Universiteit
Leiden
The Netherlands

Aspects of the analysis of cell imagery: from shape to understanding

Li, C.

Citation
Li, C. (2024, June 27). *Aspects of the analysis of cell imagery: from shape to understanding*. Retrieved from https://hdl.handle.net/1887/3765419

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3765419

Note: To cite this publication please use the final published version (if applicable).

Chapter 3

Analysis of Automatic Image Classification Methods for Urticaceae Pollen Classification

This chapter is based on the following publication:
C, Li., M, Polling., L, Cao., B, Gravendeel., F, J. Verbeek., Analysis of Automatic Image Classification Methods for Urticaceae Pollen Classification. *Neurocomputing*. Vol. 522, 2023. pp. 181-193.

Abstract:

Pollen classification is considered an important task in palynology. In the Netherlands, two genera of the Urticaceae family, named *Parietaria* and *Urtica*, have high morphological similarities but induce allergy at a very different level. Therefore, distinction between these two genera is very important. Within this group, the pollen of *Urtica membranacea* is the only species that can be recognized easily under the microscope. For the research presented in this study, we built a dataset from 6472 pollen images and our aim was to find the best possible classifier on this dataset by analysing different classification methods, both machine learning and deep learning-based methods. For machine learning-based methods, we measured both texture and moment features based on images from the pollen grains. Varied feature selection techniques, classifiers as well as a hierarchical strategy were implemented for pollen classification. For deep learning-based methods, we compared the performance of six popular Convolutional Neural Networks: AlexNet, VGG16, VGG19, MobileNet V1, MobileNet V2 and ResNet50. Results show that compared with flat classification models, a hierarchical strategy yielded the highest accuracy with 94.5% among machine learning-based methods. Among deep learning-based methods, ResNet50 achieved an accuracy of 99.4%, slightly outperforming the other neural networks investigated. In addition, we investigated the influence on performance by changing the size of image datasets to 1000 and 500 images, respectively. Results demonstrated that on smaller datasets, ResNet50 still achieved the best classification performance. An ablation study was implemented to help understanding why the deep learning-based methods outperformed the other models investigated. Using Urticaceae pollen as an example, our research provides a strategy of selecting a classification model for pollen datasets with highly similar pollen grains to support palynologists and could potentially be applied to other image classification tasks.

3.1 Introduction

The analysis of pollen grains is widely used in detection and monitoring of airborne allergenic particles. In recent years, pollen seasons are prolonged due to global warming and climate change [87]. This subsequently causes an increase of hay fever patients who are affected by rising allergenic pollen levels in the air [88]. In palynological research, identification of pollen grains plays a key role to suggest safety treatments to patients with allergic rhinitis. It helps patients and medical professionals to monitor the levels of airborne allergenic pollen and thus plan outdoor activities and medication treatments accordingly. Pollen recognition analysis is often implemented by human visual inspection under the microscope, and includes the identification of differences in shape, texture, size and other specific features of pollen categories [89]. However, merely relying on human inspection for pollen classification tasks is unrealistic as the size of image datasets is rapidly increasing due to high-throughput screening, while the expertise needed to perform this detailed analysis is rapidly disappearing. Another limitation of manual classification is that it may induce classification biases with varied inspectors when the differences among pollen categories are very subtle. Thus, automatic classification techniques are now being developed that have proven to perform well in pollen classification tasks [63][89][90][91][92].

Researchers have adopted different approaches to automate the process of pollen classification. In general, the two main technical approaches of pollen image classification tasks are machine learning-based methods [89][93] and deep learning-based methods [62][63][94][95][96].

Machine learning methods need to be fed with manually selected features before they can extract these from images. The, so called, handcrafted features used in machine learning techniques are mostly based on shape, texture and other related properties of pollen grain images. The extracted features play an important role in the performance of classification. In addition, suitable feature selection methods and classifiers are also crucial for machine learning-based classification methods.

In the work of del Pozo-Baños et al. [97], a combination of geometrical and texture characteristics was proposed as the discriminative features for a 17 class pollen dataset. Incorporation of Linear Discriminant Analysis (LDA) and Least Square Support Vector Machines (LS-SVM) accomplished the best performance of 94.92% accuracy. Marcos et al. [98] extracted four texture features including Gray-Level Cooccurrence Matrices (GLCM), log-Gabor filters (LGF), Local Binary Patterns (LBP) and Discrete Tchebychev Moments (DTM) from a pollen image dataset with 15 classes. Fisher's Discriminant Analysis (FDA) and K-Nearest Neighbour (KNN) were subsequently applied to perform dimensionality reduction and multivariate classification. It yielded an accuracy of 95%. Manikis et al. [93] used texture features obtained by GLCM and seven geometrical features computed from the

binary mask of a pollen image dataset. A Random Forest (RF) classifier was used in the classification stage; with this classifier 88.24% accuracy was achieved on 6 pollen classes. Machine learning thus show highly varying results, and is seemingly dependent on the dataset used.

Instead of manual design of the features, deep leaning methods automatically extract image features through convolutional layers of the network. In recent years, many state-of-the-art Convolutional Neural Networks (CNNs) were applied in pollen classification tasks. In the work of Sevillano et al. [63], pretrained AlexNet was used to classify a dataset with 46 different classes of pollen grains. By incorporating data augmentation and cross-validation techniques, an accuracy of 98% was achieved. In the work presented by Battiato et al. [90], both AlexNet and SmallerVGGNet were implemented to classify five classes of pollen grains, with 13,000 images. The two networks obtained a performance of 89.63% and 89.73% accuracy, respectively. A seven layer deep Convolutional Neural Network designed by Daood et al. [94], was trained on a dataset of 30 pollen classes and accomplished a 94% correct classification rate. Astolfi et al. [99] analysed a pollen dataset composed of 73 pollen categories. They compared the performance of eight state-of-the-art CNNs which included Inception-V3, VGG16, VGG19, ResNet-50, NASNet, Xception, DenseNet-201 and Inception-ResNet-V2. They showed that DenseNet-201 and ResNet-50 achieved superior performance against other CNNs with an accuracy of 95.7% and 94.0%, respectively.

Based on the analysis of related work mentioned above, both machine learning and deep learning-based methods have achieved comparable performance on pollen datasets. However, the pollen datasets used in these studies is derived from species or genera from different plant families [100]. The morphology of each class of pollen is already clearly distinctive under the microscopy by human analysts. For example, the public POLEN23E dataset [89] consists of 23 pollen classes from the Brazilian Savannah, derived from 23 genera in 15 families. Each class of pollen has a, different shape, size and texture. The other public pollen dataset from the Brazilian Savannah, called POLLEN73S, which was analysed by Astolfi et al. [99], has 73 pollen classes with clearly variable colour, shape and other morphological differences. These distinct features ensured the high performance of the classification model applied. However, in this research, we are more interested in distinguishing genera of the same family Urticaceae, namely, *Parietaria* and *Urtica* which are morphological very similar, but cause completely different allergy levels. Pollen of the two genera cannot currently be distinguished easily by a palynologist; the species *Urtica membranacea* represents the only species that can be specifically distinguished.

Parietaria and *Urtica* are two genera commonly encountered in the Netherlands. The occurrence of *Parietaria* plants is very much increasing and could induce severe allergy in hay

fever patients while *Urtica* does not [96]. Species from the genus *Parietaria* as well as *Urtica membranacea* originate from the Mediterranean area and now increase in North Europe. Due to climate change these species can maintain themselves in northern countries such as the Netherlands. The pollen grains from these taxa exhibit a similar roundness, and are all very small, but differ in the following features: 1) different number of pores: *Parietaria* and *Urtica* have 3 to 4 pores, while this is variable for *Urtica membranacea* (usually 5 to 10; i.e. pantoporate). 2) The average size of *Parietaria* pollen is slightly smaller ($\sim$11-18μm) and it with a coarser and more irregular surface than *Urtica*. *Urtica* pollen are bigger in size on average ($\sim$15.2-21.1μm), and often have a more pronounced thickened exine around the pore (annulus). The shape of *Urtica membranacea* is slightly angular and is easily distinguished because of its small size ($\sim$10-12μm) and high number of pores. Although these pollen grains have the aforementioned differences, it is not possible for experts to distinguish the three different classes by the naked-eye using a light microscope. This is mainly because of their small size. Therefore, in order to improve the accuracy and efficiency of Urticaceae pollen classification, automatic algorithms are required.

