Deductive automated pollen classification in environmental samples via exploratory deep learning and imaging flow cytometry

Claire M. Barnes[1]* (iD), Ann L. Power[2]* (iD), Daniel G. Barber[3] (iD), Richard K. Tennant[3] (iD), Richard T. Jones†, G. Rob Lee[2], Jackie Hatton[3], Angela Elliott[3], Joana Zaragoza-Castells[3], Stephen M. Haley[3] (iD), Huw D. Summers[1] (iD), Minh Doan[4] (iD), Anne E. Carpenter[5] (iD), Paul Rees[1,5] (iD) and John Love[2] (iD)

[1]College of Engineering, Swansea University, Bay Campus, Swansea, SA1 8EN, UK; [2]Biosciences, Faculty of Life and Health Sciences, University of Exeter, Exeter, EX4 4QD, UK; [3]Geography, Faculty of Environment, Science and Economics, University of Exeter, Exeter, EX4 4RJ, UK; [4]Bioimaging Analytics, GlaxoSmithKline, Collegeville, Upper Providence, PA 19426, USA; [5]Imaging Platform, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

Summary

• Pollen and tracheophyte spores are ubiquitous environmental indicators at local and global scales. Palynology is typically performed manually by microscopic analysis; a specialised and time-consuming task limited in taxonomical precision and sampling frequency, therefore restricting data quality used to inform climate change and pollen forecasting models. We build on the growing work using AI (artificial intelligence) for automated pollen classification to design a flexible network that can deal with the uncertainty of broad-scale environmental applications.
• We combined imaging flow cytometry with Guided Deep Learning to identify and accurately categorise pollen in environmental samples; here, pollen grains captured within *c.* 5500 Cal yr BP old lake sediments.
• Our network discriminates not only pollen included in training libraries to the species level but, depending on the sample, can classify previously unseen pollen to the likely phylogenetic order, family and even genus.
• Our approach offers valuable insights into the development of a widely transferable, rapid and accurate exploratory tool for pollen classification in 'real-world' environmental samples with improved accuracy over pure deep learning techniques. This work has the potential to revolutionise many aspects of palynology, allowing a more detailed spatial and temporal understanding of pollen in the environment with improved taxonomical resolution.

Introduction

Pollen and tracheophyte spores are ubiquitous in the environment and present in air, soil and sediments. Accurate classification and quantification of pollen grains is critical for a range of palynological applications (Edwards *et al.*, 2017) including the forecasting of aeroallergens, melissopalynology and forensics. Fossil pollen preserved in sedimentary archives is a powerful proxy used for palaeoenvironmental reconstructions, informing ecological succession and responses to climate change (Edwards *et al.*, 2017; Åkesson *et al.*, 2019), understanding impacts of human activity (Ledger, 2017; Qiu *et al.*, 2020) and monitoring biodiversity (van der Kaars *et al.*, 2017).

Palynology relies on highly skilled analysts to manually classify, identify and quantify intact and partial pollen grains within background debris, using light microscopy. The analysis of individual pollen grains involves characterisation of pollen size, shape, aperture number and type, and surface texture. Gross morphological variations are used to distinguish between pollen types (family level); however, differences at species level become more discrete, complicating the ability of precise taxonomical resolution. Manual classifications are limited to genus or, at times, family level (Makela, 1996), subject to human bias and are time intensive (MacLeod *et al.*, 2010). In palaeoenvironmental studies, 150–500 pollen grains are typically analysed per microscope slide in 2–10 h (Stillman & Flenley, 1996). Consequently, palynological investigations are inherently constrained by sampling size and frequency (Lacourse & May, 2012) and taxonomical precision, limiting the quality of spatio-temporal data used in climate change reconstructions (Lake *et al.*, 2017; Anenberg *et al.*, 2020) and pollen forecasting (Anenberg *et al.*, 2020).

---

*These authors contributed equally to this work.
†This paper is dedicated to the memory of our colleague and dear friend, Dr Richard Jones.

Despite these complications, manual microscopic analysis remains the benchmark method for pollen quantification and classification (Baksay *et al.*, 2020). Existing alternatives have their own challenges. Environmental DNA metabarcoding allows high-resolution, species-level identification; however, it cannot robustly quantify pollen abundance (Bell *et al.*, 2018). Conventional (nonimaging) flow cytometry and coulter counters have been used to rapidly analyse high quantities of pollen grains (Mitsumoto *et al.*, 2009; Tennant *et al.*, 2013; Heidmann *et al.*, 2016), however, lack the detailed morphological data that are the main phenotypic differentiators for pollen.

Recently, deep learning (neural networks) has provided new possibilities for automated pollen classification (Langford *et al.*, 1990; France *et al.*, 2000; Zhang *et al.*, 2004; Mander *et al.*, 2013; Marcos *et al.*, 2015; Daood *et al.*, 2016; Tcheng *et al.*, 2016; Khanzhin *et al.*, 2018). These AI-mediated analyses are accurate, exclude human bias and can, for example, distinguish obscured pollen on microscope slides, as well as damaged or partial grains (Holt *et al.*, 2011; Tello-Mijares & Flores, 2016; Bourel *et al.*, 2020; Olsson *et al.*, 2021). Recent advancements in microscope slide scanning have allowed higher pollen volumes and therefore taxa to be rapidly analysed (Olsson *et al.*, 2021; Punyasena *et al.*, 2022); however, scanning times can vary between instruments, and optimised pollen concentrations mounted onto slides are required so that pollen is sufficiently dispersed for adequate image capture and minimal obscured images.

An alternative to microscope slide scanning is imaging flow cytometry (IFC), a multispectral, data-rich technique that allows rapid acquisition of thousands of images per second. Samples are hydrodynamically focused during analysis; therefore, image captures typically containing one object are generated, preventing object overlap (as experienced using microscope slides), reducing the need for segmentation processing. High-throughput image acquisition makes IFC ideally suited to machine learning and deep learning applications (Doan *et al.*, 2018) since robust training libraries can be generated to train networks.

A target of 40 taxa in a training library has been suggested for meaningful applications in palaeoecology (Holt & Bennett, 2014), although the majority of studies are limited to fewer than a dozen taxa (Langford *et al.*, 1990; Makela, 1996; Bourel *et al.*, 2020). A recent investigation used a convolutional neural network (CNN) to classify 426 876 images from a pollen library containing 35 meadow plant species via IFC at 96% accuracy (Dunker *et al.*, 2020). However, their algorithm was tested on pollen samples used to train their network, limiting its scope for environmental application. Environmental, airborne pollen has been classified by a CNN trained on 46 pollen types (14 153 pollen images) using digitalised microscope slides (Punyasena *et al.*, 2022). Training data were manually annotated, subsequently limiting the network due to human bias and restricting taxonomical precision to genus and family level. A more robust approach was adopted whereby digitalised slides of reference pollen (83 species; 122 000 images) were used to train a CNN (Olsson *et al.*, 2021). Libraries were populated with key pollen species known to be present in bee pollen samples and classifications were determined as pollen types which included a mixture of pollen groups at species, genus and family level (Olsson *et al.*, 2021). Bee pollen is typically pristine, as indicated by the low detection of damaged pollen grains, and highly abundant; therefore, background (non-pollen) images were not classified (Olsson *et al.*, 2021).

The ability to train a network on a reference library of pollen species and then test it on environmental samples containing high levels of uncertainty such as a broad range of species including those independent of training libraries and background objects, is imperative for accelerating exploratory and applied palynology (Edwards *et al.*, 2017). Therefore, our aim was to construct a robust, deductive network that could rapidly:

(**1**) identify pollen against non-pollen (background) objects, inevitably present in environmental samples,
(**2**) classify pollen to species and genus level for pollen types of interest and
(**3**) predict at order, family or genus level the presence of pollen types outside predefined classes present in a reference library.

Consequently, we developed a unique version of deep learning, that we term 'Guided Deep Learning' which combined multi-labels related to pollen morphology and taxonomy, for training images (Fig. 1). This unique approach differs from previous studies that have either adopted purely deductive, classical machine learning techniques (Daood *et al.*, 2016), which require substantial human input and can fail to capture important subtle features crucial for classification, or purely inductive deep learning networks (Sevillano *et al.*, 2020), requiring minimal human input by creating its own rules for classification, and is therefore deemed a 'black box'. Our work sits between these two approaches, by allowing the expert to include labels deemed important for classification whilst also taking advantage of features extracted by the network, essentially, lifting the lid on the black box of deep learning. Moreover, the labels used by our network can be assigned in a weakly supervised manner as opposed to case-by-case annotations, saving time.

Training was split into subtasks that incorporated prior knowledge of pollen morphology with the advantages offered by deep learning to extract subtle features from the IFC images to phylogenetically classify pollen samples. This combination allowed prediction of pollen abundance across various levels of taxonomy, even if species were independent from the training dataset. In a true test, the system successfully classified previously unseen fossil pollen archived in *c.* 5500 Cal yr BP lake sediment, which had been manually classified to genus or family level.

The combination of IFC and flexible AI described here is widely transferable and will potentially revolutionise many aspects of palynology, allowing more precise taxonomical classifications, for example, species level, opposed to family-level detection, and higher volumes of samples to be rapidly analysed, therefore improving our understanding of pollen in the environment. Our network could be applied to enhance data used to inform climate change and pollen forecasting modelling and to provide a deeper understanding of exposure to allergenic pollen types and subsequent health effects.
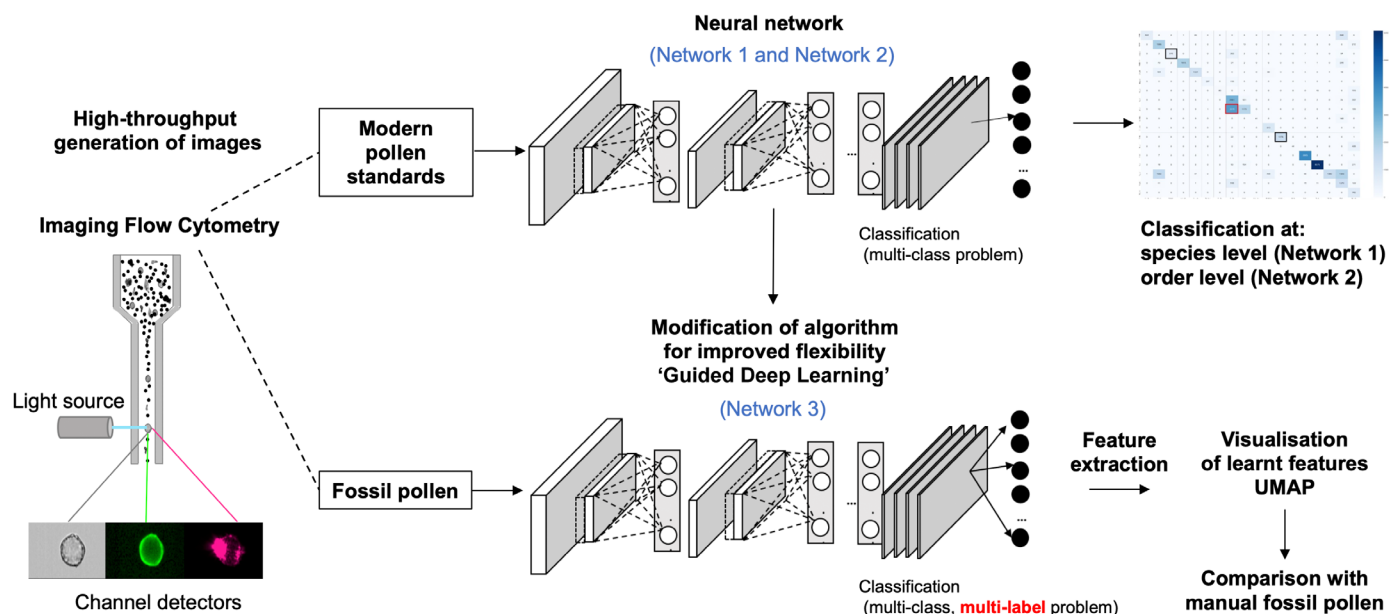
Fig. 1 Flow diagram of our exploratory approach to automated pollen classification combining imaging flow cytometry and deep learning. Imaging flow cytometry was used to analyse modern pollen and fossil pollen. Individual objects within the sediments were hydrodynamically focused and interrogated by a suite of lasers. Brightfield, side scatter and autofluorescence images were captured at specific bandwidths via channel detectors. Modern pollen samples were used to train Deepometry, an open-source workflow based on ResNet50 (Doan *et al.*, 2021). A multi-classification algorithm was initially implemented (Network 1), trained on images split into classes depending on their species. A second network (Network 2) was trained on images classified at order level. The accuracy of both networks was tested on hold-out data, and confusion matrices were used to understand misclassifications. The networks were also trained to identify and extract 'background' or 'noise' events. A more flexible approach was necessary to carry out exploratory work on fossil pollen. Therefore, Network 3 was modified to perform multi-classification, multi-label training. Images present in the modern samples were assigned labels according to their taxonomy and overall morphology and used to train the modified network. Features learnt by the neural network were extracted from the training images and a supervised method was used to perform dimension reduction and visualise features in 2D as a UMAP (Uniform Manifold Approximation and Projection) plot. Support vector machine learning was then used to learn classifications of test images based on their positions in feature space.

