Deep learning approaches to the phylogenetic placement of extinct pollen morphotypes

Marc-Élie Adaïmé [ID][a], Shu Kong [ID][b,c] and Surangi W. Punyasena [ID][a,*]

[a]Department of Plant Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
[b]Faculty of Science and Technology, University of Macau, Macau 999078, China
[c]Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843, USA
*To whom correspondence should be addressed: Email: spunya1@illinois.edu
Edited By: Gerhard Hummer

Abstract

The phylogenetic interpretation of pollen morphology is limited by our inability to recognize the evolutionary history embedded in pollen features. Deep learning offers tools for connecting morphology to phylogeny. Using neural networks, we developed an explicitly phylogenetic toolkit for analyzing the overall shape, internal structure, and texture of a pollen grain. Our analysis pipeline determines whether testing specimens are from known species based on uncertainty estimates. Features from specimens with uncertain taxonomy are passed to a multilayer perceptron network trained to transform these features into predicted phylogenetic distances from known taxa. We used these predicted distances to place specimens in a phylogeny using Bayesian inference. We trained and evaluated our models using optical superresolution micrographs of 30 extant *Podocarpus* species. We then used trained models to place nine fossil *Podocarpidites* specimens within the phylogeny. In doing so, we demonstrate that the phylogenetic history encoded in pollen morphology can be recognized by neural networks and that deep-learned features can be used in phylogenetic placement. Our approach makes extinction and speciation events that would otherwise be masked by the limited taxonomic resolution of the fossil pollen record visible to palynological analysis.

Keywords: computer vision, fossil pollen, machine learning, neural networks, phylogenetic placement

Significance Statement

Machine-learned features from deep neural networks can do more than categorize and classify biological images. We demonstrate that these features can also be used to quantify morphological differences among pollen taxa, discover novel morphotypes, and place fossil specimens on a phylogeny using Bayesian inference. Deep learning can be used to characterize and identify the morphological features with evolutionary significance. These features can then be used to infer phylogenetic distance. This approach fundamentally changes how fossil pollen morphology can be interpreted, allowing greater evolutionary inference of fossil pollen specimens. The analysis framework, however, is not specific to pollen and can be generalized to other taxa and other biological images.

Introduction

Machine learning in the computational biological literature has largely focused on biological classification and categorization, that is, developing neural networks for *K*-way classification, where *K* represents the number of taxa in a training set. Neural networks are used to identify features that define individual classes (taxa) and not the features that define the evolutionary relationships among taxa. This poses a challenge when classifying fossil specimens, many of which may be derived from extinct species that are entirely new to a trained network.

New methods are needed to detect novel fossil taxa and to accurately place them within a phylogeny. Deep learning techniques, such as convolutional neural networks (CNNs), are capable of extracting key discriminative features from a given image ([1], [2]). They can quantify phenotypic similarities between taxa, measure phenotypic distance, and define a clade's morphospace ([3]). Novelty can be identified by measuring the uncertainty within the *K*-way classification for a given specimen: high uncertainty implies a high likelihood that the specimen is from an unknown taxon ([4], [5]). For phylogenetic placement, however, neural networks must explicitly incorporate evolutionary distances. Models must be trained not only to identify the traits which define classes but also to recognize phylogenetic synapomorphies—derived features that are shared between sister taxa.

Our insight is that phylogenetic distances can be used to relearn and reweight morphological features derived from CNNs. Transformed features can be treated as a vector of continuous morphological traits for placing taxa within a known phylogeny using Bayesian inference. We obtained these revised features by

first extracting the features from classification CNNs and then

passing them to a second embedding model trained to transform features according to a phylogenetically informed distance function. This is a fundamentally new approach in applying machine learning to biological classification.

Fossil pollen is an example of a fossil record where many morphotypes have unknown or uncertain biological affinities (6–8). So although fossil pollen comprises one of the most extensive paleontological archives of terrestrial ecosystems, poor taxonomic resolution has limited evolutionary and paleoecological inference (6, 9). Our approach builds on our previous work demonstrating that well-constructed CNN analyses can transfer learning from modern pollen morphology to fossil (10) and expands the range of questions that can be addressed by pollen data, offering new insights into paleoecology, plant evolution, and phylogenetics. It allows us to recognize extinct morphotypes and reconstruct their phylogenetic relationship to extant taxa, and to reassess and recalibrate existing phylogenetic trees using microfossil data. The development of phylogenetically informed deep learning models is not specific to pollen and can be generalized to other taxa with robust phylogenies and standardized morphological imaging.
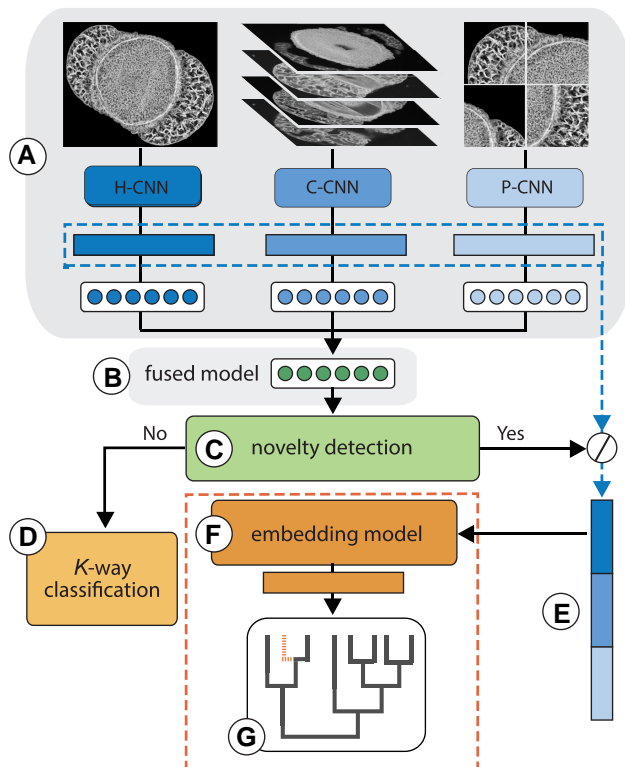

Fig. 1. Flowchart illustrating the trained multimodal neural network pipeline. Three representations of superresolution images are passed through CNNs capturing shape, internal structure, and texture (H-CNN, C-CNN, P-CNN; A). The three sets of classification scores are fused (FM; B) and the analysis determines whether a specimen belongs to a known taxon by assessing the network's uncertainty during image classification (ROC analysis; C). If the specimen is recognized as one of the K taxa with high confidence (low uncertainty), its predicted taxon is reported D). Otherwise, its features are extracted across all three image modalities and concatenated E) as input to a multilayer perceptron F), which is trained to transform these features to an embedding feature to better compute phylogenetic distances from known taxa. Embedding features are clustered with the features of known taxa for phylogenetic placement, based on Bayesian inference G).