Currently, very few studies focused on pollen classification of the Urticaceae family. Rodrìguez-Damián et al. [100] extracted both geometrical and texture features and probed three classifiers: Support Vector Machines (SVM), Multi-Layer Perceptron (MLP) and Minimum Distance Classifier (MDC). The best performance of 88% success rate was reached on a total of 291 pollen images of the three species *Parietaria judaica*, *Urtica urens* and *Urtica membranacea*. Compared with their relatively small Urticaceae dataset, we aimed to analyse a much larger dataset that includes all species (*Parietaria judaica*, *Parietaria officinalis*, *Urtica dioica*, *Urtica urens* and *Urtica membranacea*) present in the Netherlands. We grouped these five species into 3 classes: *Parietaria* (*Parietaria judaica*, *Parietaria officinalis*), *Urtica* (*Urtica urens*, *Urtica dioica*) and *Urtica membranacea*. Both *Parietaria* and *Urtica* dominate in the Netherlands but cause a totally different allergy level. *Urtica membranacea* is an exotic Mediterranean species and it is the only species can be easily distinguished. Hence our starting point for three labels and thus, our study is based on a three-class classification task. The best performance achieved in our study is 99.4% by a ResNet50. Actually, it is also possible to do a classification task over all five species (see Appendix B: Supplementary Table S1). Another challenge is that the pollen grains that we used were unacetolyzed. Acetolyzed pollen grains are those that all pollen materials are destroyed by acetolysis with the exception of sporopollenin that forms the outer pollen wall, the exine. In contrast to acetolyzed pollen grains, unacetolyzed pollen keep their original organic features which are less apparent. To the best of our knowledge, our previous work [96] was the first and the only time that CNNs were applied and compared for the

analysis of the unacetolyzed Urticaceae pollen grains. In this study, we extended this work further and aimed to find an automatic classification model with the best performance in both machine learning-based and deep learning-based methods for our unacetolyzed Urticaceae dataset. In general, for a deep learning model, a large dataset is required as input. However, there are many limitations for researchers to collect a sufficiently large dataset in practice. Subsequently, we were curious about how machine learning-based and deep learning-based methods work on a smaller sized image dataset. Therefore, two additional experiments on smaller datasets were designed to compare the performance of different classification models. For a 1000-image dataset, a ResNet50 yielded the best performance of 96.3% while for a 500-image dataset, it achieved the best accuracy of 93.3%.

3.2 Methods

3.2.1 Sample and Image Preparation

3.2.1.1 Sample Preparation of the Pollen Grains

Our pollen data included both fresh pollen specimens and dry pollen specimens [96]. Fresh pollen specimens were collected by an experienced biologist in the surroundings of Leiden and The Hague (the Netherlands) during the flowering season of 2018 and 2019. Dry pollen specimens were collected from the herbarium of Naturalis Biodiversity Center, Leiden, the Netherlands, using identification keys and descriptions. For each species in our dataset, pollen samples from 4 to 8 plants were taken, from different geographical locations in order to cover as much variation as possible (see Appendix B: Supplementary Table S2). Microscope slides were freshly prepared by aerobiological experts from Naturalis Biodiversity Center. The thecae of open flowers were carefully opened on a microscopic slide using tweezers. Non-pollen materials were manually removed. The pollen grains were mounted using a glycerin:water:gelatin (7:6:1) solution with 2% phenol and stained with Safranin (0.002% w/v). Cover slips were sealed with paraffin. Each slide contained only one plant of each species of Urticaceae.

3.2.1.2 Image Capturing and Pre-processing

The slide area rich in pollen was scanned automatically using a Zeiss Observer Z1 microscope with a Plan Apochromat 100× objective (NA 1.4), equipped with a Hamamatsu c9100 EM-CCD camera. As pollen grains are three dimensional, it is difficult to set a focal plane for pollen samples. Therefore, we captured 20 slices of images along the Z axis for pollen grains.

The step size was 1.8 µm. After obtaining a stack of images including pollen, the grains were detected and cropped; this is referred to as the 3D pollen stack. Fig. 3.1 (a) shows an example of a slice from the raw image. Fig. 3.1 (b) shows all 20 slices of different focal depths of an individual pollen grain. In total, 6472 individual pollen stack images were captured. Three categories were included for the image classification study. These were (1) *Parietaria* (including *Parietaria judaica*, *Parietaria officinalis*), (2) *Urtica* (*Urtica dioica*, *Urtica urens*), (3) *Urtica membranacea* (see Fig. 3.2).

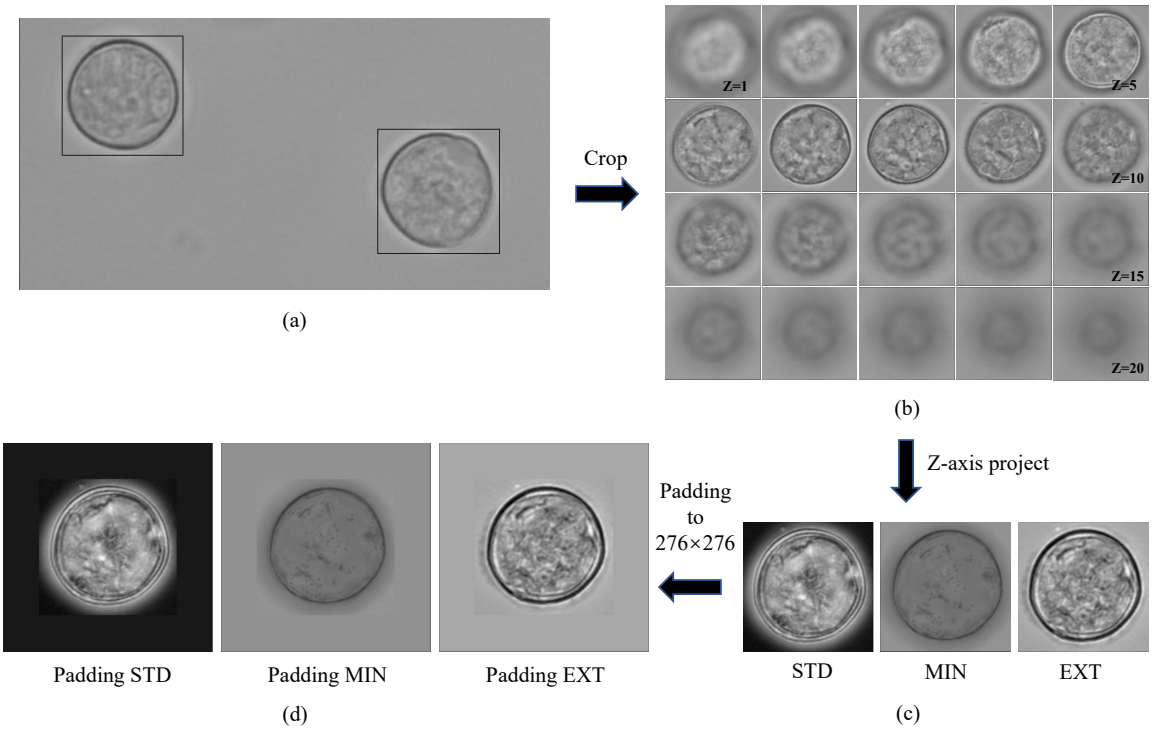

Fig. 3.1 The workflow of pollen image acquisition. (a) One plane of raw pollen image. (b) 20 slices at different focal depths of gray scale images of one individual pollen grain. (c) 3 different projections in Z axis, STD = Standard Deviation Projection, MIN = Minimum Intensity Projection, EXT = Extend Focus Projection. (d) The padding images of each projection.