Materials and Methods

Modern pollen reference sample preparation

Pollen standards from 53 plant species (Allergon AB Thermo Fisher Scientific, Ängelholm, Sweden) were analysed for a broad range of European and Northern American angiosperms and gymnosperms (Table 1). Approximately 0.1 g of pollen material was added to 40 ml of Mili Q water, vortexed, heated at 90°C for 1 h to hydrate pollen grains and remove air within the sample, and sieved at 100 μm (Corning nylon cell strainer) to prevent blockage of the IFC instrument.

Palaeoenvironmental sample preparation

Sediment cores were extracted from Mere Tarn, Cumbria, UK (54°08′12.4″N, 03°07′26.0″W), and extruded into 1 cm intervals. Known amounts of *Lycopodium* spp. grains (Lund University, Sweden) were added to samples to enable estimates of absolute concentrations (Stockmarr, 1971). Pollen grains were extracted from background lake sediment via sequential digestion using HCl, KOH and HF at 90°C, acetolysis and physical separation by sieving to collect particulate 10–106 μm in size. Samples were stored in glycerol. Before IFC analysis, glycerol was washed from the palaeoenvironmental samples via resuspension of 1 ml of sample in Milli Q, centrifugation at 11 337 × *g* for 5 min, removal of the supernatant and final storage in 1 ml of Milli Q.

Manual identification of fossil pollen in a palaeoenvironmental sample

Pollen grains were manually identified and counted using light microscopy at 400× magnification, aided by a range of pollen reference material and keys.

Imaging flow cytometry (IFC) data acquisition

Data were acquired using a fully calibrated (ASSIST tool) ImageStream X MkII (ISX; Luminex Corp., Seattle, WA, USA) configured with a single camera and 405, 488, 561, 642 and 785 nm excitation lasers, brightfield (BF) illumination, multi magnification (20×, 40× and 60×) and a six-channel detection system.

A hydrodynamically focused stream allowed interrogation of individual objects by a suite of lasers and BF illumination. Light emitted from each object was detected on a series of six channels (Ch01, Ch02, Ch03, Ch04, Ch05 and Ch06), each relating to a specific bandwidth. Brightfield images were collected on Ch04 (BF, 610/30 nm), side scatter on Ch06 (SSC, 762/35 nm) and

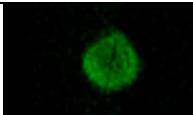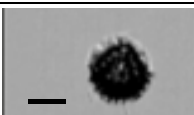**Table 1** Reference library of pollen and spore samples used to train the deep learning network.
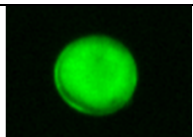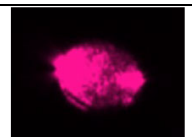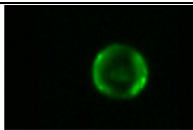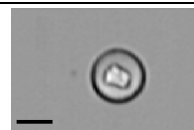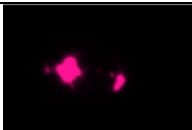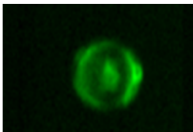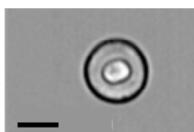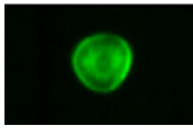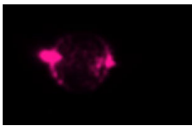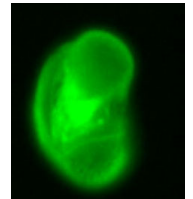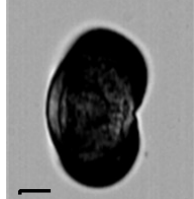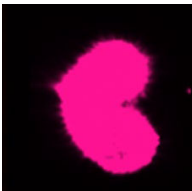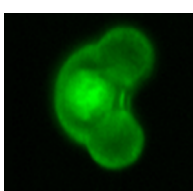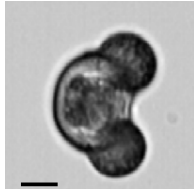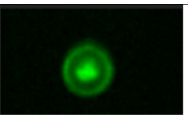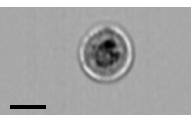
Order	Family	Species	Autofluorescence	Brightfield	Side scatter	*n*
Tracheophyta – Lycopodiophyta						
Lycopodiales	Lycopodiaceae	*Lycopodium.* spp.*				210
Tracheophyta – Polypodiophyta						
Equisetales	Equisetaceae	*Equisetum arvense* L.				1519
Tracheophyta – Gymnospermae						
Pinales	Cupressaceae	*Cupressus arizonica* Greene.				2670
		Cupressus sempervirens L.				2668
		Taxodium distichum L.				5186
	Pinaceae	*Abies alba* Mill.				391
		Pinus echinata Mill.				2532
		Pinus sylvestris L.				838
	Taxaceae	*Taxus baccata* L.				6034

Table 1 (Continued).
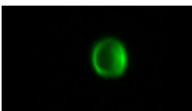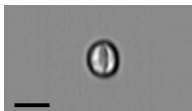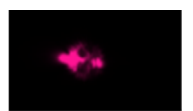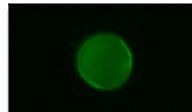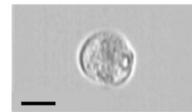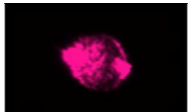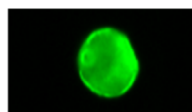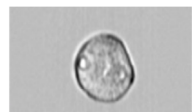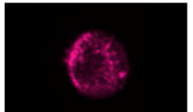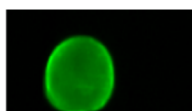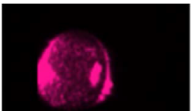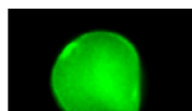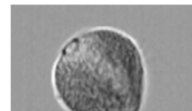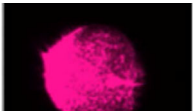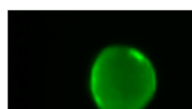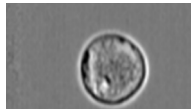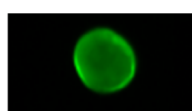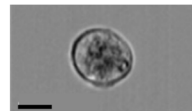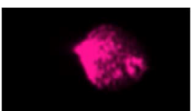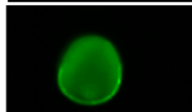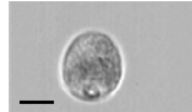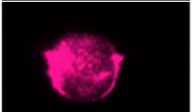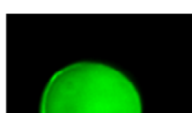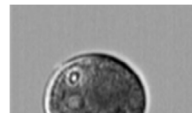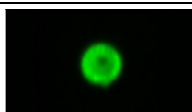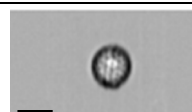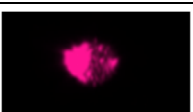
Order	Family	Species	Autofluorescence	Brightfield	Side scatter	*n*
Tracheophyta – Angiospermae – Monocots						
Arecales	Arecaceae	*Phoenix dactylifera* L.				3213
Poales	Poaceae	*Agrostis capillaris* L.				4753
		Agrostis gigantea Roth.				2600
		Anthoxanthum odoratum L.				2815
		Avena fatua L.				1953
		*Cynodon dactylon (*L.) Pers.				5171
		Holcus lanatus L.				5191
		Lolium perenne L.				4348
		Triticum aestivum L.				3739
Tracheophyta – Angiospermae – Eudicots						
		Ambrosia artemisiifolia L.				3752

Table 1 (Continued).
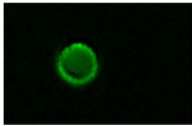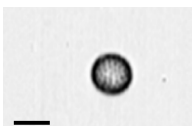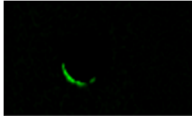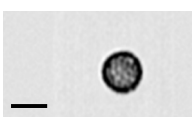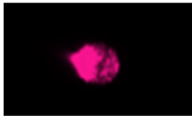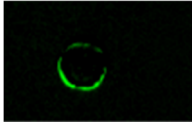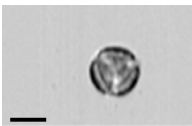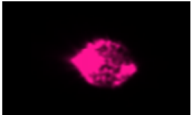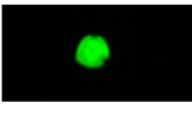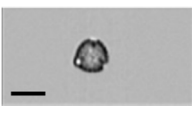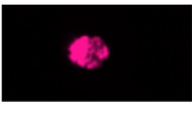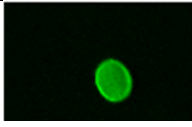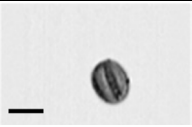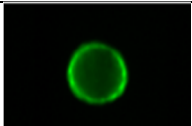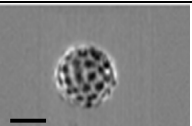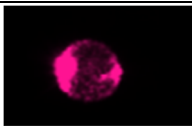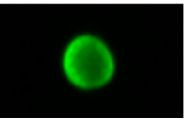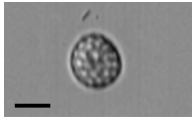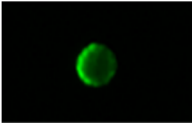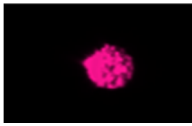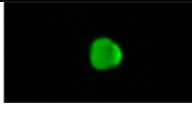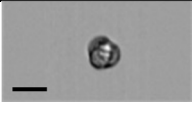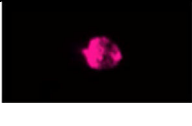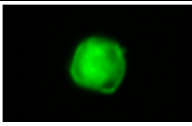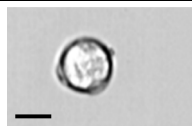
Order	Family	Species	Autofluorescence	Brightfield	Side scatter	n
Asterales	Asteraceae	*Ambrosia psilostachya* DC.				2530
		Ambrosia trifida L.				2808
		Artemisia vulgaris L.				4885
		Iva xanthifolia Nutt.				5307
Brassicales	Resedaceae	*Reseda odorata* L.				2333
Caryophyllales	Amaranthaceae	*Amaranthus retroflexus* L.				5898
		Chenopodium album L.				1128
		Salsola kali L.				3820
Dipsacales	Adoxaceae	*Sambucus nigra* L.				6447
Fabales	Fabaceae	*Vicia sativa* L.				403
		Alnus incana (L.) Moench.				4155