General analysis pipeline

Image classification

We began by training three independent CNNs for K-way classification by exploiting three individual modalities (shape, internal structure, and texture) (Fig. 1A). The first, holistic CNN (H-CNN) took as input maximum intensity projections (MIPs) of the whole grain. The second, cross-sectional CNN (C-CNN) used axial slices taken through the entirety of a pollen grain. The third, patch CNN (P-CNN), was trained on overlapping square crops of the initial MIP image, each covering ~10% of the entire grain. The result of each CNN was a K-dim vector of classification scores that captured probabilities for each known taxonomic classification. We used a probabilistic fused model (FM) (Fig. 1B) to combine these single-modal probabilities into a fused multimodal probability for better performance.

Novel taxon detection

The total uncertainty of each CNN and FM classification was calculated using entropy computed over the product of their classification probabilities. We used receiver operating characteristic (ROC) analysis (11) to establish the uncertainty threshold above which a specimen should be considered novel (Fig. 1C). ROC is widely used to evaluate the performance of binary classification tasks (11, 12). The area under the ROC (AUROC) curve (5) quantified our models' ability to discriminate between known and novel pollen taxa. We determined the threshold that optimized both the precision and recall of novel taxa. If the model's uncertainty score was below this threshold, the specimen's most likely predicted taxon was reported (Fig. 1D). Otherwise, the sample was flagged as novel, and the features extracted from the three classification CNNs (Fig. 1E) were forward-passed through an embedding model for phylogenetic placement (Fig. 1F).

Phylogenetic placement

Our phylogenetic embedding model is a trained multilayer perceptron (MLP) that transforms a specimen's classification features to embedding features based on a learned phylogenetic distance function (Fig. 1F). The resulting embedding features capture the phylogenetic information within classification features. The embedding features can then be used to place a novel specimen on an established phylogeny (Fig. 1G). We train the MLP over specimens from K known taxa by minimizing the differences in the computed Euclidean distances between pairs of specimens' embedding features and ground-truth phylogenetic distances. We then use Bayesian inference to reconstruct the original tree topology from the continuous machine-learned features, applying the Brownian motion (BM) model for character evolution (13–15).

Materials and methods
Specimens and molecular phylogeny

We used 30 extant *Podocarpus* (Podocarpaceae) species and nine fossil *Podocarpidites* specimens to train and evaluate our neural networks (16). Specimens were from the collections of Smithsonian Tropical Research Institute and Utrecht University. The image dataset included 309 extant specimens from Australasia/Indomalaya (16 species), the Neotropics (10 species), Africotropics (2 species), and the Asian Palearctic (1 species) (Fig. 2 and Table S1). Specimens of each species were from a single slide and single individual. Fossil *Podocarpidites* specimens were isolated from samples from the Late Cretaceous and Neogene of Colombia, Peru, and Panama (Fig. 2 and Table S2) (17–19).
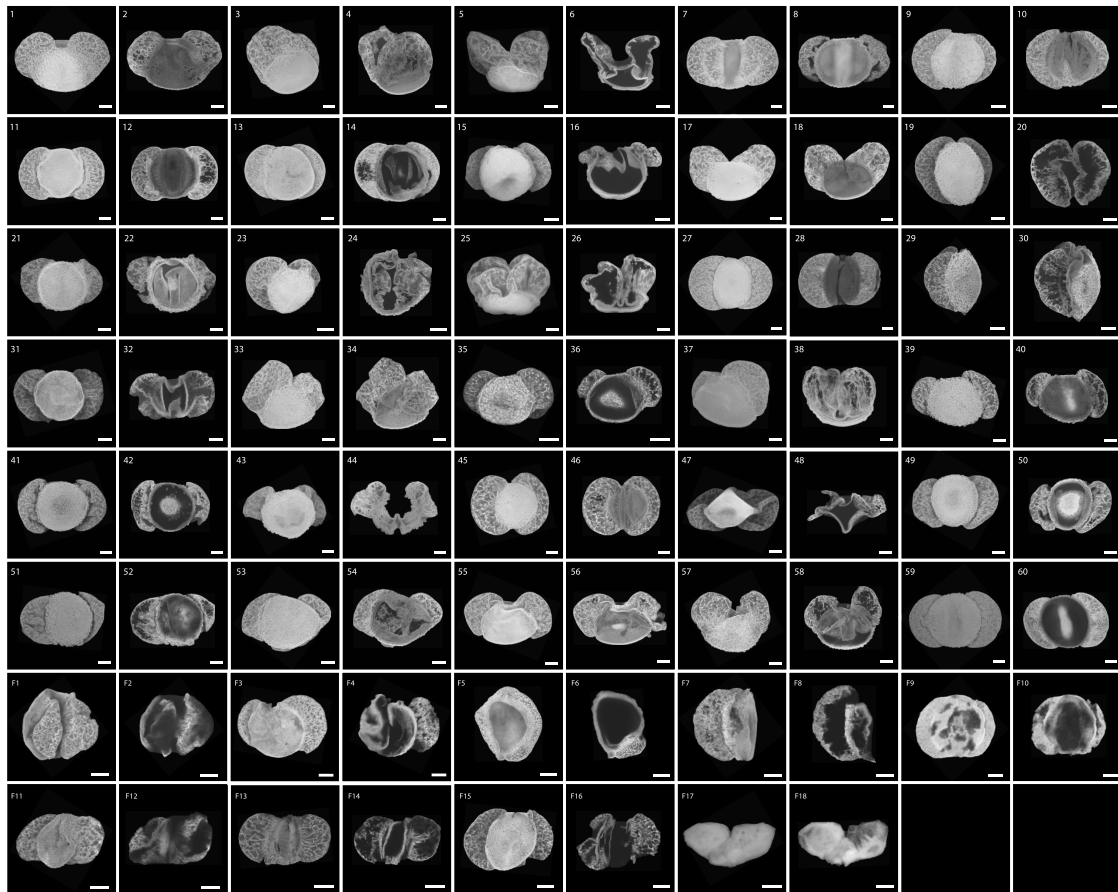
Fig. 2. Modern and fossil specimens of *Podocarpus*. Each specimen is represented by a pair of images consisting of an MIP (first image) and a cross-sectional image (second image) derived from image stacks taken with confocal superresolution (Zeiss LSM 880 with Airyscan, 63 × /1.4NA objective). Plate numbers and their corresponding species can be found in Tables S1 and S2 (scale bar: 10 μm).