As shown in Fig. 3.1 (b), not all of the 20 slices in the Z-stack were in-focus. In order to obtain as much informative features as possible, all Z-stack images were further processed using a Z-stack projection method [69]. Z-stack projection is a method of analysing and highlighting specific features from all slices in a stacked image without incorporating out-of-focus blurriness. The selected projections were Standard Deviation (STD), Minimum Intensity (MIN), and Extend Focus (EXT) [101], which are shown in Fig. 3.1 (c) and Fig. 3.2. The three projections per pollen grain were treated as three separate channel images

for the input of supervised classification models. Same-sized images are required to feed into classification models which is achieved by resizing images to the same size. However, for pollen images captured by a microscope, the morphology and details of pollen grains are expected to be changed by resizing. We did not opt for resizing as the resized images might ignore the original size differences of pollen grains, which is a potentially important diagnostic feature. There are several ways to preserve this nature of the features. One can crop images of the pollen grains to the same size from a slide-scan image. However, some pollen grains are very closely located to each other so that cropping to the same size might cause incomplete pollen separation. Therefore we chose for as another approach which is to use padding of the cropped images so that the resulting images all have the same size. For the padding size, the biggest size of all individual pollen images was selected, i.e.276 $\times$ 276 pixels. The padding value was set to the median value at the edge of each pollen image in order to make the content of the padding images more natural.

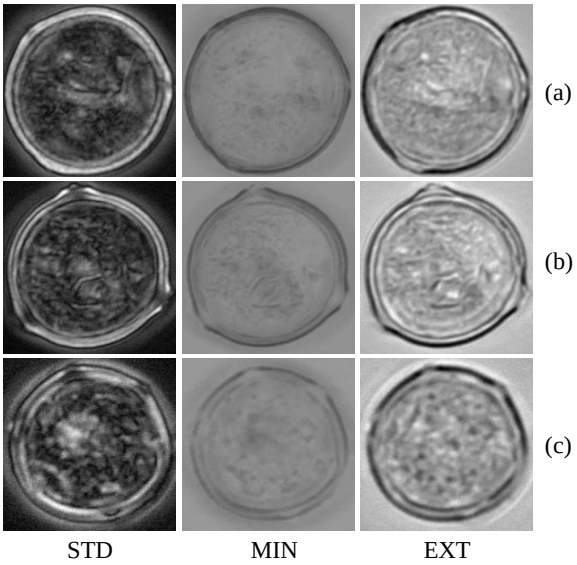

Fig. 3.2 A sample pollen grain of each category in our dataset. (a) *Parietaria judaica*. (b) *Urtica urens*. (c) *Urtica membranacea*. Each column represents the STD, MIN, EXT projections of each pollen grain, respectively.

After the pre-processing of the images, we aimed to find the best classification model for our Urticaceae pollen dataset. Machine learning and deep learning-based classification models were constructed and the performance of each model was evaluated and compared.

3.2.2 Machine Learning Methods

3.2.2.1 Feature Extraction and Selection

Machine learning methods require manual selection of relevant features before extracting these from images. One challenge is how to select an appropriate set of features for classification. By observing the characteristics of Urticaceae pollen grains, we noticed that *Parietaria* has a coarser ornamentation on the surface of its pollen grains, *Urtica* has thickened pores and *Urtica membranacea* has an angular outline. Texture attributes of surface and shape features was considered as the appropriate pollen descriptors for Urticaceae pollen grains. We aimed to include as much representative features as possible for Urticaceae pollen classification due to their high morphological similarities. The following selected features have been proven to be successful in classification tasks for pollen recognition: GLCM, LBP, Gabor filter texture features and Histogram of Oriented Gradients (HOG). These features have provided satisfactory results as reported in [90][98]. Both First Order Statistics (FOS), which are derived from statistical properties of the intensity histogram of an image, and Wavelet measurements, which is a texture analysis based on a Discrete Wavelet Transform (DWT) have been included as they have been successfully used in pattern recognition of cells [102][103]. In addition, the seven Hu invariant moments and three shape measures derived from the invariants, referred to as Extension, Dispersion and Elongation (EDE) were included as invariant descriptors for shape [104]. So, based on aforementioned image-based studies [105][106] we have selected six texture features and two moment-based features to represent the characteristics of pollen grains in our study. Table 3.1 shows the selected features with the dimensions of each feature vector.

Table 3.1 The dimension of feature vector of each feature.

Feature	Dimension
HOG	3600×3
LBP	416×3
Gabor filter	60×3
GLCM	24×3
FOS	5×3
Wavelet	9×3
Hu moments	7×3
EDE	3×3
Total	4124×3

HOG features in combination with a SVM classifier have proven to be a representative texture descriptor in the image recognition field [107]. In the procedure of HOG feature extraction, we divided an image into several small connected regions, aka cells. Each cell returns a 9×1 feature vector. In order to be more invariant in representing the changes of shadowing and illumination, a larger region, referred to as the block, is formed. The block consists of four cells and returns a 36×1 feature vector. In the experiment, a pollen image with size (276×276) can be divided into 100 blocks, consequently, a 3600×1 feature vector is returned at the end.

LBP is an invariant descriptor that can be used for texture classification. A n-digit binary number is obtained by comparing each pixel with its n neighbour pixels on a circle with radius r and used to compute the histogram. In our study, we fine-tuned the parameters and set it as $n = 24$, $r = 3$. Similar to the HOG feature extraction procedure, the image was also divided into 16 smaller blocks. In this manner a 416×1 LBP feature vector is returned.

GLCM characterizes the texture of images by considering the spatial relationship of pairs of pixels in an image. GLCM is created based on a statistical rule $P(i, j, d, theta)$, which refers to the number of times that gray-level j occurs at a distance d and at a direction $theta$ from gray-level i. Our experiment set $d = 1$ and the direction $theta = 0°, 45°, 90°, 135°$. We further calculated the properties based on the matrix which are defined by Haralick et al. [108], as the extracted feature vector. Finally, a 24×1 feature vector was returned.

A 5×1 texture feature vector of FOS named standard deviation of intensity, smoothness, skewness, uniformity and entropy was calculated [103]. These measures reflect the statistical properties of the intensity histogram of each pollen image. Wavelet-based texture measurements show the image details in different directions after DWT. We calculated the mean, standard deviation and entropy of intensity in three directions (horizontal, vertical, diagonal) of each pollen image. The dimension of wavelet measurement is 9×1.

Another commonly used texture descriptor is the Gabor filter: it reflects frequency content in a specific direction of a localized region of the image. In this study, 12 Gabor filters are designed at four directions 0°, 45°, 90°, 135° with three different frequencies $\pi/4, \pi/2, 3\pi/4$. Therefore, the dimension of the Gabor filter feature vector is 60×1.

Hu moments are normally extracted from pollen images with the property of scale and rotation invariance. A total of seven moment invariants as proposed by Hu [109] were extracted. EDE features are derived from the 1^{st} and the 2^{nd} order invariants. Even though the morphological differences of pollen between genera is subtle, it was expected that these image moment features could play a role in the pollen classification task.

Each pollen image in our dataset consisted of 3 projections (STD, MIN, and EXT) obtained by projecting 20-slice Z-stack images. Fig. 3.2 shows that the features of each pro-

jection are different, especially the texture features. In order to include as much information of the pollen grain dataset as possible, we calculated 8 features for 3 projections of pollen images and concatenated these together as the final feature vector (cf. Table 3.1). Therefore, after feature extraction, the dimension of feature vector reaches 4124×3. Compared with public pollen image datasets like POLEN23E, POLLEN73S [89][99] and the 2D Urticaceae pollen images used in [100], our dataset based on a method of projection of 3D images might intrinsically extract more representative features. This partially underlies the reason of the high performance results that we have achieved.

In order to remove redundant and irrelevant features, feature selection and dimensionality reduction techniques were applied. Feature selection returns prominent subsets of features while dimensionality reduction creates new features with lower dimension from the original features. Feature selection includes a filter method, wrapped method and embedded method [110]. These methods have been shown to improve the accuracy of classification studies [111][112].

In this study, feature selection methods including Mutual information, SelectFromModel and Principal Component Analysis (PCA) were assessed. Mutual information is a filter feature selection method. In this method a subset of the best K features which are most relevant to the target labels is chosen; the selection of the number K is mostly based on experience. The embedded method named SelectFromModel works more flexible. It selects the most relevant features according to the performance of machine learning models during the training process. This integrated approach ensures that the selected features are the best for the model. Alternatively, PCA is widely used in feature dimensionality reduction. It is a process of computing the principal components and preserving the first few of them that maximize the variances between different classes. PCA is known for obtaining lower-dimensional features and improving the accuracy of machine learning model in many fields such as image recognition [97][113].

In short, these selected feature selection and dimensionality reduction methods were applied after feature extraction. Subsequently, classification models were used as the last step to classify pollen grains.