Table 1 (Continued).
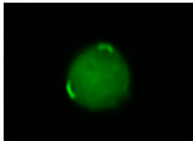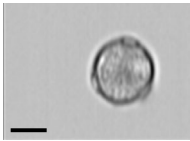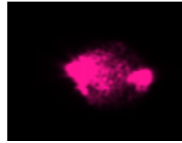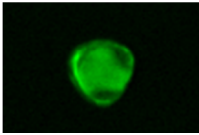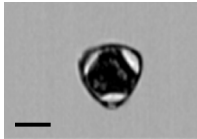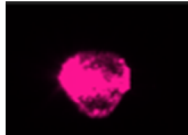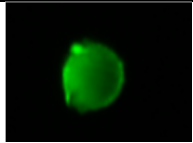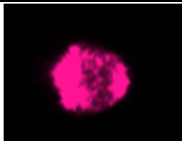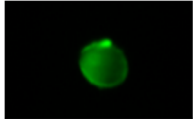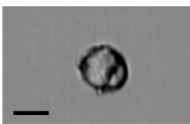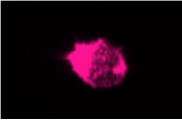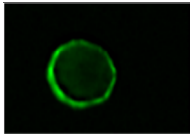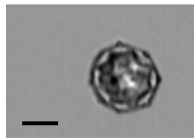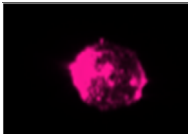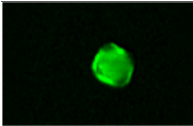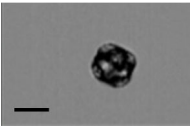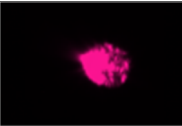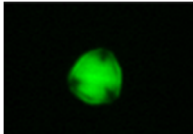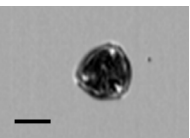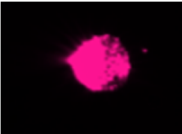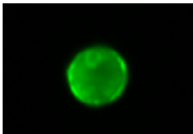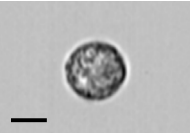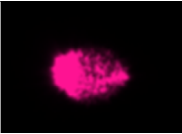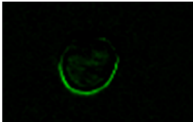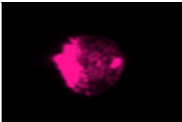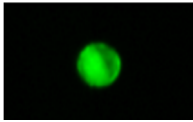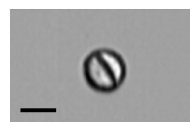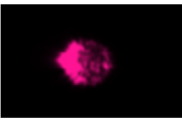
Order	Family	Species	Autofluorescence	Brightfield	Side scatter	*n*
Fagales	Betulaceae	*Betula pendula* Roth.				2224
		Corylus avellana L.				8150
	Fagaceae	*Quercus alba* L.				6486
		Quercus robur L.				7850
	Juglandaceae	*Juglans californica* S. Watson.				8011
Lamiales	Oleaceae	*Fraxinus americana* L.				4264
		Fraxinus excelsior L.				2794
	Plantaginaceae	*Plantago lanceolata* L.				3379
Malpighiales	Salicaceae	*Populus alba* L.				5044
		Salix viminalis L.				2431

Table 1 (Continued).
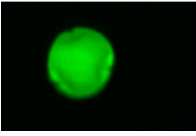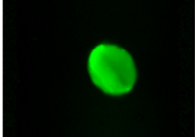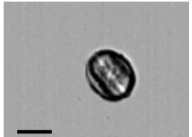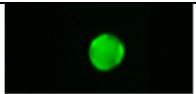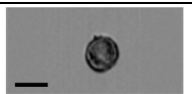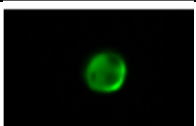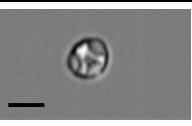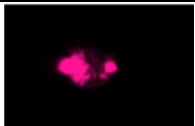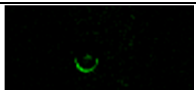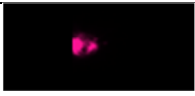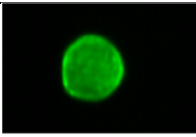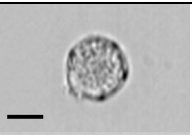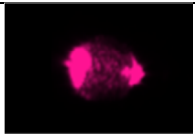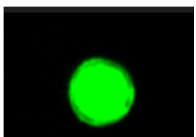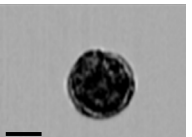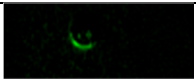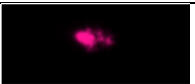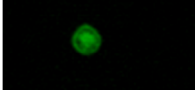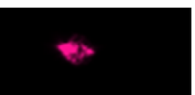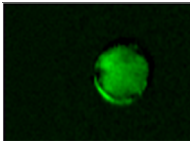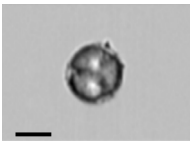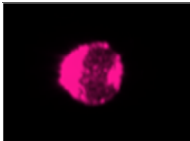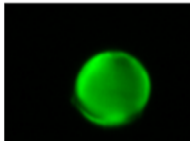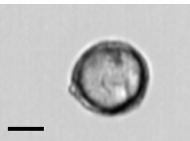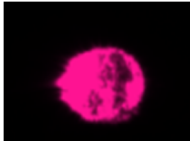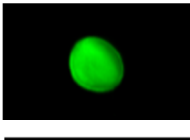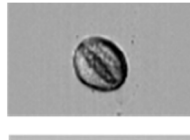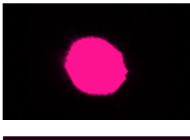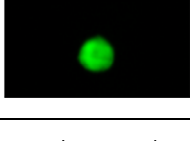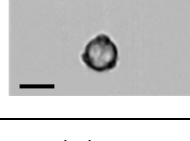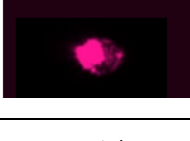
Order	Family	Species	Autofluorescence	Brightfield	Side scatter	*n*
Malvales	Malvaceae	*Tilia cordata* Mill.				6958
Myrtales	Onagraceae	*Chamaenerion angustifolium* (L.) Holub.				117
Proteales	Platanaceae	*Platanus x hispanica* Mill.				6935
Rosales	Cannabaceae	*Humulus lupulus* L.				3592
	Moraceae	*Broussonetia papyrifera* (L.) L'Her. Ex Vent.				1246
	Ulmaceae	*Ulmus americana* L.				3972
		Ulmus glabra L.				4172
	Urticaceae	*Parietaria officinalis* L.				4393
		Urtica dioica L.				5612

Table 1 (Continued).

Order	Family	Species	Autofluorescence	Brightfield	Side scatter	*n*
Sapindales	Sapindaceae	*Acer negundo* L.				5286
		Acer pseudoplatanus L.				3617
		Acer rubrum L.				2315
		Aesculus hippocastanum L.				404

Exemplar pollen grains from each of the 53 plant species used to train the network are presented, showing autofluorescence (Ch02: 533/55 nm), brightfield (BF: 610/30 nm) and side scatter (SSC: 762/35 nm) images. To test the ability of the network to identify a range of pollen types, the library included species from 18 orders and 26 families to encompass a wide taxonomical range and varied morphological features including shape (spheroidal, oval and bisaccate), aperture numbers (0–> 6) and aperture types (e.g. colporus, porus and leptoma ulcus; Supporting Information Table S1). To test the prediction of pollen at both high and low taxonomical resolution, we also included multiple species within the same genus. Our library was also tailored to represent some of the most abundant pollen types identified manually in the environmental sample. The number of pollen grains analysed for each species is reported, totalling 595 647 images from 198 549 pollen grains. Bars, 20 μm.
*The *Lycopodium* spp. standard was derived from a tablet routinely used to spike palaeolimnological samples as a counting estimate.

autofluorescence on Ch01 (457/45 nm), Ch02 (533/55 nm), Ch03 (577/36 nm) and Ch05 (702/85 nm).

For maximum resolution and high sensitivity, fluidics were set at low speed. Magnification was set at 20× (1 μm^2 pixel resolution) for optimum image capture of objects up to 120 μm in length to encompass a wide range of pollen types. Brightfield illumination and excitation lasers 488 nm (at 1.00 mW) and 785 nm (at 2.03 mW) were applied to determine autofluorescence and side scatter, respectively.

Sample volumes of 200 μl were prepared for analysis. For the modern pollen standards, images were collected until *c.* 1000–20 000 objects of interest were analysed or until acquisition reached 30 min (sample volume analysed ranged from 42.65 μl (*n* = 558; *Abies alba* Mill.) to 3.66 μl; *n* = 8150; *Corylus avellana* L.). Pollen grains were visually isolated from background material based on their high SSC and fluorescence properties.

For the palaeoecological samples, a similar acquisition template was applied to gate the large, autofluorescent population within the sample that contained pollen. Data were acquired for a duration of 30 min to obtain sufficient fossil pollen and analysed in replicate (*n* = 4). Approximately 14–47 μl of sample was analysed. To determine any potential batch effect, we analysed five different reference pollen types over subsequent days (Supporting Information Fig. S1). A compensation matrix was applied to adjust for spectral overlap between

channels, calculated from data acquired excluding BF and SSC excitation.

Imaging flow cytometry (IFC) data analysis

Postacquisition, data analysis was performed using IDEAS® software (v.6.2; EMD Millipore, Seattle, WA, USA). To ensure high-quality data, background objects, unfocused images and images containing multiple objects were excluded from the 53 modern pollen reference samples by setting thresholds for selected features (Fig. S2). Variations in brightfield pollen images (focus and rotation) are demonstrated in Fig. S3.

Autofluorescence (Ch01) vs SSC (Ch06) biplots were used to identify pollen grains against background, non-fluorescent and small particulate contained within the standardised pollen extracts (Fig. S2). Compensated image files containing data from all six channels were extracted; however, we selected BF, autofluorescence detected by Ch02 and SSC images to train our deep learning networks (Table 1) since they yielded the highest classification accuracies on hold-out data.

Fossil pollen was visually isolated (postdata acquisition) from nonpollen, minerogenic particulate in the environmental samples based on the high SSC and fluorescence properties of pollen (Fig. S4). Of the most fluorescent particulate, those with an aspect ratio > 0.35 contained pollen grains (Fig. S4D) and were

analysed by the network, which was also trained to identify background events, by including nonpollen populations such as speed beads (small 1 μm plastic beads used during image acquisition to focus the IFC camera), minerogenic particulate from the sedimentary matrix, fluorescent but nonpollen particulate, and unidentifiable background events.

Deep learning

Dataset Our initial dataset was compiled from analysis of 198 549 pollen grains with three channel (BF, Ch02 and SSC) images (595 647 images in total) across 58 classifications which included 53 pollen species (Table 1) and five background types. A training/validation/test split of 70/20/10 was applied. The test (hold-out) data were not used to inform training in anyway. All confusion plots presented were based on hold-out data to assess the accuracy of the network after training and to detect overfitting. A modified workflow was subsequently developed and tested on palaeoenvironmental samples to classify fossil pollen.

IFC had been used to broadly gate the pollen from the palaeoenvironmental samples on which we ultimately tested our network (Fig. S4); however, background events (e.g. minerogenic and biogenic particulate) were still present and so our network was trained to recognise a 'nonpollen or background' classification by presenting examples of these images to the network during training. Preprocessing of the data included normalisation of images and centre cropping along with augmentation of images in the form of resizing, rotation and translation to encourage a robust approach and guard against overfitting.

Our network also mitigates for potential batch effects from intrainstrument variability. Five reference pollen species were also prepared using acid digest techniques, replicating the fossil pollen treatment, to assess effects from sample preparation (Fig. S1).

Network Our network used a modification of Deepometry, an open-source workflow based on ResNet50, designed to apply both supervised and Guided Deep Learning analysis to IFC data (Doan *et al.*, 2021), however, was not pretrained. ResNet-50 is a deep neural network, which applies residual learning to image classification. Deepometry formed the basis of our workflow, as was deemed appropriate for the complex task of learning at different levels of the phylogenetic tree.

Residual convolutional networks have been previously employed successfully to classify images for a wide range of applications (Dhungel *et al.*, 2017; Sarwinda *et al.*, 2021; C. Li *et al.*, 2022). These types of networks use convolutional layers to learn features from images. Skip connections are used to jump layers, improving gradient flow through the network during backpropagation. These skip connections help to deal with issues associated with training deeper networks, for example vanishing gradients, whilst yielding the benefits of using the additional layers to extract more complex features from images.

Feature extraction Our workflow used the network to classify images against a multi-labelling system, whereby each image can be assigned a number of labels. This guided the network's learning process to key parts of feature space, deemed important yet robust by the expert. The network is made up of multiple layers and, due to the application of a multi-labelling classification system, is trained with a complex loss function (which may be tuned for a range of purposes, giving higher weighting to some feature labels over others). The network learns which are the most informative features during the training stage. Due to the increased number of layers, the network is capable of learning more specialised and complex features (likely to be specific to a particular genus or species of pollen) at the deeper levels of the network along with more general features such as size and edge formation, extracted by the top layers (Radovanović & Đukanović, 2020; Tan *et al.*, 2021). Such an approach gives the workflow necessary flexibility to handle a mixture of testing pollen made up of previously seen pollen and potentially new pollen only linked to the training data at higher levels of taxonomy.

The Guided Deep Learning approach we propose can be applied in place of strongly supervised techniques. Moreover, as an alternative to manually annotating each image used to train the network, labels were assigned to subsets of images and therefore not typically assigned ground truths. This form of weak supervision has been found to yield important advantages, not least the time saved when preparing training data (Xu *et al.*, 2014; Yao *et al.*, 2016; Y. Li *et al.*, 2022).