Modern taxon names were reviewed and amended according to the taxonomic nomenclature in Tropicos (20).

For our phylogenetic model, we relied on a published conifer-wide phylogeny based on a concatenated dataset of two chloroplast DNA (cpDNA) protein-coding genes, rbcL and matK, and the 18S nuclear ribosomal DNA (rDNA) gene (21). The reconstruction used a partitioned maximum likelihood analysis and assumed a GTR+Γ nucleotide substitution model. Most clades were well supported, with high bootstrap support. The phylogeny was time-calibrated using a Late Eocene specimen from New South Wales (a leafy shoot with a mature female cone and an immature cone) (21, 22). Using a time-calibrated phylogeny allowed us to work with absolute ages (in millions of years, Ma) rather than relative phylogenetic distances. Of the 45 species in the *Podocarpus* image dataset (16), only 30 were included in the phylogeny and were therefore selected for this study.

Image modalities

Specimens were imaged using a Zeiss LSM 880 with Airyscan confocal superresolution at 630× magnification (63×/1.4 NA oil objective, 0.08 μm per pixel resolution) (16). Images were taken as a series of axial focal planes (Z-stacks) at 0.19 μm increments. We manually cropped and masked grains to remove background debris.

For our first imaging modality, we generated images of the external structure and shape of an entire pollen grain using MIPs of the entire specimen image stack (9). We applied contrast-limited adaptive histogram equalization (CLAHE) to standardize all images (23).

For the second modality, we used cross-sectional images. For the modern specimens, we generated a series of internal cross-sections by producing MIPs of staggered subsections of the image stack (10). Each substack included 15 axial cross-sectional planes (2.85 μm depth), offset by 10 planes (1.90 μm). Because our fossil specimens were highly compressed, instead of subsections we used each individual plane of the image stack as our cross-sections.

The final modality was image patches that sampled a small portion of the grain. The whole-grain MIP for each specimen was divided into 10–13 smaller square patches, each of which covered ~10% of the original MIP image and overlapped ~25% with adjacent patches (24).

Deep learning models

Network architectures

We trained three separate *K*-way classification CNNs for each of our three modalities: an H-CNN, a C-CNN, and a P-CNN. All three CNNs incorporated ResNeXt-101 architecture (25) with the exception of the last layer, which was modified to output *K*-dim logit vectors. We normalized the logit vectors using softmax to produce *K*-dim probability vectors, which indicated the probability of the input image being classified as each of the *K* pollen taxon classes.

We fine-tuned the ResNeXt CNNs (pretrained on ImageNet images) using our modern reference pollen dataset (26, 27). For a

given specimen, the H-CNN produced a K-dim logit vector (i.e. the feature prior to the K-dim softmax scores). The C-CNN and P-CNN used multiple cross-sections or patches and produced multiple logit vectors, respectively. We next computed the mean of the per-patch logit vectors and the mean of the per-cross-sectional logit vectors. Lastly, we summed the three logit vectors and normalized by softmax to obtain a K-dim classification probability vector.

Training details

Specimens of the K known taxa were randomly split into training (70%) and validation (30%) sets. We augmented the training data by adopting common augmentation techniques including random flip, random rotation (with probability $p = 0.5$ by degrees in $[-90°, +90°]$), and random translation (within pixel displacement in $[-30, +30]$). We resized the augmented images to 224×224×3 pixel resolution using bilinear interpolation.

We trained the K-way classification CNNs by minimizing the cross-entropy loss:

$$\mathcal{L}^{CE} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} T_{i,k} \times \log(P_{i,k}), \tag{1}$$

where N is the number of specimens in the training set, K is the number of taxa, and $P_{i,k}$ is the softmax probability of specimen i classified as taxon k. $T_{i,k}$ is a binary indicator:

$$T_{i,k} = \begin{cases} 1, & \text{if specimen } i\text{'s label is taxon } k \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

We used the stochastic gradient descent optimizer with momentum (0.9). We adopted a stagewise learning rate schedule with an initial learning rate of 0.0009, decreased by half every two epochs. All models were trained for 20 epochs with a batch size of 10.