3.2.2.2 Classifiers

Once features are extracted from images, an efficient classification model is required so that it can perform well on the pollen classification task. A large number of classification approaches exist. In this study, SVM, RF, MLP and Adaboost classifiers were used as they have shown to perform well in previous pollen classification studies [90][93][100].

Combined with extracted features and feature selection methods, the classifiers have been trained and the hyperparameters were tuned based on the performance of the experiment.

3.2.2.3 Hierarchical Strategy

A flat classification model is a straight-forward approach for taxonomic classification tasks. Only one classifier is used to classify all classes. However, the process ignores potential hierarchical structure among different classes which could reduce performance [114]. Hierarchical classification can be seen as a particular tree-structured approach. It merges the classes which are more similar into subgroups and classifies these subgroups separately. Varied classifiers are used to classify classes at different hierarchical levels until reaching the leaf nodes. Hierarchical strategy has been used in many classification tasks [103][115] and has proven to increase the performance compared with flat classification models.

For our work, we structured the three classes of pollen grains as a hierarchical tree as shown in Fig. 3.3. We used a local classifier per parent node approach to train a two-stage classifier for each parent node in the hierarchical tree. In the first stage we merged *Parietaria* and *Urtica* into one subgroup based on high morphological similarity. *Urtica membranacea* is a distinct species that can already be clearly distinguished under the light microscope and it was therefore treated as the other subgroup. In the second stage, *Parietaria* and *Urtica* were subsequently classified. At both stages we selected the best classifier and feature selection method for each parent node in order to get a better performance of the hierarchical classification model.

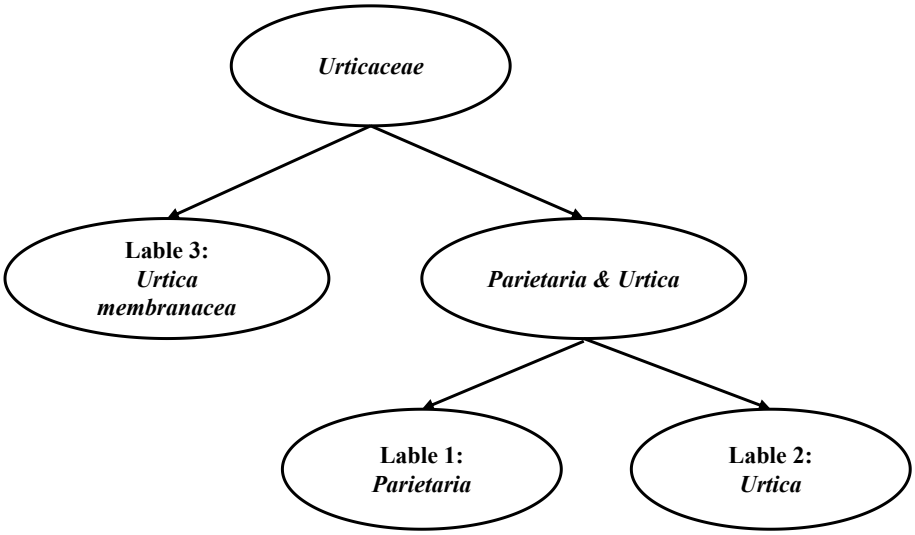

Fig. 3.3 Hierarchical tree of pollen classification.

3.2.3 Deep Learning Methods

3.2.3.1 Convolutional Neural Networks

We selected several well-established deep learning models for pollen classification. AlexNet achieved the best performance on an ImageNet classification task in 2012 [116]. In order to prove that convolutional network depth affects image classification accuracy, Simonyan et al. [117] proposed VGGNet for large-scale image recognition. A 16-layer VGG16 network and 19-layer VGG19 network have proven to be the two best-performing convolutional neural networks in other studies. ResNet introduces a residual learning framework to ease the training process of very deep networks—up to 152 layers [118]. Even though ResNet has lower complexity than VGGNets, it still has millions of parameters making the network computational heavy. A more light-weighted set of neural networks, named MobileNets, was designed in order to embed it into mobile devices or other applications [72]. In this study, we selected the aforementioned models: i.e., AlexNet, VGG16, VGG19, ResNet50, and MobileNet V1 and V2 [73] to classify our pollen dataset.

In addition, we used a transfer learning technique to alleviate the computational burden of training from scratch. With our three classes pollen dataset, we fine-tuned the pretrained AlexNet based on the ImageNet dataset in the PyTorch framework. The other pretrained networks implemented in the Keras Library were fine-tuned on the TensorFlow platform. All experiments were executed on a dedicated server equipped with two NVidia GeForce GTX 2070 with 8 GB GPUs using Linux Ubuntu operating system.

3.2.3.2 Data Augmentation

Deep learning models need a large number of image datasets covering diverse scenarios. Data augmentation techniques play an important role in increasing the variety of images. Furthermore, if appropriate transforms are applied to a dataset, data augmentation can greatly improve the performance and reduce overfitting. In our case, differences between pollen of Urticaceae genera are very subtle and slight configurational changes during image capturing may affect the classification performance. Therefore, a large amount of training data was needed for our study.

In order to simulate the possible transforms of pollen data, brightness and flip transforms were most obvious and straightforward to select and therefore applied as augmentation options. Other transforms like rotation, zoom range, etc., were not selected.

3.2.3.3 Cross-validation and Hard Voting

Cross-validation is applied in an image training process to improve the effectiveness, robustness and generalization ability of deep learning models, as well as to prevent overfitting. In this study, K values of 5 or 10 were used [62][119]. A 5-fold cross-validation means the ratio of training data and validation data is 8:2 while 10-fold cross-validation means the ratio is 9:1. We compared the performance of deep learning models with 5-fold cross-validation and 10-fold cross-validation, respectively. After 5/10-fold cross-validation, 5/10 models were obtained and tested on test datasets. In this study, hard voting was adopted to calculate the final accuracy rather than the average accuracy of 5/10 models. Hard voting sums the votes for class labels from each model for predicting the class with the majority votes. The experimental results show that the hard voting technique further improves the classification performance on the test dataset.

3.2.4 Performance Evaluation

Before addressing the results, we first introduce the performance measures that we used. These were:

$$precision = \frac{TP}{TP + FP} \tag{3.1}$$

$$recall = \frac{TP}{TP + FN} \tag{3.2}$$

$$F1\ score = \frac{2 \times precision \times recall}{precision + recall} \tag{3.3}$$

Where *TP* refers to true positives, *TN* represents true negatives, *FP* is false positives and *FN* is false negatives. High precision and recall values are able to verify good performance against false positives and false negatives of a model [63]. The F1 score is an overall measurement which combines precision and recall together. A high F1 score means that a model retrieved both low false positives and low false negatives, which proves the consistency of those measures and the reliability of the model. Precision, recall and F1 score were calculated as the average weighted by the number of true instances for each class in our experiments. The accuracy of the classification model was also calculated by the number of true predictions divided by the total number of samples.

The performance measurements mentioned above are commonly applied in flat classification models. However, they are not suitable for hierarchical classification models since they do not differentiate the misclassification errors among different hierarchical stages.

Instead, we adopted the measures suggested in [115], which include hierarchical precision (*hP*), hierarchical recall (*hR*) and hierarchical f-measure (*hF*). These are defined as follows:

$$hP = \frac{\sum_i |\widehat{C}_i \cap \widehat{C}'_i|}{\sum_i |\widehat{C}'_i|} \tag{3.4}$$

$$hR = \frac{\sum_i |\widehat{C}_i \cap \widehat{C}'_i|}{\sum_i |\widehat{C}_i|} \tag{3.5}$$

$$hF = \frac{2 \times hP \times hR}{hP + hR} \tag{3.6}$$

Where $\widehat{C}_i$ is a set of real classes with all of its ancestors and $\widehat{C}'_i$ is a set of predict classes with all of its ancestors. Ancestors here refer to all the nodes which are connected to the specific real/predict class node in the hierarchical tree structure.

3.3 Experiment Results and Discussion

We derived results from two parts: a comparison of pollen classification performance with different machine learning algorithms and an analysis of performance of deep learning neural networks. Two additional experiments show how machine learning-based and deep learning-based methods work on smaller-size image datasets. Our results are based on performance measures.

