E.C., Flenley, J.R.: The needs and prospects for automation in palynology. Quat. Sci. Rev. **15**(1), 1–5 (1996). https://doi.org/10.1016/0277-3791(95)00076-3

64. Viertel, P., König, M., Rexilius, J.: PollenGAN: Synthetic pollen grain image generation for data augmentation (forthcoming). In: 20th IEEE International Conference on Machine Learning and Applications (ICMLA) (2021)

65. Wang, Z., Bao, W., Lin, D., Wang, Z.: A local feature descriptor based on SIFT for 3D pollen image recognition. IEEE Access **7**, 152658–152666 (2019). https://doi.org/10.1109/ACCESS.2019.2948088

66. Xie, Y., OhEigeartaigh, M.: 3D Discrete spherical fourier descriptors based on surface curvature voxels for pollen classification. In: 2010 WASE International Conference on Information Engineering, pp. 207–211 (2010). https://doi.org/10.1109/ICIE.2010.56

67. Yaslan, Y., Cataltepe, Z.: Co-training with adaptive Bayesian classifier combination. In: 2008 23rd International Symposium on Computer and Information Sciences, pp. 1–4 (2008). https://doi.org/10.1109/ISCIS.2008.4717971

68. Zhang, Y., Fountain, D., Hodgson, R., Flenley, J., Gunetileke, S.: Towards automation of palynology 3: pollen pattern recognition using Gabor transforms and digital moments. J. Quat. Sci. **19**, 763–768 (2004). https://doi.org/10.1002/jqs.875

Philipp Viertel received his M.Sc. in computer science from Bielefeld University of Applied Sciences, Germany, in 2018, where he is currently employed as a research assistant and Ph.D. candidate. His research interests include computer vision and machine learning for pollen analysis.

Matthias König received the M.Sc. equivalent diploma degree in computer science from the University of Paderborn, Germany, and the Ph.D. in computer science from the OvG-University of Magdeburg, Germany, in 1997 and 2001, respectively. He received the MBA degree from the Open University Business School, UK, and the Ph.D. in management science from the University of Lancaster, UK, in 2005 and 2010, respectively. After several positions in industry as software developer and patent engineer between 2001 and 2012, he joined the Bielefeld University of Applied Sciences, Germany, in 2012, where he is currently a professor in computer science. His main areas of research interest are embedded systems, internet of things, and service robotics.