Fused model

Given a specimen X and its three modalities (x_1, x_2, and x_3), we fused the predictions of the three modalities by assuming conditional independence given the class label (28), i.e. $p(x_1, x_2, x_3 \mid y = k) = p(x_1 \mid y = k) \times p(x_2 \mid y = k) \times p(x_3 \mid y = k)$. Further, we assumed a uniform prior, i.e. $p(y = 1) = \cdots = p(y = K)$. Below is the fusion result:

$$p(y \mid x_1, x_2, x_3) \tag{3}$$

$$= \frac{p(x_1, x_2, x_3 \mid y)p(y)}{p(x_1, x_2, x_3)} \tag{4}$$

$$= \frac{p(x_1 \mid y)p(y)p(x_2 \mid y)p(y)p(x_3 \mid y)p(y)}{p(x_1, x_2, x_3)p(y)p(y)} \tag{5}$$

$$\propto p(y \mid x_1)p(y \mid x_2)p(y \mid x_3) \tag{6}$$

Equation 4 applies the Bayes' theorem, Eq. 5 assumes conditional independence, Eq. 6 assumes uniform prior w.r.t. class labels. To derive the probability after fusion, we normalized Eq. 6 to sum-to-one over the K classes, so that:

$$p(y = k \mid x_1, x_2, x_3) = \frac{p(y = k \mid x_1)p(y = k \mid x_2)p(y = k \mid x_3)}{\sum_{j=1}^{K} p(y = j \mid x_1)p(y = j \mid x_2)p(y = j \mid x_3)}.$$

Model validation

We assessed our classification models using five random splits of training (70%) and validation (30%) image sets. We report the mean of per-taxon accuracies. Confusion matrices were presented to provide a detailed report of per-taxon accuracies (Fig. 3).

Novelty detection

Uncertainty estimates

We used entropy (12) to assess whether a specimen should be considered as novel to the trained model. Given specimen X and its probability of being classified as taxon-k, $p(y = k \mid X)$ for $k = 1, \ldots, K$, we computed entropy $H(p)$ as

$$H(p) = -\sum_{k=1}^{K} p(y = k \mid X) \log p(y = k \mid X). \tag{7}$$

We compared novelty detection performance in each of the three CNN models and the FM. To measure the models' abilities to determine true novel morphologies, we used the ROC curve between true-positive rate (TPR) and false-positive rate (FPR). ROC is a graph showing the performance of a binary classification model at different thresholds. We computed the AUROC curve as the summary number to benchmark methods. Higher AUROC means better performance for novelty detection (5).

Optimal threshold selection

Although AUROC effectively summarizes model performance with varied thresholds, real-world systems require a single operation threshold. We used the method introduced in Unal (11) to find an optimal threshold $c*$ for fossil pollen analysis. We analyzed the ROC curve to find $c*$, using the TPR and FPR for a binary classification model designed to detect specimens of novel pollen types. We evaluated the performance of the model by using our training set to represent the known specimens and the pseudonovel set to represent the novel specimens. We optimized the following objective function (11) to obtain $c*$, the value at which TPR and
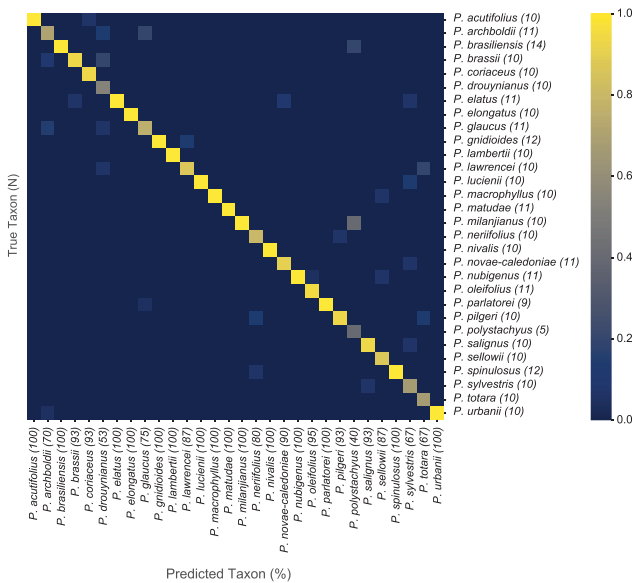

Fig. 3. Average confusion matrix depicting the classification accuracy of the FM for the modern *Podocarpus* species. Rows correspond to the true taxon (with number of specimens in parentheses), while columns represent the model's predictions (with the prediction accuracy in parentheses).

(1−FPR) values are closest to AUROC while minimizing their absolute difference:

$$c^* = \arg\min_{c \in \mathbb{R}} (|\text{TPR}(c) - \text{AUROC}| + |1 - \text{FPR}(c) - \text{AUROC}|). \quad (8)$$

Phylogenetic placement
Model and loss function

Specimens identified as novel were placed within an existing phylogeny using a trained embedding model, implemented as an MLP. MLPs are a special form of CNNs in which all the convolution layers have 1×1 kernels. For a novel specimen, the input to the MLP was a concatenation of the specimen's three classification feature vectors (with 2,048 dimensions each), extracted from the three classification CNNs. The MLP outputs a low-dimensional embedding feature vector $\mathbf{v} \in \mathbb{R}^{256}$ used to place the specimen within a phylogeny. Our MLPs have five convolutional layers. The first layer learns three nonnegative parameters to weigh the three classification features learned by individual CNNs. Between every two consecutive layers, we inserted a rectified linear activation unit (ReLU) for nonlinearity (29) and dropout layers, with a dropout rate of 0.2 to prevent overfitting (30). The complete network architecture is described in supplementary material (Fig. S1 and Table S3).

Our MLP is lightweight relative to a CNN, allowing us to train with all training specimens' features as a single batch. We used the Adam optimizer (31) with a constant learning rate of 0.00001 and a weight decay of 10^{-4}. We trained the MLP for 20,000 iterations (∼10 min running time on a GPU).

We developed a simple loss function to train the MLP to learn phylogenetic patterns from classification features. The loss function minimizes the error between ground-truth and predicted evolutionary distances separating pairs of specimens. Because the MLP outputs an embedding feature $\mathbf{v}_i$ for specimen i, between specimens i and j, we compute their Euclidean distance.

With the known ground-truth phylogenetic distance $D_{i,j}$, we minimize the mean square error (MSE) between $D_{i,j}$ and $\hat{D}_{i,j}$:

$$\mathcal{L}^{\text{MSE}} = \sum_{i,j} \left(D_{i,j} - \hat{D}_{i,j} \right)^2. \quad (9)$$

All the ground-truth and predicted phylogenetic distances in D and $\hat{D}$ were rescaled to the range [0, 1].