3.3.1 Results with Machine Learning Methods

For machine learning methods, the 6472 pollen images were divided into training and test datasets in a ratio of 9:1. In this experiment, we compared the performance of each classification model using 5-fold and 10-fold cross-validation, respectively. In addition, a grid search technique [120] was applied in the training process to help search for optimal hyperparameters automatically. With this technique, a list of hyperparameter values is defined beforehand and the optimal set of parameters, that can maximize the accuracy of the model, is returned.

Table 3.2 shows the performance of each classifier with corresponding hyperparameter settings and feature selection methods. In Table 3.2 we present the results of a SVM with a Radial Basis Function (RBF). This SVM, with a penalty parameter C=4, was shown to be optimal for this type of data. These parameters were determined by a grid search technique. Two dictionaries which included kernel functions (linear, rbf, poly) and penalty values (from

1 to 10) were set and the best parameter combination was returned. Other parameters in SVM were selected by default values. The highest score with an accuracy of 91.5% (σ= ±0.008) and F1 score of 0.915 was achieved in combination with a PCA threshold of 0.8 with 5-fold cross-validation. The threshold 0.8 of the PCA means that the first few principal components with the ratio of accumulated data variation and total data variation greater than 0.8 are preserved while all others are discarded. In this case, the final size of the selected feature vector is 179 $\times$ 1 (see Appendix B: Supplementary Table S3). For the RF classifier, the number of trees was set to 500. The SelectFromModel function of a threshold 'mean' embedded in a classifier can achieve the best performance of 88.6% (±0.015) with 10-fold cross-validation. The threshold 'mean' was set according to the importance of each feature. It means that a feature whose importance is greater or equal to the 'mean' is kept while others are discarded. The final size of the selected feature vector is 2064 $\times$ 1. Similarly, all of these parameters are fine-tuned by the grid search technique.

Table 3.2 Performance comparison of different flat classification models. Standard deviation, of each subset via cross-validation, is given in brackets.

Classifiers	Hyperparameters	Feature selection with threshold	Cross-validation	Precision	Recall	F1 score	Accuracy
SVM	kernel='rbf', C=4	PCA (0.8)	10-fold	0.913 (±0.012)	0.913 (±0.012)	0.913 (±0.012)	0.913 (±0.012)
			5-fold	0.915 (±0.008)	0.915 (±0.008)	0.915 (±0.008)	0.915 (±0.008)
RF	Estimators=500	SelectFromModel (mean)	10-fold	0.886 (±0.015)	0.886 (±0.015)	0.886 (±0.015)	0.886 (±0.015)
			5-fold	0.884 (±0.010)	0.884 (±0.010)	0.884 (±0.010)	0.884 (±0.010)
MLP	Solver='sgd', Maxiter=300	PCA (0.85)	10-fold	0.898 (±0.011)	0.898 (±0.011)	0.898 (±0.011)	0.898 (±0.011)
			5-fold	0.890 (±0.008)	0.890 (±0.008)	0.890 (±0.008)	0.890 (±0.008)
Adaboost	Estimators=500, LR=0.5	Mutual Information (2000)	10-fold	0.789 (±0.014)	0.754 (±0.021)	0.749 (±0.025)	0.754 (±0.021)
			5-fold	0.784 (±0.015)	0.745 (±0.024)	0.743 (±0.028)	0.748 (±0.024)

Furthermore, we carried out experiments with MLP and Adaboost classifiers. The MLP led to the best accuracy of 89.8% (±0.011) with the following settings: the optimizer was Stochastic Gradient Descent (SGD); the number of maximal iterations was 300; PCA feature reduction was at a threshold of 0.85. The final size of feature vector after PCA feature reduction becomes 337 $\times$ 1. Adaboost reached a performance of 75.4% (±0.021) with the number of Estimators set to 500 and a Learning Rate (LR) of 0.5. Mutual Information plays an important role in the accuracy of the Adaboost classifier because it selected 2000

features most relevant to target classes of pollen datasets. In addition, 5-fold and 10-fold cross-validation obtained a comparable performance with different machine learning-based classification models.

In order to improve the performance of the flat classification model further, we applied a hierarchical strategy classification. We have implemented different combinations of flat models to form a two-level hierarchical structure which includes SVM + SVM, SVM + MLP, MLP + SVM, SVM + RF, RF + SVM. Table 3.3 shows all permutations with SVM for the hierarchical classification model except for Adaboost. This is because SVM achieved the best performance of the flat classification models while Adaboost achieved the lowest performance. In Table 3.3 the best combination is SVM + SVM which obtained 94.5% of accuracy and 0.941 of all of the *hP*, *hR* and *hF*. The reason why *hP*, *hR* and *hF* are equal is that our simple hierarchical tree structure only has 3 layers and for each parent node, it only has 2 children. According to the definition of *hP* and *hR* (cf. eq. (5), (6)), in this case, the calculation of *hP*, *hR* and *hF* is equal. Based on our experiments, a hierarchical model which combined SVM + PCA and SVM + PCA at both 2 levels was considered as the best model among machine learning-based methods.

Table 3.3 Performance comparison of hierarchical classification models.

Hierarchy Level 1		**Hierarchy Level 2**					
Classifier with Hyperparameters	Feature selection with threshold	Classifier with Hyperparameters	Feature selection with threshold	**hP**	**hR**	**hF**	**Accuracy**
SVM (kernel='rbf',C=6)	PCA (0.8)	SVM (kernel='rbf',C=4)	PCA (0.8)	0.941	0.941	0.941	0.945
SVM (kernel='rbf',C=4)	PCA (0.8)	MLP (Solver='sgd', Maxiter=300)	PCA (0.85)	0.920	0.920	0.920	0.916
MLP (Solver='sgd', Maxiter=300)	PCA (0.85)	SVM (kernel='rbf',C=4)	PCA (0.8)	0.929	0.929	0.929	0.933
SVM (kernel='rbf',C=4)	PCA (0.8)	RF (Estimators=500)	SelectFromModel (mean)	0.907	0.907	0.907	0.897
RF (Estimators=500)	SelectFromModel (mean)	SVM (kernel='rbf',C=4)	PCA (0.8)	0.913	0.913	0.913	0.911

3.3.2 Results with Deep Learning Methods

Our starting point has been to work with commonly available deep learning methods, the AlexNet, VGG16, VGG19, ResNet50, MobileNet V1 and MobileNet V2. First of all, a total of 6472 pollen grain images were divided into a training set and test set in a ratio of 9:1 randomly. The test set was composed of images that were not seen by the model during the training process and that were used to test the trained classification model. Secondly, considering that deep learning models require a huge amount of data, a data augmentation

technique was applied. Thirdly, similar with machine learning models, 5-fold and 10-fold cross-validation were used to prevent overfitting and increase the robustness as well as the generalization ability of the deep learning models. Data augmentation process was performed for each cross-validation set independently.

Based on these aforementioned procedures, we fine-tuned six representative deep learning classification models. The pretrained AlexNet was implemented in the PyTorch framework. The whole five convolutional layers and three fully connected layers were fine-tuned with a learning rate 0.001. We set the batch size to 128 and the number of batches to 4000. Fig. 3.4 (a) shows the performance plot of AlexNet. The plot shows the accuracy and loss for both training and validation dataset. The accuracy drastically increases in the first 700 batches and then converges gradually. AlexNet achieved an accuracy of 94.1% with a standard deviation of (±0.002) using 10-fold cross-validation while 5-fold cross-validation retrieved a comparable accuracy of 92.4% (±0.002) (see Table 3.4). The average accuracy and standard deviation were calculated by training each model three times. The accompanied precision, recall and F1 score achieved with the same 0.941. The reason why these three measurements are so similar is that, for this case, False Positive (FP) samples are nearly equal to the number of False Negatives (FN). The consistency of these measurements shows the reliability of the model. Six positive samples and six negative samples among three classes of pollen grains performed by AlexNet are shown in Fig. 3.5 (a) and (b). The actual label, predicted label and confidence score of each sample are indicated. Label 1 to 3 represent the 3 classes of pollen: *Parietaria*, *Urtica*, *Urtica membranacea*. In Fig. 3.5 (a), positive samples clearly show the distinguished properties. *Urtica* pollen has obviously thickened pores compared with pollen of *Parietaria* and *Urtica membranacea*. Pollen of *Urtica membranacea* has more angular outlines than pollen of the other two genera. In Fig. 3.5 (b) illustrates that, when the properties of the 3 classes of pollen are not clearly displayed, the network will misclassify these samples because of high similarities among the 3 classes.