For the flexible part of our network, we used a multi-classification, multi-label approach to train the network. Images were assigned multiple labels in a weakly supervised way linked to their taxonomy, for example, order and family labels along with labels based on morphological features (Table S1). Whilst training the network, each classification made up its own subtask and was treated as a binary classification problem. A cross-entropy loss function is used by the network to combine performance across the various subtasks and inform the weights of the different layers of the network:

$$\text{BCE}_{1\ \text{(large in size)}} = -\big(p(x)\log(q(x)) + (1-p(x))\log(1-q(x))\big)$$

$$\text{BCE}_{2\ \text{(spherical in shape)}} = -\big(p(x)\log(q(x)) \\ + (1-p(x))\log(1-q(x))\big)$$

$$\text{BCE}_{N\ \text{(Taxonomy Order Pinales)}} = -\big(p(x)\log(q(x)) \\ + (1-p(x))\log((1-q(x)))\big)$$

where x is the classification, p is the probability of class x in the target vector, and q is the probability of class x in the prediction vector.

The loss functions for each subtask were summed together for all images in a batch and used to update the weights of the network. Each image may have up to 110 different assigned classifications to allow the network to learn more flexible features that could scale well across the different taxonomy levels. Whilst mutually exclusive implementations were trialled, this network with the additional flexibility, yielded superior performance and high accuracy when tested on a typical deep learning hold-out task (Fig. S5).

Clustering techniques Clustering and visualisation techniques were applied to features extracted from images by the network. Features learnt by the network were extracted from the global pooling layer by activations. This layer pools input features over all spatial locations and summarises information learnt by the whole network. These features were subsequently visualised for further analysis.

Features learnt by the network were fully analysed for exploratory purposes. The ability to visualise high-dimensional features learnt during training in 2 or 3 D offered important advantages, particularly when dealing with some of the challenges presented by our test data, composed of both previously seen and unseen classifications. Uniform Manifold Approximation and Projection (UMAP), a dimension reduction technique, was applied (Stolarek *et al.*, 2022). Given high-dimensional data, UMAP produces a lower-dimensional representation for purposes of data visualisation and exploration. This technique can be used for general nonlinear dimension reduction (McInnes *et al.*, 2018) and allows manifolds in feature space to be modelled with a fuzzy topological structure. Moreover, whilst an unsupervised approach can be taken, topological structures were found and plotted based on external labels or classifications such as pollen order or genus. This technique was applied to features extracted from the images by the network trained with our modified version of weak supervision. UMAP visualisations were produced based on labels linked to the order, family, genus and species levels of the taxonomy, which allowed us to interrogate the data for clusters and compare these against classification labels at different levels of the phylogenetic tree. Support vector machine learning was then used to classify images based on these representations in low-dimensional space. Whilst this provided classifications to nearest neighbours in features space, levels of uncertainty were also determined and easily visualised for further discussion and exploration.

Support vector machine learning Support vector machine learning (SVM) was used to find optimised hyperplanes in N-dimensional space that distinctly classify the data points. In our case, this technique was applied to our UMAP representations of features learnt by the network. Supervised learning according to various taxonomy labels was used and the models generated were tested on our hold-out data and palaeoenvironmental samples.

Results

Testing the accuracy of the deep learning network on training data

A reference library of 53 modern pollen species (Table 1) and background (nonpollen images) was prepared and manually annotated by an expert (Table S1). We trained an initial network (Network 1), which identified pollen with 93% accuracy at the species level (Figs 2, S6) and successfully identified background noise, when tested on hold-out data.

Network 1 was trained to extract specifically defined features to assign species classification. However, to develop a flexible system to predict unseen pollen that belongs to the same family or order as species used to train the network, a broader classification at higher taxonomy is required. Therefore, we trained Network 2 to identify pollen at order level. The lower accuracy (67.9% on hold-out data made up of species previously seen by the network) compared with Network 1 (Figs 3, S7) reflects feature variability of species within higher taxonomical levels. For example, some orders were well populated in the training library (e.g. Poales), whilst others (e.g. Malvales) were represented by only one species. Therefore, when tested on hold-out data, high levels of accuracy were achieved for those classifications composed of a single species or several highly similar species. Classifications containing multiple genera or species, particularly those from different families, showed higher heterogeneity within the training data and, therefore, reduced accuracy (e.g. Rosales and Fagales), highlighting the challenges presented by a dataset with different classification imbalances at different levels of taxonomy.

Although higher accuracies were achieved at species level by Network 1, which learnt specific features from the images for the task of classification, lower accuracy results from Network 2 suggested that more general, flexible features were required to predict pollen types at higher levels of taxonomy. A flexible approach is crucial for environmental applications, where unseen pollen types will be inevitable. We therefore designed a guided version of deep learning that focused on robust features performing well to identify pollen across multiple levels of taxonomy, namely genus, family and order levels (Table 2).

Our Guided Deep Learning utilised broad (species-wide) labels of morphological features and taxonomy, to train Network 3 which then extracted features from images. UMAP was applied to this resulting high-dimensional data to visualise a lower-dimensional representation of these features (Fig. 4a). UMAP is a popular clustering technique that may be used in a supervised or unsupervised way. Here, the placement of manifolds (topological space in which each species is plotted) is guided by order labels via supervised learning to minimise intravariation of a particular order whilst maximising intervariations. Exemplar pollen images (Fig. 4b) demonstrate how orders are represented in the training set. Whilst Network 3 added flexibility, it did not compromise on accuracy when performing typical deep learning tasks with *c.* 96% classification accuracy of the hold-out data noted (Fig. S5).

Classification of fossil pollen

To demonstrate the applicability of Network 3 to environmental samples, it was used to classify fossil pollen grains deposited within *c.* 5500 Cal yr BP lake sediments (Fig. 5). Results were compared with classifications by a human analyst (Fig. 6). These manual counts were performed as standard practice, at genus level or, for some pollen types, at family level (Poaceae and Cyperaceae) due to the laborious and difficult task of pollen identification. The network successfully detected the most abundant pollen types within the environmental sample such as *Corylus*, *Quercus* and *Alnus*, as well as relatively rare pollen grains such as *Pinus* (Fig. 6a; Table 3). These key pollen types were represented in the training library and the UMAP demonstrates that fossil
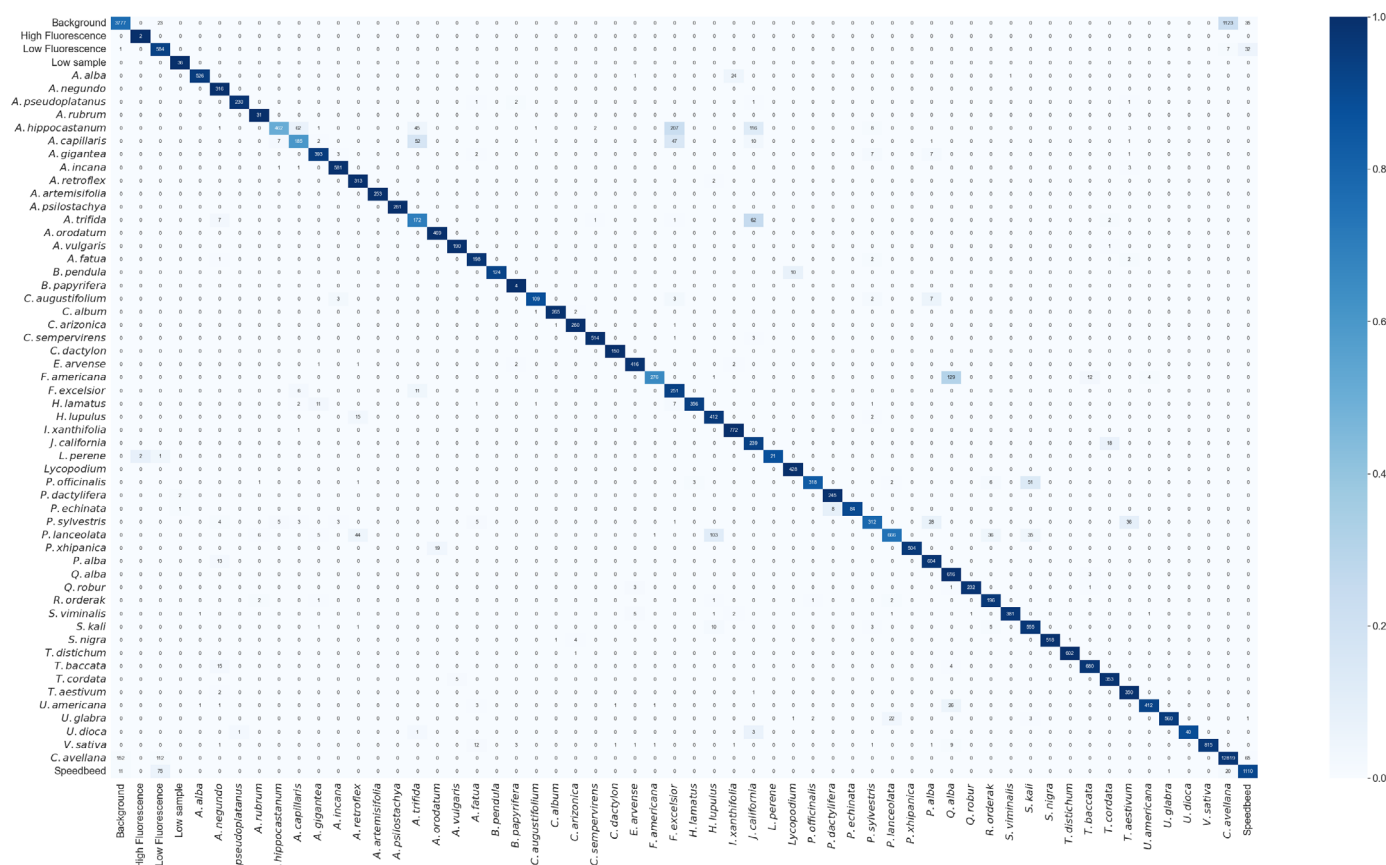
Fig. 2 Classification accuracy for modern pollen standards at species level using Network 1. In this confusion matrix, the number of correctly identified images is given in the corresponding rows and columns. Quantities outside of the diagonal represent misclassifications. The dataset was split into training/validation/test images (70/20/10). Deepometry was trained by a fully supervised method to label images according to their species and tested on hold-out data, achieving an overall instance-level accuracy of 93%. Matrices depicting class-level accuracies are presented in Supporting Information Fig. S6.

pollen closely clustered with the modern reference images (Fig. 5). The network performed very well at order level classification (Fig. 6c), demonstrating the ability of the network to detect at higher taxonomy the presence of pollen types not represented in the training library at genus level.

The prediction of pollen types at order, family and genus level by Network 3 when compared to manual counts revealed that 100% of order, 77% of family and 96% of genus types were successfully identified. For meaningful comparisons, all taxonomical labels identified manually but not present in our training set were removed.

Not represented in the reference library but identified by a human analyst were *Polypodium*, *Sphagnum*, *Hedera* and *Calluna*. As the Network was tasked with classifying all images, these species will account for some misclassifications. It is important to also note that in general, deep learning performs best when classifying high-quality images, in cases where networks successfully deal with noise, this is often accounted for during training. Network 3, however, was trained on standard images and tested on data that included some low-quality, noisy data from the environment, which may also account for some of the misclassifications. However, the UMAP visualisation tool clearly indicates the

presence of pollen that fail to cluster with those represented during training (Fig. 5).

Our workflow also outputs the probability that an image belongs to a particular class. An average probability of 0.9714 was achieved for the fossil pollen classifications. Probabilities less than 0.95 indicate a previously unseen order of pollen failing to tightly cluster with training images. In the palaeoenvironmental sample, 9.38% of pollen grains achieved probabilities of less than 0.95. In a practical setting, the probability scores along with the UMAP visualisation tool would highlight an unseen pollen type within the sample, indicating the need for further interrogation. It is important to note that our workflow initially removes images deemed as background noise or unsuitable for classification due to poor quality, potentially removing some low-quality pollen images at this stage.