Tree construction

We constructed phylogenetic trees from the transformed features using Bayesian inference and assuming the Brownian motion (BM) model for character evolution (13–15). Analyses were performed using RevBayes (32). We performed Markov chain Monte Carlo analyses consisting of two independent runs with four chains each, running for 1 million generations. Trees were sampled every 1,000 generations, and initial prestationary generations were discarded according to the burn-in value determined with Tracer (v.1.7.1). Tree files from both independent runs were combined using LogCombiner (v.1.10.4) to generate the final phylogeny. The maximum clade credibility (MCC) tree, which maximizes the posterior probability of each individual clade, was chosen as the final model using TreeAnnotator (v.2.4.2). The resulting trees were visualized and edited using the functions plotTree and plotFBDTree in the R package "RevGadgets" (v.1.1.1) (33). Branches in the tree with a posterior probability ≥ 0.95 were considered significantly supported.

Pseudonovel evaluation experiments
Evaluation of novelty detection

To evaluate our models' ability to discriminate among known and novel taxa, we adopted a "leave-one-out" protocol (34), where all specimens of an extant species were excluded from the training dataset and treated as novel morphotypes. We ran five separate experiments where we excluded all specimens of a given species. We selected one of five taxa: *Podocarpus drouynianus*, *P. elongatus*, *P. neriifolius*, *P. oleifolius*, and *P. totara*. The five species were taken from five biologically meaningful subclades within the reference phylogeny. We refer to these excluded species as "pseudonovel taxa."

We forward-passed all specimens of a pseudonovel taxon where the species was excluded from model training) and the known validation specimens (where the species were included in model training) through the trained classification CNNs. We computed entropy over their class probability distributions and used ROC analysis to assess the trained model's performance in novelty detection. In each run, we also computed an optimal threshold. We averaged these thresholds to obtain a single threshold for identifying new fossil morphotypes.

Evaluation of phylogenetic placement

We extracted the CNN-learned features of all pseudonovel specimens and passed them to the trained MLPs. We then concatenated the resulting output features with the average feature results for the known species. This produced a continuous character matrix of pseudonovel specimens and known species. We used the character matrix to construct Bayesian inference trees under a single rate Brownian motion model. The trees were simulated under a birth–death model, with all parameters defined following Parins-Fukuchi (15). Because our modern dataset sampled ∼30% of extant *Podocarpus* taxa, the sampling rate was set to 0.3.

Visualization and quantification of feature transformations

We used *t*-distributed stochastic neighbor embedding (*t*-SNE) (35) to visualize the underlying structure of our FM classification features and our MLP-transformed features. Feature vectors were taken from the penultimate layer of both models. We analyzed the training and validation datasets independently to distinguish between the transformation results during training vs. evaluation (Fig. 4). We used Blomberg's *K* statistic (36) to measure the strength of the phylogenetic signal in our machine-learned features. Blomberg's *K* allowed us to directly compare the signal in classification features from the FM and features from the MLP embedding model.

Fossil application

We forward-passed the nine fossil *Podocarpidites* specimens through our fully trained CNN-MLP pipeline, where the model had been trained on all modern specimens. This produced a continuous character matrix of fossil specimens and extant species, which were then used to construct a Bayesian inference tree.

We performed tip-dating under the fossilized birth–death (FBD) process, enabling the simultaneous analysis of both modern and fossil morphotypes while calibrating divergence times estimates in a Bayesian framework (37, 38). Tip ages were included as single-point occurrence times corresponding to the most likely ages of the fossil specimens. Fossil ages were used to constrain the estimated ages of nodes in the tree.

The diversification and turnover rates were drawn from a uniform distribution over an interval ranging from 0 to 1. Speciation was defined as the sum of diversification and turnover rates, while
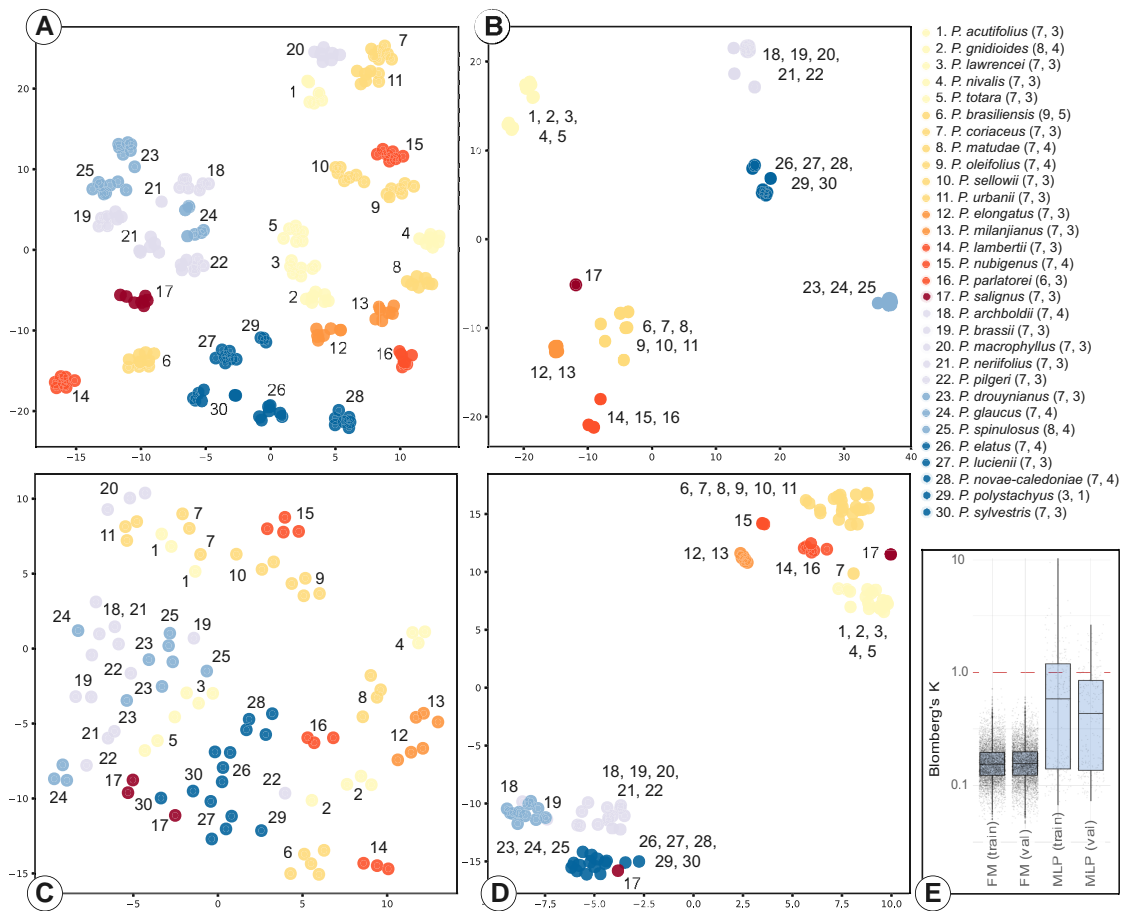
Fig. 4. Visualization of the FM and MLP features for both the training and validation sets, with a comparison of phylogenetic signal between FM and MLP features. Legend lists the species names and number of specimens in the training and validation sets. Species numbered 1–17 are within the subgenus *Podocarpus Podocarpus*. Species numbered 18–30 are within *P. Foliolatus*. The results demonstrate the degree to which the MLP transformation accentuated the phylogenetic signal within the features. A) *t*-SNE plot of FM features of the training set. B) *t*-SNE plot of MLP-transformed features of the training set. C) *t*-SNE plot of FM features of the validation set. D) *t*-SNE plot of MLP-transformed features of the validation set. E) Box plot of the strength of the phylogenetic signal of FM and MLP features quantified using Blomberg's *K*.