Similarly, the pretrained VGG16 and VGG19 models were fine-tuned with a batch size of 64 and the number of epochs set to 30. The whole network was fine-tuned using learning rate 2e-5 without freezing any layers. Fig. 3.4 (b) and (c) show the performance plots of VGG16 and VGG19, respectively. Both plots show that the models converge well in the training process. Table 3.4 lists detailed measurements of these two models. For 10-fold cross-validation, VGG16 obtained an average accuracy of 98.3% with a standard deviation of (±0.001) while VGG19 achieved a comparable average accuracy of 98.6% with (±0.002).

The pretrained ResNet50, MobileNet V1 and MobileNet V2 models were constructed based on Keras as well. And all of these models were fine-tuned on our pollen dataset. Table 3.4 shows that light-weights MobileNets achieved a comparable accuracy with VGGNets

but it had a slightly higher standard deviation. This means that light models are not as robust as heavy-weight models. ResNet50 obtained the highest performance of 99.4% with 10-fold cross-validation among the six models investigated due to its deeper network layers and a creative residual structure. Consequently, ResNet50 was selected as the best performing model among all the models that we implemented. In addition, both 5-fold and 10-fold cross-validation achieved comparable performance for all of the deep learning models studied.

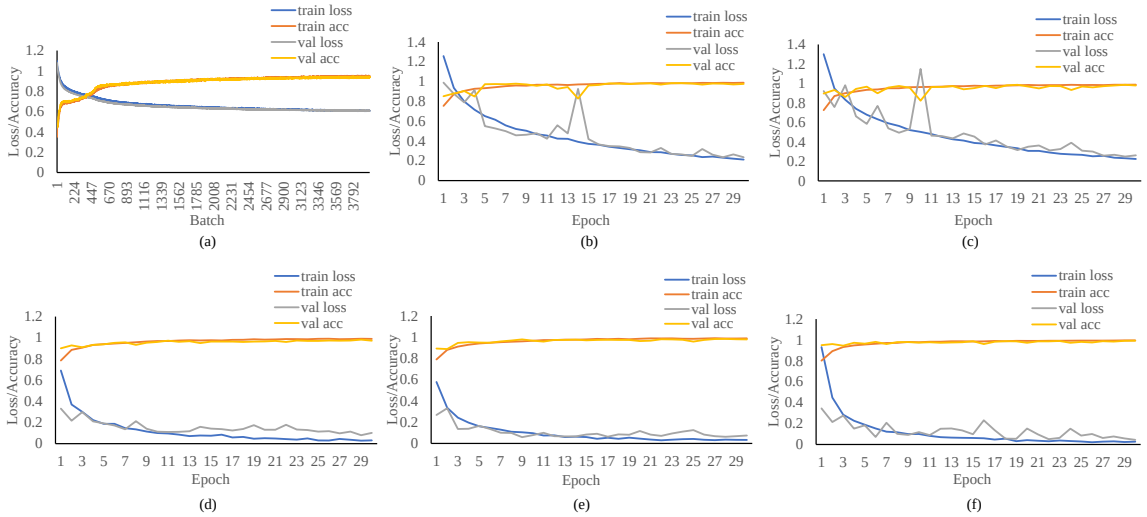

Fig. 3.4 The performance plots of (a) AlexNet, (b) VGG16, (c) VGG19, (d) MobileNet V1, (e) MobileNet V2, and (f) ResNet50, in terms of training loss, etc., with respect to the number of epochs.

Table 3.4 Classification performances of different deep learning classification models. Standard deviation, training each model three times, is given in brackets.

	Cross-validation	**Precision**	**Recall**	**F1 score**
AlexNet	10-fold	0.941(±0.002)	0.941(±0.002)	0.941(±0.002)
	5-fold	0.924(±0.002)	0.924(±0.002)	0.924(±0.002)
VGG16	10-fold	0.983(±0.001)	0.983(±0.001))	0.983(±0.001)
	5-fold	0.985(±0.002)	0.985(±0.002)	0.985(±0.002)
VGG19	10-fold	0.986(±0.002)	0.986(±0.002)	0.986(±0.002)
	5-fold	0.988(±0.003)	0.988(±0.003)	0.988(±0.003)
ResNet50	10-fold	0.994(±0.002)	0.994(±0.002)	0.994(±0.002)
	5-fold	0.993(±0.002)	0.993(±0.002)	0.993(±0.002)
MobileNet V1	10-fold	0.981(±0.003)	0.981(±0.003)	0.981(±0.003)
	5-fold	0.980(±0.002)	0.980(±0.002)	0.980(±0.002)
MobileNet V2	10-fold	0.985(±0.003)	0.985(±0.003)	0.985(±0.003)
	5-fold	0.984(±0.003)	0.984(±0.003)	0.984(±0.003)

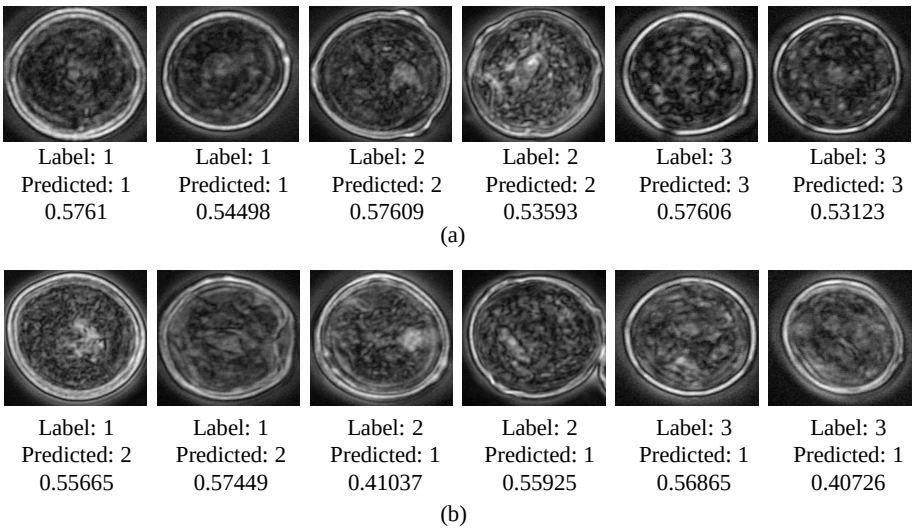

Fig. 3.5 Examples of classification performed by AlexNet. (a) Positive samples with their predicted label and confidence score. (b) Negative samples with their predicted label and confidence score.

Fig. 3.6 and 3.7 show the positive and negative samples classified by VGG16 and ResNet50, respectively. Compared with AlexNet (Fig. 3.5), the confidence score of the classified pollen of VGG16 was higher. The reason is that VGG16 has deeper layers which results in the extraction of more detailed and distinct features of pollen data. ResNet50 has a much deeper and complex network structure and the classification accuracy was higher than that of VGG16 (Table 3.4). For positive samples in Fig. 3.7 (a), the confidence score of ResNet50 was almost 1.00 which was higher than that of VGG16. And in the test dataset, only three negative samples, shown in Fig. 3.7 (b), were misclassified due to the high performance of the ResNet50 model.

After analysing automatic classification models based on both machine learning and deep learning methods on our pollen dataset, we observed that the ResNet50 neural network reached an accuracy of 99.4% (±0.002) which is 4.9% higher compared to the hierarchical machine learning model. Deep learning-based methods perform better to classify our Urticaceae pollen grains. In addition to our pollen images dataset, we have used the deep learning classifiers to other pollen image datasets available to us. These have not been used in the training/testing but are used as unseen samples to probe the classifiers from our study. The classification results with these additional datasets confirm the findings from study. Early results with these extra datasets, based on VGG16, have already been reported in [96]. With our ResNet50 model, the results with unseen data are even better. In Appendix B: supplementary Table S4 these results are summarized.