We also examined the effectiveness of the labels used to train Network 3 for classification purposes (Fig. S8) and found that both taxonomy and morphology labels were highly weighted and thus important to classify pollen at order, family and genus levels. Moreover, this result demonstrated the need for a deep learning, as opposed to a classical machine learning approach, to incorporate both morphological and taxonomical labels.
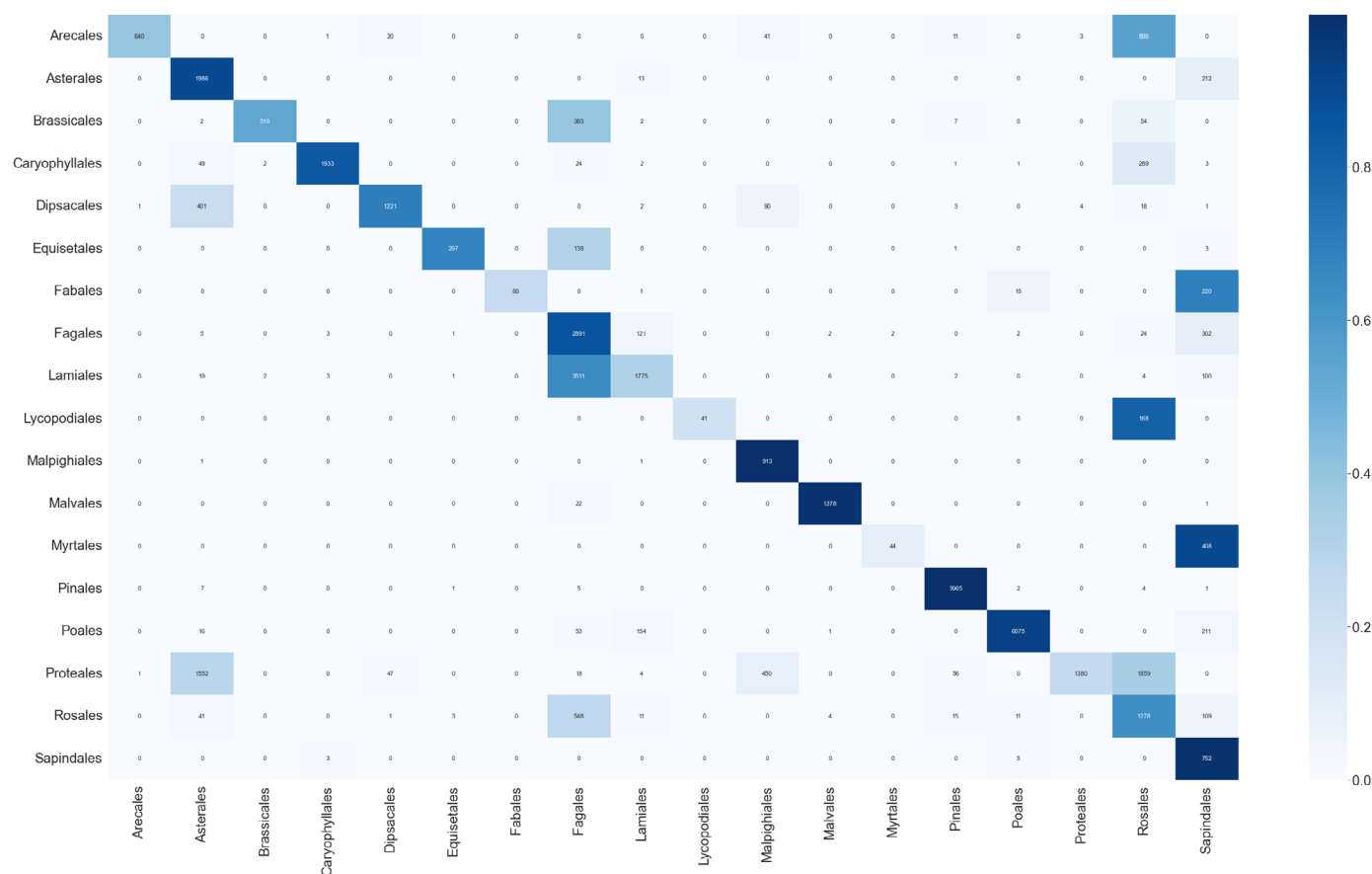
Fig. 3 Confusion matrix for modern pollen standards at order level using Network 2. The number of correctly identified images are given in the corresponding rows and columns. Quantities off-diagonal are misclassifications. All background images were removed from the training data and the remaining images were split into training/validation/test images of 70/20/10, achieving an overall instance-level accuracy of 67.9%. Matrices depicting class-level accuracies are presented in Supporting Information Fig. S7.

Discussion

We designed a flexible network that can rapidly analyse and accurately classify a broad range of pollen in 'real world' environmental samples. We applied high-throughput images acquired via IFC to a modified Guided Deep Learning network to accurately categorise pollen from sediments *c.* 5500 Cal yr BP. Previously unseen pollen species were either identified at higher levels of taxonomy (genus, family or order) or noted as belonging to an unrepresented pollen type based on our visualisation tool and output statistics.

Fundamental to the success of our network are the highly robust training libraries of pollen images we acquired via IFC. Consisting of 595 647 brightfield, autofluorescence and side scatter images from 198 549 individual modern pollen grains, the library incorporated 53 pollen types from 26 families, spanning 18 orders (Table 1). The high-throughput, rapid analysis of pollen performed by IFC is impossible to achieve via human analysts and is unparalleled by other techniques including light microscopy (Langford *et al.*, 1990; Khanzhin *et al.*, 2018), SEM (Mander *et al.*, 2013; Daood *et al.*, 2016) and classifynder (Holt *et al.*, 2011), and non-image-based techniques such as light

scattering (Miki *et al.*, 2021) and traditional flow cytometry (Tennant *et al.*, 2013). The high volume of pollen grains analysed by IFC, addressed issues that complicate pollen classification, for example variations in the appearance of pollen grains due to rotation. Also, IFC was able to distinguish pollen from background material based on their autofluorescence properties, further promoting accurate identification by the network.

Our initial, conventional deep learning network (Network 1) accurately classified 53 modern species and excelled at tasks that are notoriously difficult for a human operator, for example, successfully discriminating between morphologically similar pollen such as *Agrostis capillaris* and *Agrostis gigantea*, and *Quercus alba* and *Quercus robur* (Fig. 2). This further supports the accuracy for deep learning to perform specific tasks, that is to classify known pollen types using controlled images (Dunker *et al.*, 2020). Such a network can be very useful in instances when the classes are known in advance and are included in training data (Olsson *et al.*, 2021). However, testing the network using hold-out data generated at the same time as the library samples, with minimal experimental variation, although typical for deep learning, is a key limitation in exploratory analysis and the application of deep learning to environmental samples with high uncertainty (i.e.

Table 2 Strategy used to develop a Guided Deep Learning network for the classification of pollen in palaeoenvironmental samples.

Step	Process
1	Images were given multiple labels based on their morphology and taxonomy. Labels related to morphology provided an opportunity for the expert to direct training to the most important parts of feature space whilst taxonomy labels allowed the network to learn more subtle features, linking images at various levels of the phylogenetic tree.
2	The network was trained on this multi-label problem using an adjusted cross-entropy loss function to inform learning of important features.
3	Features were extracted from the global pooling layer and visualised via a supervised learning implementation of Uniform Manifold Approximation and Projection (UMAP). The technique visualises data, placing manifolds according to external labels and can preserve both local and global information with manifolds placed near those with maximum similarity.
4	The process was tested on a range of datasets including palaeoenvironmental data to explore aspects of our network to assess the applicability of the method to real life, environmental samples and exploratory work. The fully trained network was used to extract features from these images, and UMAP was used to visualise these features in relation to those extracted from the training images.
5	Support Vector Machine learning (SVM) was used to learn the classification of images at various levels of their taxonomy according to their position in the 2D plot. The model was used to estimate classifications present in our test mixes.

background noise and non-pristine pollen grains). Deep learning is often criticised for its highly specific approach, and this was reflected in the inflexibility of Network 1 and Network 2 to successfully deal with poor image quality and previously unseen pollen types (Fig. S9A).

However, our Guided Deep Learning approach used to train Network 3 maintains the advantages of a CNN with the addition of expert knowledge of key morphological features and taxonomy. Assigned labels were highly weighted against classification at all levels of the phylogenetic tree (Fig. S8), making the network robust to noise and background images, allowing accurate classification of pollen in an environmental sample at different taxonomical ranks (Fig. 6). The influence of batch effects, often

problematic for networks trained for highly specific tasks (Doan *et al.*, 2020), were minimal (Fig. S1A–C) as were effects from different sample preparations (Fig. S1D,E). We achieved classification accuracies >97% in all cases for a subpopulation of pollen analysed over different days and therefore conclude that our approach was robust against intrainstrument variability. Although reduced, an accuracy of *c.* 73%, demonstrates the flexibility of Network 3 to accurately classify pollen despite slight morphological differences resulting from acid treatment (Fig. S1).

Our Guided Deep Learning algorithm also removed images identified as background noise or nonpollen events. This is an important step for all types of work, removing the need for extensive gating via IFC and, therefore, further speeding up classification. The inclusion of background reference images helped to optimise classification and prevent the network forcing erroneous predictions of unseen (not included in the library) background objects.

As an exciting exploratory tool, our network can not only accurately predict the presence of pollen types at genus or species level when represented in the training library, achieving an accuracy of >96% for such typical tasks (Fig. S5), but crucially predict pollen types at family or order level if not included in training datasets.

The flexibility Network 3 demonstrates to deal with complex environmental samples is also superior to supervised machine learning. Our approach involved labelling images in a weakly supervised way, with general high-level morphology labels assigned to all pollen belonging to a particular species. This is unlike classical machine learning that typically involves the manual inspection of individual images and the generation of standard features along with feature engineering (Kim & Choi, 2019; Lv *et al.*, 2022), which can be time consuming, requiring a degree of *a priori* expert knowledge and may still fail to capture all information. By exploiting the advantages of deep learning, our network can quickly learn a range of features at different taxonomical resolutions and has the potential to learn subtle abstract features that would be difficult to summarise as a metric.

Table 3 shows the representation of pollen types in the training libraries that were manually identified in the environmental sample. Our training libraries exhibit imbalances in their