extinction was set to be equal to the turnover rate. The root age was drawn from a uniform distribution with a lower bound of 67.2 Ma, corresponding to the age of our oldest fossil, and an upper bound of 82 Ma, based on the maximum root age estimated in Quiroga et al. (39). Similarly, the fossil recovery rate was drawn from a uniform distribution ranging from 0 to 1. The sampling fraction of extant taxa was set at 0.3. The resulting time-calibrated MCC tree provided us with the most likely placements of our fossil specimens within the reference phylogeny and their corresponding posterior probabilities.

Results

Baseline classification accuracy

We measured taxon classification accuracies for the three imaging modalities (H-CNN, C-CNN, and P-CNN), and the FM. The FM and C-CNN produced the highest average accuracies, 90.40 and 90.60% average accuracy, respectively. The P-CNN scored lower (79.80%) and the H-CNN lower still (55.00%) (Figs. 3 and S2; Table S4).

Novel taxon detection

The C-CNN ($\mu = 0.9425$) and FM ($\mu = 0.8987$) markedly outperformed the P-CNN ($\mu = 0.8407$) and H-CNN ($\mu = 0.7104$) in detecting novel pollen morphotypes. The FM had the smallest variability overall ($\sigma = 0.0325$). The P-CNN and H-CNN had higher variability ($\sigma = 0.0725$ and 0.0591, respectively) (Fig. S3 and Table S5).

Phylogenetic placement of pseudonovel specimens

The results of the pseudonovel experiments demonstrated the method's ability to place taxa that have been previously unseen by the model. For three pseudonovel taxa (*P. drouynianus*, *P. oleifolius*, and *P. totara*), all specimens were placed in their correct respective subclade, with high support values ($P = 1$ in all three experiments) (Figs. S4.1, S4.4, and S4.5). For *P. elongatus*, the accuracy of the placement varied by specimen (Fig. S4.2). Six specimens were correctly placed with *P. milanjianus*, with fairly high support ($P = 0.935$). Three were placed elsewhere within subgenus *Podocarpus* and the remaining one was placed within subgenus *Foliolatus*. In the *P. neriifolius* experiments, seven specimens were correctly placed within the subclade comprising *P. brassii*, *P. archboldii*, *P. macrophyllus*, and *P. pilgeri*, with fairly high support ($P = 0.910$) (Fig. S4.3). The remaining three specimens were placed as sister to the entire *Foliolatus* subgenus.

The Bayesian inference trees constructed from the learned MLP features largely replicated the *Podocarpus* molecular phylogeny and faithfully reconstructed two well-supported subgenera

(*Podocarpus* subg. *Podocarpus* and *P.* subg. *Foliolatus*), with strong branch support for most clades across the trees (Figs. S4.1–S4.5). The MLP determined the optimal weights for combining features from the three modalities (H-CNN, C-CNN, and P-CNN). In all five pseudonovel experiments, the MLPs learned nearly equal weights for each modality (Table S6).

Feature transformations

The *t*-SNE plots illustrate how the MLP embedding model transformed the original FM classification features into phylogenetically informed ones (Fig. 4). There was little phylogenetic structure in the original FM features (Fig. 4A and C). Following the MLP transformation, specimens clustered in two distinct groups corresponding to the two *Podocarpus* subgenera and phylogenetically meaningful subgroupings within these larger clusters (Fig. 4B and D). The results for the training data show the effect of the MLP transformation (Fig. 4C). The results for the validation data show how well the trained embedding model generalized to new

Fig. 5. Combined tip-calibrated fossil tree simulated under the FBD process using the MLP-transformed features. Posterior probability values and ranges of nodes are illustrated by the colored bars. The tree topology replicates that of the reference phylogeny proposed by Leslie (21). Branch lengths were estimated based on our fossil tip ages. Fossils are italicized with ages in Ma in parentheses. The three oldest Panamanian specimens were placed as sister to the Central-South American tropical clade formed by *Podocarpus sellowii*, *P. coriaceus*, *P. brasiliensis*, *P. urbanii*, *P. oleifolius*, and *P. matudae*, while the youngest specimen was placed with the Chilean *P. salignus*. Likewise, the four Peruvian specimens are placed with *P. salignus*. Finally, the late Cretaceous Colombian specimen is placed as sister to the subgenus *P. Podocarpus*.

images. Only one image was misclassified in the validation dataset, an overexposed image of *P. salignus* (salignus08) (Fig. 4D).