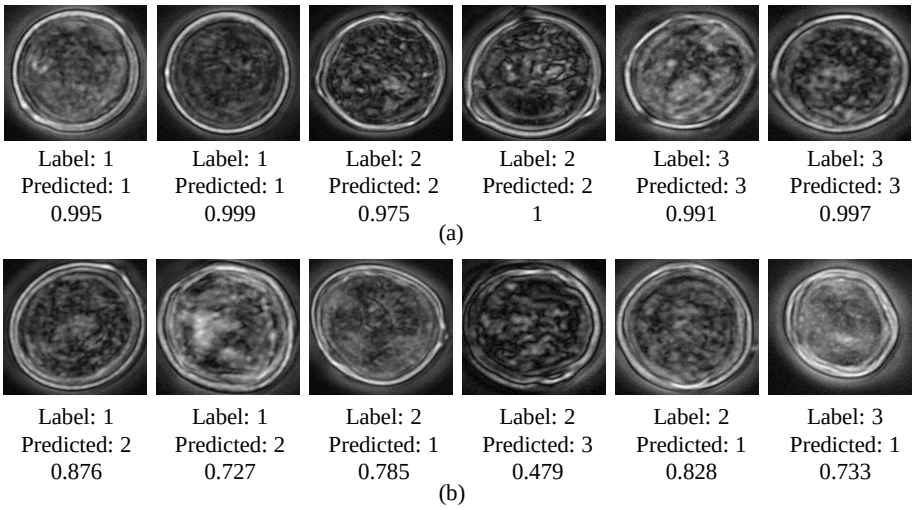

Fig. 3.6 Examples of classification performed by VGG16. (a) Positive samples with their predicted label and confidence score. (b) Negative samples with their predicted label and confidence score.

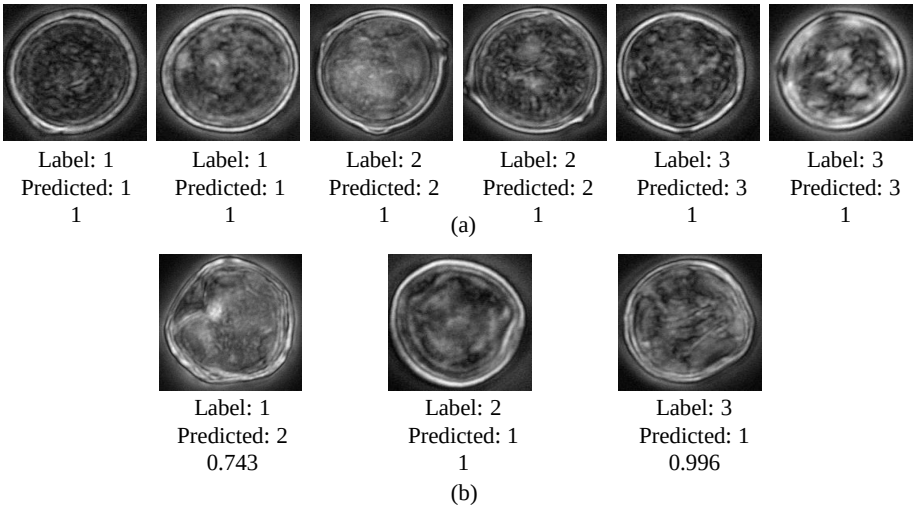

Fig. 3.7 Examples of classification performed by ResNet50. (a) Positive samples with their predicted label and confidence score. (b) Negative samples with their predicted label and confidence score.

3.3.3 Results on Smaller-size Image Datasets

It is common knowledge that the training process of deep learning model requires the use of a large data set. However, in daily practice, there are limitations in the collection of sufficient samples and images. Therefore, we examined the robustness of both machine learning-based and deep learning-based methods when facing a smaller dataset. Are machine learning-based method and deep learning-based method comparable in performance? To answer this question, starting from the original data, two smaller pollen image datasets consisting of 1000-sized and 500-sized image subsets, were constructed. These image subsets were randomly selected from 6472 images. And the ratio of the 3 classes was 1:1:1. The experimental results on smaller datasets shown in Table 3.5 was based on one round of selection.

On both smaller pollen datasets (1000 and 500 images), the same six deep learning-based models were applied. For machine learning models, we refine-tuned the hyperparameters of the best performed flat model (SVM) and hierarchical model (SVM+SVM). Table 3.5 shows the performance of both machine learning-based and deep learning-based methods on the two smaller image datasets. Compared with the 88% accuracy of the flat model on the 1000-image dataset, the 93.9% accuracy obtained by the hierarchical model demonstrated that hierarchical strategy improves the performance. The performance of the hierarchical model was, however, still lower than the deep learning models, except for AlexNet. We obtained similar results with the larger 6472-image dataset (Table 3.3 and Table 3.4). This is probably due to the fact that AlexNet has a shallow layer-structure which includes only five convolutional layers and three fully connected layers. The results indicate that extracting as many features as possible manually as well as using a hierarchical strategy outperforms a shallow deep learning neural network such as AlexNet. For the 500-image dataset we obtained similar results.

Table 3.5 Performance comparison of different methods on smaller-size image datasets. Standard deviation, training each model three times, is given in brackets.

	Deep learning-based						**Machine learning-based**	
	AlexNet	VGG16	VGG19	ResNet50	MobileNet V1	MobileNet V2	Flat model	Hierarchical model
Accuracy of 1000 images	0.916 (±0.006)	0.943 (±0.006)	0.943 (±0.012)	0.963 (±0.012)	0.947 (±0.015)	0.950 (±0.010)	0.880	0.939
Accuracy of 500 images	0.861 (±0.032)	0.920 (±0.000)	0.920 (±0.020)	0.933 (±0.012)	0.927 (±0.012)	0.907 (±0.023)	0.760	0.896

In order to obtain sufficient information for a statistical analysis of the performance of the models, we used a cross-validation approach over the entire image set with two

differently sized groups of subsets [1]. We implemented the 5-fold cross-validation to select five image subsets from the 6472-image dataset, as well as the 10-fold cross-validation to select ten smaller image subsets. With this selection method, the average performance of all subsets from different models was compared, the results of which are given in Appendix B: Supplementary Table S5. The experiments confirmed that ResNet50 achieved the best performance on both 1000- and 500-sized dataset.

Deep learning-based methods show a better performance on both the large and smaller pollen datasets. An ablation study was conducted to help understand why this difference was retrieved. Convolutional layers of deep neural networks can catch more representative features compared with extracting handcrafted features manually. We visualized intermediate feature maps of VGG16 and ResNet50 in Fig. 3.8 and Fig. 3.9 to provide extra insight in the procedure of feature extraction. For each different layer of the model, different features were extracted. In Fig. 3.8, the feature maps of convolutional layers 1, 4 and 7 of VGG16 are shown. In Fig. 3.9, we show the feature maps of the convolutional layer in stage 1 and three bottlenecks in stage 2 of ResNet50. The structure of ResNet50 can be divided into five stages [118]. Stage 1 consists of 1 convolutional layer and stage 2-5 consist of a different number of bottleneck structures. From both feature maps, we can conclude that, in the first several convolutional layers, basic pollen features such as edges and textures (surface ornamentation) are clearly displayed as was also found in [96]. With an increase of the number of network layers, more and more complex and abstract features influence the performance of pollen classification. For example, in convolutional layer 4 of VGG16 and the first bottleneck in stage 2 of ResNet50, other important parts of pollen such as the pores are highlighted. In the higher layers of the network, only the most representative features are retained but these features are difficult to grasp. With the help of a deep convolutional network that can extract different features from low level (detail) to high level (abstract), the best score in pollen classification tasks is achieved.

In addition, all of the techniques, i.e., transfer learning, data augmentation and hard voting, clearly contributed to improve the performance of the deep learning models under study. Table 3.6 shows to what extent the accuracy can be improved by different techniques applied on the around 1000-sized image subset. Five 1000-sized image subsets were selected via 5-fold cross-validation. The average performance of five subsets was calculated and the results are shown in Table 3.6. The first row shows that ResNet50 achieved 81.4% accuracy if the model was trained from scratch. Transfer learning improved the accuracy to 95.0% using pre-trained parameters which were trained on the ImageNet dataset. Based on the

[1]Two differently sized groups of subsets consist of 1000- and 500-sized image subsets, they are given as an indication; the real number is slightly higher.