Fig. 4 Uniform Manifold Approximation and Projection (UMAP) plot of features extracted from training images by Network 3. (a) Each point in the UMAP scatter plot represents an image present in the training data, colour-coded according to the order to which they belong. Each order investigated has a different number of representative samples: Lycopodiales, 1 (highlighted by dotted circle); Arecales, 1; Poales, 8; Asterales, 5; Brassicales, 1; Caryophyllales, 3; Dipsacales, 1; Fabales, 1; Fagales, 6; Lamiales, 2; Malpighiales, 2; Malvales, 1; Myrtales, 1; Proteales, 1; Rosales, 5; Sapindales, 4; Pinales, 7; Equisetales, 1. (b) Selected orders are numbered with exemplar images of the pollen types representing these orders. Family groups are noted, and dashed boxes highlight species belonging to the same genus. (bi) The seven species representing the Pinales order show morphological variability within the order. The distinct saccate grains of the Pinaceae family result in a clustering of similar features that are morphologically distinct from the spheroidal grains of the Cupressaceae family, within the same order. (bii) The Poales order is represented by eight species from the Poaceae family, which have spheroidal grains with one ulcus pore, reflected by a tight clustering of five of these eight species. Distance between the species, therefore, potentially reflects size variations and highlights the ability of the network to distinguish between species in the Poaceae family, which is notoriously difficult via manual classification. (biii) The Caryophyllales order is represented by three species from the Amaranthaceae family, that are morphologically similar: spheroidal pantoporate grains with echinate surface ornamentation. However, the UMAP plots these morphologically similar grains apart, highlighting the potential for deep learning to identify subtle differences that are not easily detectable by a human scorer. (biv) The Asterales order is represented by five species from the Asteraceae family, three of which belong to the same genus. Two of these species (*Ambrosia psilostachya* and *Ambrosia artemisiifolia*) cluster together, highlighting close similarities in the grains; however, the network also separates out *Ambrosia trifida* as being morphologically distinct.
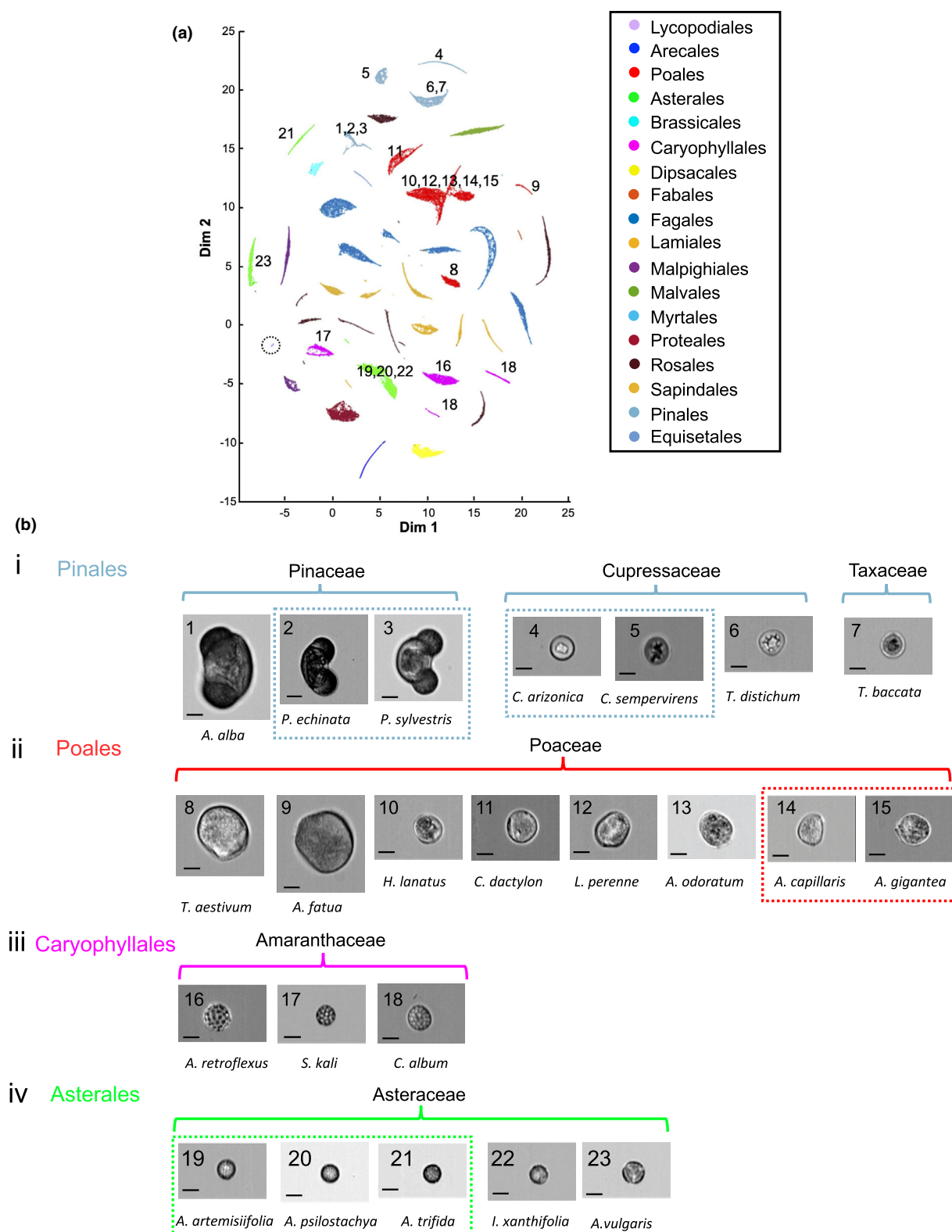
representation. Despite this, Network 3 successfully extracted features flexible enough to disregard subtle differences between species and correctly identify fossil pollen at genus level in the palaeoenvironmental sample when they were represented in the training set. A key example is *Corylus*, which was accurately predicted in high abundance. Moreover, pollen types present in the environmental sample but not represented in our training library were successfully detected at higher taxa, for example, *Carpinus*
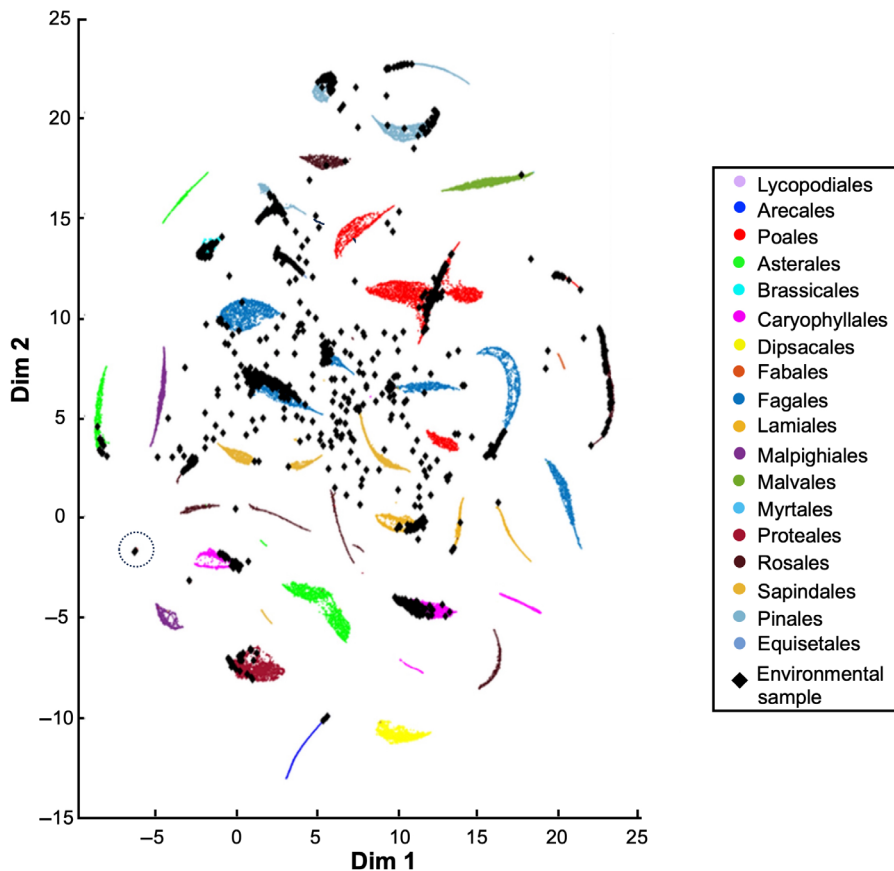
Fig. 5 Classification of fossil pollen within the sediments of the Mere Tarn core using Guided Deep Learning. The classification of fossil pollen is visualised by Uniform Manifold Approximation and Projection (UMAP). Black dots representing fossil pollen are superimposed over the modern pollen training dataset. Lycopodiales are highlighted by dotted circle.

(Betulaceae) and *Sparganium* (Poaceae) were successfully predicted at family level. Whilst there were no examples of pollen belonging to the Rosaceae family in our training library, the algorithm correctly identified the presence of Rosales, the order to which *Sorbus* (Rosaceae) belongs. Furthermore, the network also successfully detected pollen grains that were present in low abundance in the environmental sample such as *Pinus* at genus level. The detection of, for example, *Abies alba* by the network may have been a misclassification at species level, however, highlights a correct classification at family level, indicating the network's ability to learn connections between species. Since the truth populations for the environmental sample are derived from manual classifications (at genus/family level), it is impossible to quantify the accuracy of our network at species level; however, our work shows the potential for species-level discrimination.

A potential limitation of the network was the use of a modern pollen reference library for fossil pollen classification; however, the UMAP visualisation tool revealed that fossil pollen plotted tightly over the modern reference samples, demonstrating the suitability of our reference library for palaeoecological classifications (Fig. 5). Potential alterations of grain morphology during compaction in the sediment archive and the shrinking or swelling of grains from chemical extraction (Makela, 1996) were effectively handled by the flexibility of the network (Fig. S1).

Misclassifications observed in Fig. 6 likely reflect poor image quality or the presence of an 'unclassified' category of pollen. Our fully trained algorithm did not have such a category since training a network to identify such a heterogeneous class is challenging. However, an additional benefit of the UMAP visualisations is that the presence of unseen pollen types is indicated by

Fig. 6 Comparison of predicted fossil pollen assemblages with manual counts within the sediments of the Mere Tarn core. Pollen grains were classified via Guided Deep Learning at genus (a), family (b) and order level (c). Manual classifications resolved to genus or family level performed by a human operator (black bars) are presented alongside the predicted pollen assemblage (green bars) by Network 3. Manual counts could only be resolved to family level by human analyst for Poaceae and Cyperaceae. Predictions made at genus level for Poaceae are highlighted by *. Of the 21 main pollen types within the palaeoenvironmental sample (manual pollen count >3 grains), 15 were represented in our training library at genus level. *Sorbus*, *Carpinus* and *Sparganium* were represented at order level (Rosaceae, Betulaceae and Poaceae, respectively). However, some species (*Polypodium*, *Sphagnum*, *Hedera* and *Calluna*) were not represented at genus, family or order level. Classifications of other pollen types predicted by the network but not counted by the human analyst are presented in Supporting Information Fig. S10. Average level of certainty associated with a particular classification by the system was 97.14%. Some images achieved as low as 0.216 level of certainty associated with their classification with 9.38% having an estimated certainty of less than 95%, indicating that previously unseen pollen species are present within the sample. Exemplar images of correct classifications for selected pollen types are presented.
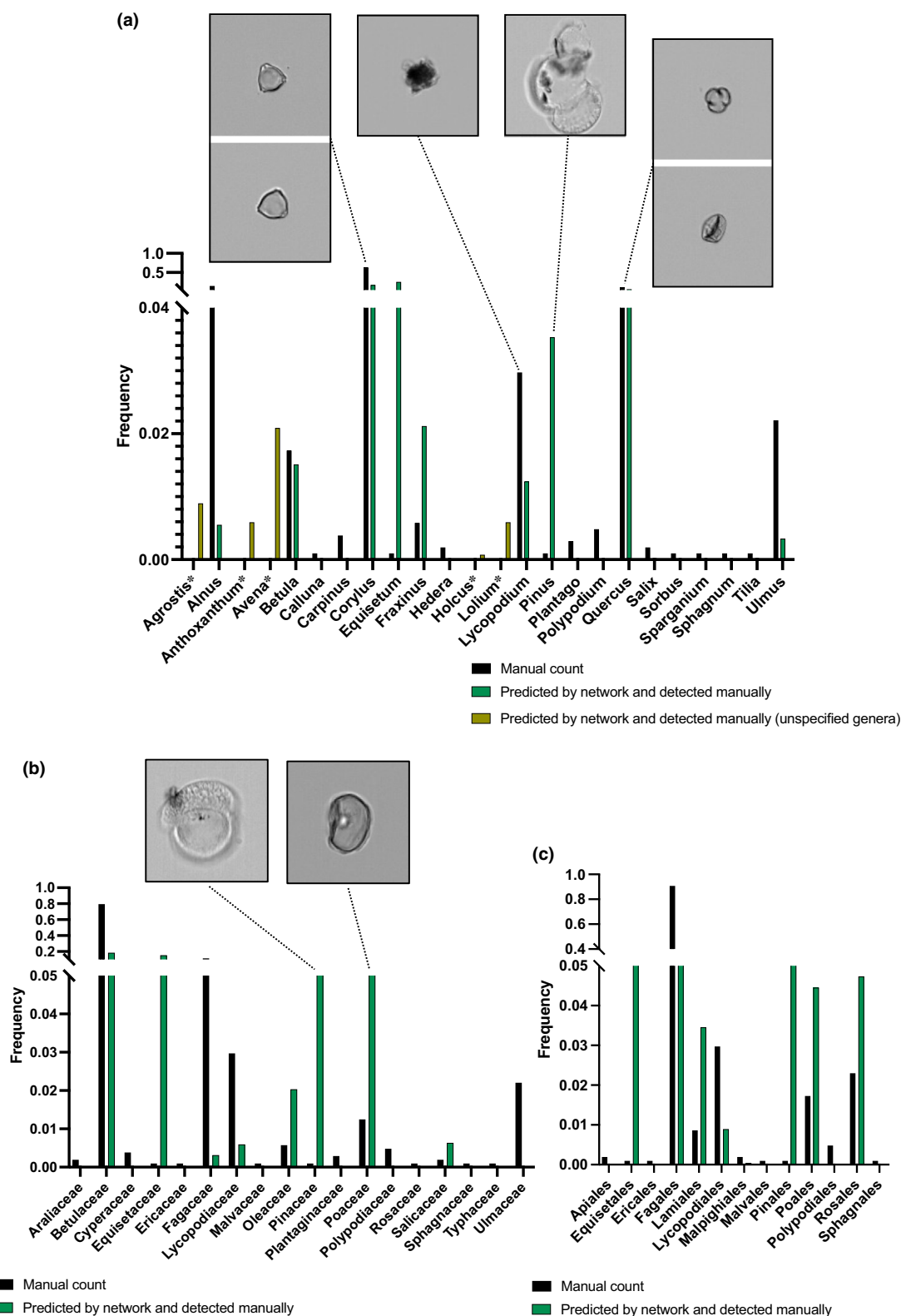
the spread of events across the feature space and confirmed by statistics related to estimated certainty of classification for each pollen image (Fig. 5). This will form the basis of future work which may also explore different combinations of the features used to guide learning. We chose features that scaled well across taxa

(Fig. S8); however, these could be refined to prioritise noise reduction and image quality. Our strategy allows the expert the unique opportunity to incorporate preferences into training.