Blomberg's *K* corroborates that there is more phylogenetic structure in the embedding model features. All 6,144 FM classification features had values <1 (therefore, not phylogenetically informative). *K* values for the 256 MLP embedding features ranged from 0.07 to 10.28 for the training dataset and 0.07 to 2.65 for the validation dataset. While the majority of MLP features were <1, 32.8% of the features of the training set and 18.4% of the validation set had values <1.

Fossil application

All nine fossil *Podocarpidites* specimens were placed with *Podocarpus* subgenus *Podocarpus* (Fig. 5). Notably, the fossil placements were less certain than the pseudonovel placements (Figs. 5, S4.1–S4.5, and S5).

The oldest specimen (Colombia, 67.2 Ma) was placed as sister to the entire subgenus, with strong support values (*P* = 1). The three oldest Panamanian specimens (10.15, 4.25, and 3.55 Ma) were placed as sister to the Neotropical clade formed by *P. matudae*, *P. urbanii*, *P. oleifolius*, *P. brasiliensis*, *P. sellowii*, and *P. coriaceus* (*P* = 1).

The youngest fossil and most taphonomically altered fossil (Panama, 2.05 Ma) was placed as sister to the Andean Chilean species *P. salignus*. The four Peruvian specimens (12 Ma) were also placed as sisters to *P. salignus*. The *P. salignus* and *Podocarpidites* clade was poorly supported (*P* = 0.5426).

Discussion

CNNs are powerful tools for the analysis of biological morphology. Both our results and previous research demonstrate their ability to capture the variability of pollen features and accurately classify species, even when there are limited morphological differences (9, 10, 40, 41).

Our study takes the application of machine learning in biological classification further, demonstrating that features captured by neural networks can be interpreted and used within an evolutionary framework to place fossil specimens within a phylogeny. Our phylogenetic embedding model uses the topology of our original molecular phylogeny to transform the classification features output by our fused classification model. This approach differs from existing total evidence analyses of pollen, e.g. (42), in that our morphological features are continuous and quantitative, and the original molecular data are intrinsic to the selection of our phylogenetic embedding features and the training of our MLP neural networks. While this specific study uses a reference phylogeny derived exclusively from molecular data, the embedding model could utilize any well-supported tree, including total evidence phylogenies constructed using both molecular and morphological data.

Molecular phylogenetic studies place the origin of *Podocarpus* in the Late Cretaceous (39) or Paleocene (21, 43, 44), with more uncertain timing for the divergence of the two subgenera, *P.* subg. *Podocarpus* and *P.* subg. *Foliolatus*. The underlying uncertainty in published phylogenies introduced noise into our MLP models. Not all nodes of the original reference tree were well supported or fully resolved (21), and there were disagreements between the phylogeny used to train our models (21) and an alternative recent phylogeny (39). Most notably, the African clade within *P. Podocarpus* is more closely related to the tropical Central-South American clade in our reference phylogeny (21), while in Quiroga et al. (39), the clade is more closely related to subtropical
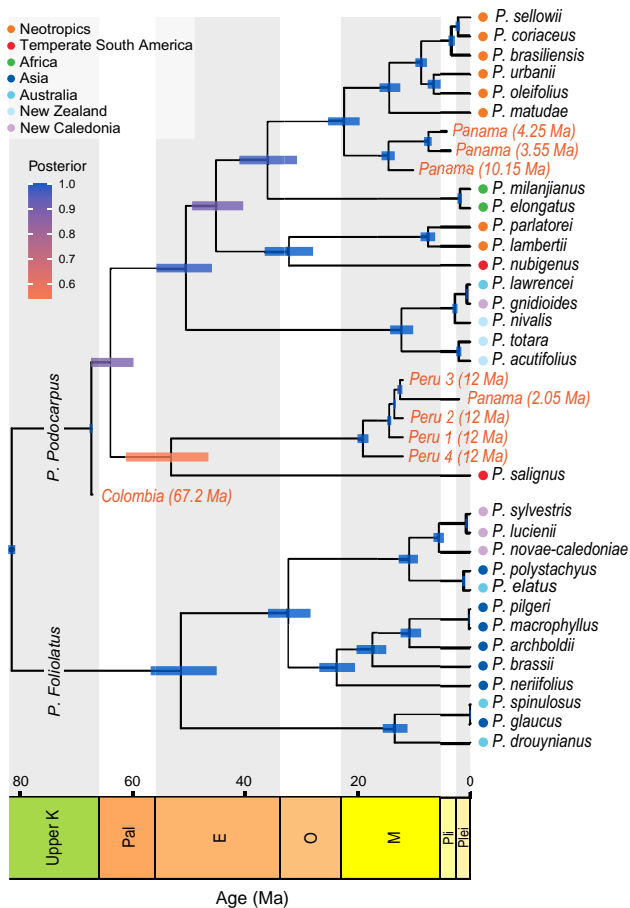
South American species. Additionally, in Quiroga et al. (39), the temperate South American species *P. nubigenus* and *P. salignus* are more closely related to the Australasian taxa of *P. Podocarpus* (*P. nubigenus*, *P. lawrencei*, *P. acutifolius*, *P. totara*, *P. gnidioides*, and *P. nivalis*). In contrast, in our reference phylogeny, *P. salignus* is placed as sister to the subgenus *P. Podocarpus* and *P. nubigenus* is more closely related to the tropical–subtropical South American species of *P. parlatorei* and *P. lambertii* (21).

Podocarpus pollen is bisaccate with large sacci that are coarsely endoreticulate. Visual differences among species are subtle and difficult to consistently identify. As a result, palynologists generally do not attempt to identify individual species (45–47). However, we achieved high classification accuracies with few per-species examples (~10 per species) by using cropped patches and cross-sectional images in addition to whole-grain images.