95.0% accuracy of transfer learning, the ResNet50 model with data augmentation improved the accuracy to 96.2%. The accuracy is 1.2% higher than without data augmentation which is shown in the second row. Data augmentation helped to increase the variety and the size of image data. Hard voting predicted the class labels with the majority votes of different classification models. The combination of all these techniques significantly improved the accuracy of the ResNet50 model to 97.5%. The third row of Table 3.6 shows that without transfer learning, the accuracy of the ResNet50 model was only 86.1%. We can conclude that, in this study, transfer learning plays a more important role in the performance of deep learning models comparing with data augmentation and hard voting. Because of these advanced techniques, we achieved great success with our deep learning models in the classification of our Urticaceae pollen data. Appendix B: Supplementary Table S6 shows the ablation study of ResNet50 based on 10-fold cross-validation selection method.

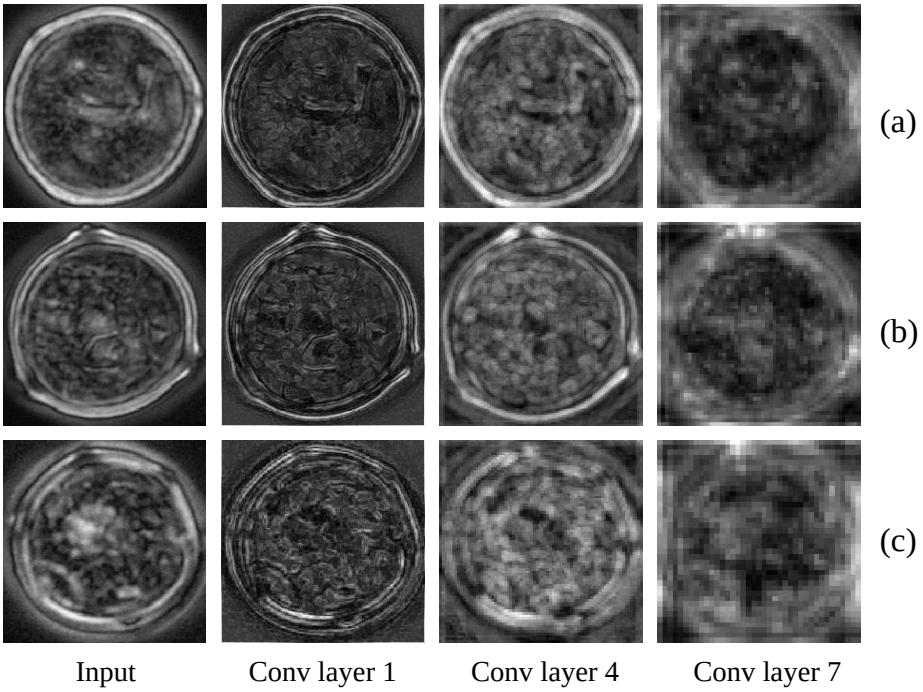

Fig. 3.8 Example of feature maps of VGG16. (a) *Parietaria*. (b) *Urtica*. (c) *Urtica membranacea*. Column 1 represents the input data, column 2-4 are the output of convolutional layer 1, 4, and 7, respectively.

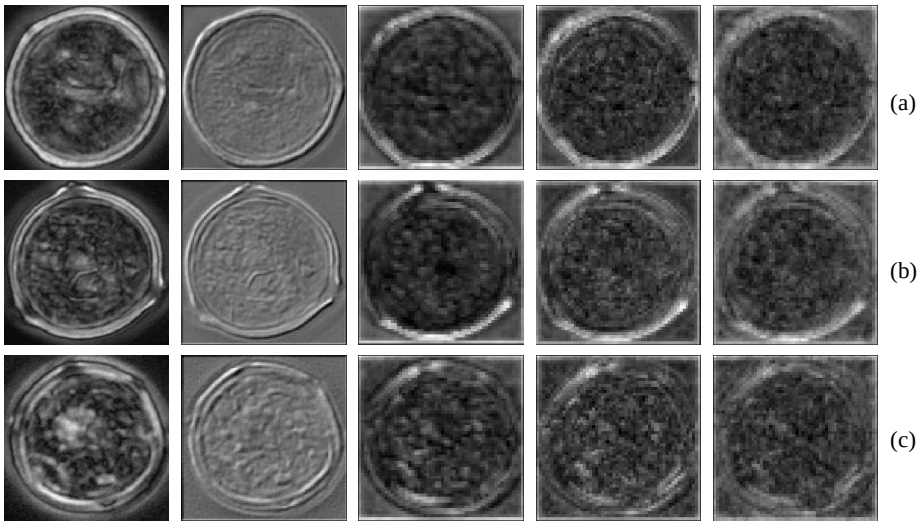

Fig. 3.9 Example of feature maps of ResNet50. (a) *Parietaria*. (b) *Urtica*. (c) *Urtica membranacea*. Column 1 represents the input data, column 2-5 are the output of convolutional layer in stage 1, and output of three bottlenecks in stage 2, of ResNet50, respectively.

Table 3.6 Ablation study with ResNet50. The average performance of ResNet50 based on five (about) 1000-sized image subsets via 5-fold cross-validation selection method is given. Standard deviation, of five subsets, is given in brackets. Numbers in italics refer to training without transfer learning and data augmentation, respectively.

	Training from scratch	**With/*without* transfer learning**	**With/*without* data augmentation**	**With hard voting**
Accuracy	0.814 (±0.025)	0.950 (±0.017)	0.962 (±0.004)	0.975 (±0.002)
	0.814 (±0.025)	0.950 (±0.017)	*0.950 (±0.017)*	0.971 (±0.022)
	0.814 (±0.025)	*0.814 (±0.025)*	0.837 (±0.026)	0.861 (±0.023)

3.4 Conclusions

This study aimed to find the automatic classification model with the best performance to classify Urticaceae pollen grains. Pollen grains of this family have high morphological similarity while they induce different allergenic levels. Few researchers focused on classification of pollen of the Urticaceae nettle family to genus and species level. For our research, a pollen

grain image dataset of the Urticaceae family was constructed, consisting of 6472 images. The pollen grains were unacetolyzed and to our knowledge, these had not yet been used before the analysis of pollen image classification tasks except for our own previous work [96]. Two approaches in image classification techniques including machine learning-based methods and deep learning-based methods were implemented and analysed. For machine learning-based methods, six texture features and two moment features were extracted. Subsequently, several popular feature selection techniques and classifiers were applied. Compared with flat classification models, a hierarchical strategy was confirmed to achieve great success with the classification task. Among the different machine learning methods, the highest performance of 94.5% accuracy was achieved by hierarchical classification models. For deep learning-based methods, six well-established deep Convolutional Neural Networks were used to perform a classification task. Together with data augmentation, cross validation and hard voting techniques, the pretrained ResNet50 model, which achieved an accuracy of 99.4% (±0.002) was considered the best classification model among the six models investigated.

From our comparison of machine learning-based with deep learning-based methods, we conclude that deep learning-based methods perform better for pollen image classification. Two additional experiments demonstrated that deep learning models are more successful for both large and smaller sized datasets. One reason is that deep learning models can extract more representative features of pollen images from low (detailed) to high (abstract) level. The performance of machine learning methods is, however, highly dependent on the quality of features that are extracted from the image dataset. In addition, transfer learning, data augmentation and hard voting techniques drastically improved the performance of deep learning models. An ablation study showed that the accuracy of deep learning models is improving step by step. Deep nets such as Inception-V3, DenseNets, and NASNets have shown to perform well on datasets in the public domain. Nevertheless, ResNet50 has already yield an accuracy of 99.4% on our dataset. We may apply these deeper networks on a larger dataset in the future. Our work clearly demonstrates what automatic classification methods can accomplish for highly similar images of pollen species in the Urticaceae family. This technique can be broader applied to similar pollen from other families. This method could also potentially be extended to cope with other image classification tasks.

CRediT Authorship Contribution Statement

Chen Li: Conceptualization, Investigation, Methodology, Validation, Writing-Original Draft **Marcel polling:** Conceptualization, Investigation, Resources, Writing-Review Editing **Lu Cao:** Conceptualization, Investigation, Writing-Review Editing, Supervision **Barbara**

Gravendeel: Resources, Writing-Review Editing **Fons J. Verbeek:** Conceptualization, Writing-Review Editing, Supervision

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the Chinese Scholarship Council through Leiden University and the European Union's Horizon 2020 research and innovation programme under H2020 MSCA-ITN-ETN Grant agreement No 765000 Plant.ID. We would like thank Xiaoqin Tang, Zhan Xiong, Shima Javanmardi and other colleagues who assist us through feedback during discussions and meetings.