Optimal training libraries are crucial for an exploratory network and should be populated with balanced representations of

Table 3 Summary of notable results from the classification of fossil pollen in the palaeoenvironmental sample from *Mere Tarn*, Cumbria, UK by Network 3.

Pollen type detected by Network 3	Representation in the training library (number of references at taxa-level)	Result summary
Genus		
Equisetum	1 species	Although manual counts of *Equisetum* spores were relatively low, the network detects high numbers in the sample. This is likely due to the homogenous morphology of *Equisetum* spores, with fewer distinguishing features. Consequently, the network may have confused them with morphologically similar pollen types (spheroidal, medium sized, one pore) such as those of the Poaceae family. This misclassification is however tracked across all levels of taxonomy.
Corylus	1 species	Identified as the most abundant pollen species by the network (after *Equisetum*), confirmed by manual counts.
Alnus	1 species	Accurately predicted in relatively high abundance by the network, verified by manual counts.
Quercus	2 species	Accurately predicted as the third most abundant pollen type by the network, verified by manual counts.
Lycopodium	1 genus	A *Lycopodium* spp. spore tablet was added to the sample to estimate pollen concentrations and was successfully classified by the network at genus level.
Ulmus	2 species	*Ulmus* exhibits a distinct morphology making it readily distinguishable by the network. Identified at family and genus level at relatively low concentration, reflecting manual counts.
Betula	1 species	Accurately detected in relatively low abundance.
Carpinus	0 species	Unable to detect at genus level due to poor representation in the training library, however, successfully predicted at family level (Fagales).
Pinus	2 species	Network detects this relatively rare pollen type.
Family		
Equisetaceae	1 species	Estimated at relatively high concentrations by the network, reflecting the misclassification of grains as *Equisetum* (as described above).
Betulaceae	3 species	Successfully predicted at high abundance. The network distinguishes between genera within this family: *Betula*, *Alnus* and *Corylus*.
Fagaceae	2 species	Estimated in relatively low abundance indicating that some pollen grains from this family may have been misclassified.
Poaceae	8 species	Grass species could only be resolved to family (Poaceae) level via manual counts due to similarities in pollen grain morphology. The network successfully classified pollen types belonging to the Poaceae family and furthermore, resolved identification to genus level, detecting the presence of *Agrostis*, *Anthoxanthum* and *Avena*.
Order		
Equisetales	1 family – 1 species	Estimated at relatively high concentrations by the network, reflecting the potential misclassification of grains as *Equisetum* (as described above).
Fagales	3 family – 6 species	Detected as the most dominant order, mirroring the manual counts. Also, likely to include *Carpinus* grains which were not represented in the library at genus level.
Lycopodiales	1 family – 1 genus	Successful detection of *Lycopodium* at order level.
Rosales	4 family – 6 species	
Poales	1 family – 8 species	Accurate detection at low abundance.
Pinales	3 family – 7 species	

The representation of pollen types in the training library (number of species at closest level of taxa) is noted.

pollen types at family, genus and species level. We show the incorporation of sufficient pollen types to reflect heterogeneity within a family is required for accurate predictions at this level. The inclusion of new training data may alter the UMAP clustering as more species are added to the reference library; however, our network could be quickly retrained for specific applications. Also, to prevent the network forcing an erroneous prediction of unseen pollen types or background objects, libraries should also contain background reference images and include adequate representation of pollen types of interest.

Despite some misclassifications, which can be easily identified and addressed by incorporating pollen types of interest within the training library, our modified network demonstrates excellent results for exploratory analysis, providing a rapid alternative to manual classification. Analysts manually identify pollen based on the similar gross morphologies of pollen types at family level (Leopold *et al.*, 2012) and painstakingly measure grain size, surface features and aperture characteristics to discern subtle differences between genera or species. At times, manual classifications can only be resolved at family level (Leopold *et al.*, 2012). Our network, however, can resolve taxonomical classifications beyond the ability of a human operator.

Although improved taxonomical resolution can also be achieved through eDNA analysis (Brennan *et al.*, 2019; Parducci *et al.*, 2019), there are several limitations. Firstly, the extraction of sufficient DNA from samples with low pollen concentrations. This is especially important for palaeolimnological samples where pollen grains are heavily diluted within the lake sedimentary

matrix. Taxonomical classifications are then reliant upon incomplete reference databases, which are biased towards commonly used indicator species (Pedersen *et al.*, 2015; Bell *et al.*, 2022) and are reported as relative abundance. Also, pollen cannot be discerned from plant debris (Johnson *et al.*, 2019), preventing the detection of atmospherically derived pollen signals from other (e.g. catchment) vegetation inputs (Parducci *et al.*, 2019). The classification of a visual pollen database, however, offers the advantages of distinguishing between pollen grains from plant debris and the quantification of absolute pollen counts, as well as relative abundance.

The Guided Deep Learning network removes human bias and reduces analysis time when classifying a visual pollen database. For example, it took a highly skilled expert more than 2 h to count 1043 pollen grains from one palaeoenvironmental sample. IFC, however, generated 5415 pollen images in 30 min, which were then applied to our network, preprocessed and classified in 30 min.

Classification of environmental pollen from digitalised microscope slides via deep learning boasts rapid image capture at higher magnification than our approach (Olsson *et al.*, 2021; Punyasena *et al.*, 2022). Punyasena *et al.* (2022) utilised training datasets annotated by a human operator and therefore precise, species-level classifications were limited (Punyasena *et al.*, 2022). High numbers of pollen reference samples and taxa (83 species, 122 000 images) have been used to classify pristine environmental samples (bee pollen) that are highly abundant in pollen grains (Olsson *et al.*, 2021). Although rapid image acquisition is reported (Olsson *et al.*, 2021), only 2 µl of sample was analysed per slide (five slides were loaded per scan requiring 30 min analysis time), presumably at an optimal concentration for sufficient pollen grain dispersal to prevent grains becoming obscured on the slide. Each digitalised slide required z stacking of five focus layers (15 min per image) before classification (105 min total for 10 µl). For our environmental samples, 30–47 µl of sample containing relatively rare pollen populations was analysed in 30 min with no need for object detection via segmentation.

The rapid analysis our technique allows offers the opportunity for high-resolution spatial and temporal sampling in palaeoecological investigations and therefore more robust, replicated historical reconstructions of human activity (Deza-Araujo *et al.*, 2020; e.g. occurrence of early agriculture) and climate change throughout the Holocene (Kaufman *et al.*, 2020). Issues such as low pollen concentrations in samples can be addressed by performing several IFC data acquisitions to produce more statistically reliant datasets. With inclusion of specific species of interest in our training libraries, higher taxonomical precision can also be achieved, for example distinguishing wild grass pollen from domesticated cereals, which is extremely difficult for a pollen analyst (Joly *et al.*, 2007). Since most palaeoenvironmental samples are stored in glycerol, there is potential to rapidly re-analyse fossil pollen records using our technique, or determine broad pollen assemblages down core, from which important stratigraphic horizons requiring more detailed investigation could be selected. The rapid generation of multiple down-core pollen records at one site may also be used as a tool for intrasite core correlation.

Our flexible network could be tasked with additional reference pollen types to expand classification and resolve regional pollen signatures. Alternatively, bespoke training libraries that reflect specific pollen types of interest may also be applied to classify distinct localised pollen signals comprised of fewer pollen types. Furthermore, the ability of our network to classify background (nonpollen) events also suggests that other environmental proxies such as tephra and diatoms could also be rapidly analysed simultaneously by our technique.

The high versatility of our network makes it widely transferable to revolutionise many other aspects of palynology, to enhance taxonomical precision and provide rapid, high-resolution temporal and spatial sampling. For instance, resolving uncertainties currently experienced in aeroallergen forecasting (Erbas *et al.*, 2012; Lake *et al.*, 2017) and to better understand spatial and temporal variations in exposure to pollen from highly allergic grass species that are notoriously difficult to identify manually (Brennan *et al.*, 2019); forensic applications (Webb *et al.*, 2018); archaeological reconstructions; or determining pollen in honey to assess food quality and foraging behaviour of bees (Brodschneider *et al.*, 2019; Hoffmann *et al.*, 2020); investigations into biodiversity (Leontidou *et al.*, 2021); pollination and the migratory patterns and habitats of pollinators (Doyle *et al.*, 2020); and the morphological traits of pollen (phenotypes) and plant evolution (Heinze *et al.*, 2017).

Competing interests

None declared.

Author contributions

ALP, CMB, DGB, PR, HDS, RTJ and JL conceptualised the study. CMB, ALP and DGB planned and designed the research. RKT, GRL, AE, HDS, SMH, JZ-C and MD provided resources. AE provided technical support. Formal experiments and analyses were performed by CMB, ALP and JH. ALP and CMB wrote the manuscript with editing and reviewing by ALP, CMB, DGB, AEC, JL and PR. ALP and CMB contributed equally to this work as first authors.

ORCID

Daniel G. Barber https://orcid.org/0000-0002-5022-2846
Claire M. Barnes https://orcid.org/0000-0003-1031-7127
Anne E. Carpenter https://orcid.org/0000-0003-1555-8261
Minh Doan https://orcid.org/0000-0002-3235-0457
Stephen M. Haley https://orcid.org/0000-0002-5928-3063
Ann L. Power https://orcid.org/0000-0002-7651-5276
Paul Rees https://orcid.org/0000-0002-7715-6914
Huw D. Summers https://orcid.org/0000-0002-0898-5612
Richard K. Tennant https://orcid.org/0000-0003-3033-1858

Data availability

The data that support the findings of this study are available from Biostudies: https://www.ebi.ac.uk/biostudies/studies/S-BSST1152 (data embargoed until 1 September 2024).

References

Åkesson CM, Matthews-Bird F, Bitting M, Fennell C-J, Church WB, Peterson LC, Valencia BG, Bush MB. 2019. 2,100 years of human adaptation to climate change in the High Andes. *Nature Ecology & Evolution* 4: 66–74.

Anenberg SC, Haines S, Wang E, Nassikas N, Kinney PL. 2020. Synergistic health effects of air pollution, temperature, and pollen exposure: a systematic review of epidemiological evidence. *Environmental Health* 19: 130.

Baksay S, Pornon A, Burrus M, Mariette J, Andalo C, Escaravage N. 2020. Experimental quantification of pollen with DNA metabarcoding using ITS1 and trnL. *Scientific Reports* 10: 1–9.

Bell KL, Burgess KS, Botsch JC, Dobbs EK, Read TD, Brosi BJ. 2018. Quantitative and qualitative assessment of pollen DNA metabarcoding using constructed species mixtures. *Molecular Ecology* 28: 431–455.

Bell KL, Turo KJ, Lowe A, Nota K, Keller A, Encinas-Viso F, Parducci L, Richardson RT, Leggett RM, Brosi BJ et al. 2022. Plants, pollinators and their interactions under global ecological change: the role of pollen DNA metabarcoding. *Molecular Ecology*, 1–18. doi: 10.1111/mec.16689.

Bourel B, Marchant R, de Garidel-Thoron T, Tetard M, Barboni D, Gally Y, Beaufort L. 2020. Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. *Computers & Geosciences* 140: 104498.

Brennan GL, Potter C, de Vere N, Griffith GW, Skjoth CA, Osborne NJ, Wheeler BW, McInnes RN, Clewlow Y, Barber A et al. 2019. Temperate airborne grass pollen defined by spatio-temporal shifts in community composition. *Nature Ecology & Evolution* 3: 750–754.

Brodschneider R, Gratzer K, Kalcher-Sommersguter E, Heigl H, Auer W, Moosbeckhofer R, Crailsheim K. 2019. A citizen science supported study on seasonal diversity and monoflorality of pollen collected by honey bees in Austria. *Scientific Reports* 9: 891–812.

Daood A, Ribeiro E, Bush M. 2016. Pollen Grain Recognition Using Deep Learning. In: Bebis G, Boyle R, Parvn B, Koracin D, Porikli F, Skaff S, Entezari A, Min J, Iwai D, Sadagic A, Scheidegger C, Isenberg T, eds. *Advances in visual computing, vol. 10072*. ISVC 2016. Lecture Notes in Computer Science. Cham, Switzerland: Springer, 321–330.

Deza-Araujo M, Morales-Molino C, Tinner W, Henne PD, Heitz C, Pezzatti GB, Hafner A, Conedera M. 2020. A critical assessment of human-impact indices based on anthropogenic pollen indicators. *Quaternary Science Reviews* 236: 106291.

Dhungel N, Carneiro G, Bradley AP. 2017. Fully automated classification of mammograms using deep residual neural networks. In: *2017 IEEE 14th International symposium on biomedical imaging (ISBI 2017)*. Melbourne, VIC, Australia: IEEE, 310–314.

Doan M, Barnes C, McQuin C, Caicedo JC, Goodman A, Carpenter AE, Rees P. 2021. Deepometry, a framework for applying supervised and weakly supervised deep learning to imaging cytometry. *Nature Protocols* 16: 3572–3595.

Doan M, Sebastian JA, Caicedo JC, Siegert S, Roch A, Turner TR, Mykhailova O, Pinto RN, McQuin C, Goodman A et al. 2020. Objective assessment of stored blood quality by deep learning. *Proceedings of the National Academy of Sciences, USA* 117: 21381–21390.

Doan M, Vorobjev I, Rees P, Filby A, Wolkenhauer O, Goldfeld AE, Lieberman J, Barteneva N, Carpenter AE, Hennig H. 2018. Diagnostic potential of imaging flow cytometry. *Trends in Biotechnology* 36: 649–652.

Doyle T, Hawkes WLS, Massy R, Powney GD, Menz MHM, Wotton KR. 2020. Pollination by hoverflies in the Anthropocene. *Proceedings of the Biological Sciences* 287: 20200508.

Dunker S, Motivans E, Rakosy D, Boho D, Mäder P, Hornick T, Knight TM. 2020. Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytologist* 229: 593–606.

Edwards KJ, Fyfe RM, Jackson ST. 2017. The first 100 years of pollen analysis. *Nature Plants* 3: 17001.

Erbas B, Akram M, Dharmage SC, Tham R, Dennekamp M, Newbigin E, Taylor P, Tang MLK, Abramson MJ. 2012. The role of seasonal grass pollen on childhood asthma emergency department presentations. *Clinical & Experimental Allergy* 42: 799–805.

France I, Duller AWG, Duller GAT, Lamb HF. 2000. A new approach to automated pollen analysis. *Quaternary Science Reviews* 19: 537–546.

Heidmann I, Schade-Kampmann G, Lambalk J, Ottiger M, Di Berardino M. 2016. Impedance flow cytometry: a novel technique in pollen analysis. *PLoS ONE* 11: e0165531.

Heinze B, Kriebel R, Khabbazian M, Sytsma KJ. 2017. A continuous morphological approach to study the evolution of pollen in a phylogenetic context: an example with the order Myrtales. *PLoS ONE* 12: e0187228.

Hoffmann RD, Portes MT, Olsen LI, Damineli DSC, Hayashi M, Nunes CO, Pedersen JT, Lima PT, Campos C, Feijó JA et al. 2020. Plasma membrane H^+-ATPases sustain pollen tube growth and fertilization. *Nature Communications* 11: 615.

Holt K, Allen G, Hodgson R, Marsland S, Flenley J. 2011. Progress towards an automated trainable pollen location and classifier system for use in the palynology laboratory. *Review of Palaeobotany and Palynology* 167: 175–183.

Holt KA, Bennett KD. 2014. Principles and methods for automated palynology. *New Phytologist* 203: 735–742.

Johnson MD, Cox RD, Barnes MA. 2019. The detection of a non-anemophilous plant species using airborne eDNA. *PLoS ONE* 14: e0225262.

Joly C, Barillé L, Barreau M, Mancheron A, Visset L. 2007. Grain and annulus diameter as criteria for distinguishing pollen grains of cereals from wild grasses. *Review of Palaeobotany and Palynology* 146: 221–233.

van der Kaars S, Miller GH, Turney CSM, Cook EJ, Nürnberg D, Schönfeld J, Kershaw AP, Lehman SJ. 2017. Humans rather than climate the primary cause of Pleistocene megafaunal extinction in Australia. *Nature Communications* 8: 1–7.

Kaufman D, McKay N, Routson C, Erb M, Davis B, Heiri O, Jaccard S, Tierney J, Datwyler C, Axford Y et al. 2020. A global database of Holocene paleotemperature records. *Scientific Data* 7: 115.

Khanzhin N, Putin E, Filchenkov A, Zamyatina E. 2018. Pollen grain recognition using convolutional neural network. In: *ESANN 2018 proceedings, European symposium on artificial neural networks, computational intelligence and machine learning*, Bruges, Belgium, 25–27 April 2018, 409–414.

Kim S-H, Choi H-L. 2019. Convolutional neural network-based multi-target detection and recognition method for unmanned airborne surveillance systems. *International Journal of Aeronautical and Space Sciences* 20: 1038–1046.

Lacourse T, May L. 2012. Increasing taxonomic resolution in pollen identification: sample size, spatial sampling bias and implications for palaeoecology. *Review of Palaeobotany and Palynology* 182: 55–64.

Lake IR, Jones NR, Agnew M, Goodess CM, Giorgi F, Hamaoui-Laguel L, Semenov MA, Solomon F, Storkey J, Vautard R et al. 2017. Climate change and future pollen allergy in Europe. *Environmental Health Perspectives* 125: 385–391.

Langford M, Taylor GE, Flenley JR. 1990. Computerized identification of pollen grains by texture analysis. *Review of Palaeobotany and Palynology* 64: 197–203.

Ledger PM. 2017. Are circumpolar hunter-gatherers visible in the palaeoenvironmental record? Pollen-analytical evidence from Nunalleq, southwestern Alaska. *The Holocene* 28: 415–426.

Leontidou K, Vokou D, Sandionigi A, Bruno A, Lazarina M, De Groeve J, Li M, Varotto C, Girardi M, Casiraghi M *et al.* 2021. Plant biodiversity assessment through pollen DNA metabarcoding in Natura 2000 habitats (Italian Alps). *Scientific Reports* 11: 18226.

Leopold EB, Birkebak J, Reinink-Smith L, Jayachandar AP, Narváez P, Zaborac-Reed S. 2012. Pollen morphology of the three subgenera of *Alnus*. *Palynology* 36: 131–151.

Li C, Zhen T, Li Z. 2022. Image classification of pests with residual neural network based on transfer learning. *Applied Sciences* 12: 4356.

Li Y, Jiang T, Xie W, Lei J, Du Q. 2022. Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning. *IEEE Transactions on Geoscience and Remote Sensing* 60: 1–11.

Lv Q, Zhang S, Wang Y, Li Q. 2022. Deep learning model of image classification using machine learning. *Advances in Multimedia* 2022: 1–12.

MacLeod N, Benfield M, Culverhouse P. 2010. Time to automate identification. *Nature* 467: 154–155.

Makela EM. 1996. Size distinctions between Betula pollen types – a review. *Grana* 35: 248–256.

Mander L, Li M, Mio W, Fowlkes CC, Punyasena SW. 2013. Classification of grass pollen through the quantitative analysis of surface ornamentation and texture. *Proceedings of the Royal Society B: Biological Sciences* 280: 20131905.

Marcos JV, Nava R, Cristóbal G, Redondo R, Escalante-Ramírez B, Bueno G, Déniz Ó, González-Porto A, Pardo C, Chung F *et al.* 2015. Automated pollen identification using microscopic imaging and texture analysis. *Micron* 68: 36–46.

McInnes L, Healy J, Melville J. 2018. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv preprint*: 1802.03426.

Miki K, Fujita T, Sahashi N. 2021. Development and application of a method to classify airborne pollen taxa concentration using light scattering data. *Scientific Reports* 11: 22371.

Mitsumoto K, Yabusaki K, Aoyagi H. 2009. Classification of pollen species using autofluorescence image analysis. *Journal of Bioscience and Bioengineering* 107: 90–94.

Olsson O, Karlsson M, Persson AS, Smith HG, Varadarajan V, Yourstone J, Stjernman M, Freckleton R. 2021. Efficient, automated and robust pollen analysis using deep learning. *Methods in Ecology and Evolution* 12: 850–862.

Parducci L, Alsos IG, Unneberg P, Pedersen MW, Han L, Lammers Y, Salonen JS, Väliranta MM, Slotte T, Wohlfarth B. 2019. Shotgun environmental DNA, pollen, and macrofossil analysis of lateglacial lake sediments from southern Sweden. *Frontiers in Ecology and Evolution* 7: 1–15.

Pedersen MW, Overballe-Petersen S, Ermini L, Sarkissian CD, Haile J, Hellstrom M, Spens J, Thomsen PF, Bohmann K, Cappellini E *et al.* 2015. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 370: 20130383.

Punyasena SW, Haselhorst DS, Kong S, Fowlkes CC, Moreno JE. 2022. Automated identification of diverse Neotropical pollen samples using convolutional neural networks. *Methods in Ecology and Evolution* 13: 2049–2064.

Qiu Z, Jiang H, Ding L, Shang X. 2020. Late Pleistocene-Holocene vegetation history and anthropogenic activities deduced from pollen spectra and archaeological data at Guxu Lake, eastern China. *Scientific Reports* 10: 1–14.

Radovanović D, Đukanovic S. 2020. Image-based plant disease detection: a comparison of deep learning and classical machine learning algorithms. In: *2020 24th international conference on information technology (IT)*, Zabljak, Montenegro, 1–4.

Sarwinda D, Paradisa RH, Bustamam A, Anggia P. 2021. Deep learning in image classification using residual network (ResNet) variants for detection of colorectal cancer. *Procedia Computer Science* 179: 423–431.

Sevillano V, Holt K, Aznarte JL. 2020. Precise automatic classification of 46 different pollen types with convolutional neural networks. *PLoS ONE* 15: e0229751.

Stillman EC, Flenley JR. 1996. The needs and prospects for automation in palynology. *Quaternary Science Reviews* 15: 1–5.

Stockmarr J. 1971. Tablets with spores used in absolute pollen analysis. *Pollen et Spores* 13: 615–621.

Stolarek I, Samelak-Czajka A, Figlerowicz M, Jackowiak P. 2022. Dimensionality reduction by UMAP for visualizing and aiding in classification of imaging flow cytometry data. *iScience* 25: 105142.

Tan L, Lu J, Jiang H. 2021. Tomato leaf diseases classification based on leaf images: a comparison between classical machine learning and deep learning methods. *AgriEngineering* 3: 542–558.

Tcheng DK, Nayak AK, Fowlkes CC, Punyasena SW. 2016. Visual recognition software for binary classification and its application to spruce pollen identification. *PLoS ONE* 11: e0148879.

Tello-Mijares S, Flores F. 2016. A novel method for the separation of overlapping pollen species for automated detection and classification. *Computational and Mathematical Methods in Medicine* 2016: 1–12.

Tennant RK, Jones RT, Brock F, Cook C, Turney CSM, Love J, Lee R. 2013. A new flow cytometry method enabling rapid purification of fossil pollen from terrestrial sediments for AMS radiocarbon dating. *Journal of Quaternary Science* 28: 229–236.

Webb JC, Brown HA, Toms H, Goodenough AE. 2018. Differential retention of pollen grains on clothing and the effectiveness of laboratory retrieval methods in forensic settings. *Forensic Science International* 288: 36–45.

Xu Y, Zhu JY, Chang EI, Lai M, Tu Z. 2014. Weakly supervised histopathology cancer image segmentation and classification. *Medical Image Analysis* 18: 591–604.

Yao X, Han J, Cheng G, Qian X, Guo L. 2016. Semantic annotation of high-resolution satellite images via weakly supervised learning. *IEEE Transactions on Geoscience and Remote Sensing* 54: 3660–3671.

Zhang Y, Fountain DW, Hodgson RM, Flenley JR, Gunetileke S. 2004. Towards automation of palynology 3: pollen pattern recognition using Gabor transforms and digital moments. *Journal of Quaternary Science* 19: 763–768.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Determining batch effects.

Fig. S2 Selection of high-quality pollen images from pollen standards.

Fig. S3 Variations in focus and rotation for individual pollen grains.

Fig. S4 Gating strategy for identifying pollen in palaeoenvironmental samples.

Fig. S5 Classification accuracy for pollen standards using Guided Deep Learning.

Fig. S6 Classification accuracy for modern pollen standards at species level (Network 1).

Fig. S7 Classification accuracy for modern pollen standards at order level (Network 2).

Fig. S8 Assessment of extracted features by networks for pollen classification.

Fig. S9 Classification of pollen in Mere Tarn sediments using Network 1 and Network 3.

Fig. S10 Pollen types predicted by Network 3 in the sediments of Mere Tarn core.

Table S1 Morphological and taxonomical features used to inform Guided Deep Learning.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.