When a limited number of images are naively used as a training sample, the results of a CNN analysis will suffer from overfitting (48, 49). That is, the trained model optimizes to these few training images and generalizes poorly to testing images. In our analysis, this was the case for the H-CNN. With 309 whole-grain images for 30 species, the H-CNN only achieved 55.00% accuracy (Table S2). Including patches and cross-sectional images derived from our 3D scans increased the size of our image dataset to 3,384 patches and 5,483 cross-sections. As a result, our P-CNN and C-CNN models were able to effectively mitigate the issue of overfitting and achieved 90.60 and 79.80% accuracy, respectively.

Many misclassifications were of closely related species (Figs. 3 and S2). *P. archboldii*, *P. lawrencei*, *P. neriifolius*, *P. sylvestris*, *P. drouynianus*, *P. pilgeri*, and *P. totara* were confused with their sister species or species within the same subclade. Several misclassifications, however, were among distantly related species. *P. polystachyus* was misclassified as *P. milanjianus* and *P. brasiliensis*, likely due to its small sample size ($n = 5$). Additionally, *P. glaucus* was confused for *P. archboldii* and *P. parlatorei*, *P. drouynianus* was confused for *P. brassii*, *P. archboldii*, and *P. lawrencei*, and *P. totara* was confused for *P. pilgeri*. Although we cannot identify specific morphological features which explain these misidentifications, we suspect that they are due to similarities in the endoreticulate patterning of their sacci, resulting from convergence or homoplasy.

ROC analysis successfully established an uncertainty threshold for identifying taxa that were new or unknown to the model. In our pseudonovel experiments, where all specimens of a species with a known phylogenetic placement were left out of the training of our models, the C-CNN and FM were the least variable in recognizing pseudonovel specimens, with consistently high AUROC values. We expected the internal structure of the pollen wall to retain phylogenetically informative characters for identifying novel morphotypes based on prior research (10). The P-CNN produced both the highest (i.e. *P. oleifolius*) and lowest (i.e. *P. elongatus*) AUROC values and was, therefore, the least reliable in detecting new species.

The MLP embedding model correctly placed 44 out of 51 pseudonovel specimens in our Bayesian inference trees, including all specimens of *P. drouynianus*, *P. oleifolius*, and *P. totara* (Figs. S4.1–S4.5). We cannot be certain why four *P. elongatus* specimens were inaccurately placed, but the sacci of the three erroneously placed *P. neriifolius* grains were obscured by their orientation.

MLP-transformed features closely reproduced the molecular phylogeny used to train our models. The embedding model transformation effectively extracted the phylogenetic structure within the original FM CNN features (Fig. 4). Our fully trained models also produced plausible placements of fossil specimens. The fossil pollen morphospecies *Podocarpidites* is associated with *Podocarpus*, although other taxa—namely the extant New Zealand species *Lepidothamnus laxifolius* and Laurasian Paleogene Pinaceae species—have similar morphologies (47, 50). Our *Podocarpidites* fossils are all Neotropical and Neogene or younger, with the exception of one Late Cretaceous specimen, so we were confident in assuming that they were fossil *Podocarpus*.

The limited morphological differentiation among *Podocarpus* species makes interpreting the placement of our fossil specimens difficult. However, the spatiotemporal pattern that we observe is consistent with the biogeographic history of *P. Podocarpus* (39). In our tip-calibrated tree, the 67.2 Ma Colombian specimen was placed as sister to the entire subgenus *P. Podocarpus*, suggesting a lineage now extinct in the Neotropics. The Miocene Peruvian specimens were placed with the early-diverging Chilean species, *P. salignus*. The four Panamanian specimens were the youngest fossils in our analysis. As would be predicted, three specimens (3.55, 4.25, and 10.15 Ma) were placed with high support with the main Neotropical lineage (*P. sellowii*, *P. coriaceus*, *P. brasiliensis*, *P. urbanii*, *P. oleifolius*, and *P. matudae*).

The youngest Panamanian specimen (2.05 Ma) was highly corroded, with few visible features on the sacci or corpus. We suspect that its unlikely placement with *P. salignus* is the result of this poor preservation. Its placement, and the placement of the four Miocene (12 Ma) Peruvian specimens, with the Chilean species *P. salignus* had poor support. However, excluding the specimen did not affect the placement of the Peruvian fossils (Fig. S5). The low node support may instead reflect the uncertainty in the placement of *P. salignus* within *P. Podocarpus* (21, 39).

A potential consequence of the limited number of examples per species in our training (one individual per species) is the underestimation of intraspecific variability within *Podocarpus* and corresponding overestimation of the novelty of new specimens. Pollen morphology is known to vary between individuals as a result of differences in environment and genotype, although documentation of this variability in the literature is limited. One recent study of 23 genotypes and four populations of *Abies alba* showed limited difference in exine morphology, but variation in the length of the polar axis (51). This suggests that size was the primary source of intraspecific variation.

Our analysis is size independent, with all images uniformly scaled to 224×224 pixels, so our results should be less sensitive to variability in size. Despite this, larger image datasets would better characterize the range of morphological diversity within species and only improve the accuracy of trained models for phylogenetic assessment. Current public repositories of high-resolution pollen images, e.g. PalDat (52), often only record images of a single pollen grain per species. Pollen databases for machine learning, e.g. POLLEN73S (53), are primarily of lower quality transmitted light images. The expansion of machine learning in palynological research will rely on the development and investment in shared high-resolution image databases that capture the full range of pollen diversity that exist in herbaria and research collections around the world.

Conclusion

Palynologists have long used the conservation of morphological features within clades to classify pollen taxa. However, our results suggest that an even deeper evolutionary history is contained within pollen morphology. The vocabulary of palynology is vast, but was developed to describe the features prominent in transmitted light microscopy (54). Confocal superresolution microscopy captures a larger range of the gradations and permutations in external

shape and ornamentation and internal pollen structure (55). The extensive number of features captured by superresolution images facilitate training of deep neural networks (10). More features allow deep learning to connect morphology to molecular phylogenies, learn evolutionary distances from existing phylogenies, and transform characters that define individual taxa to the synapomorphies that connect them. Deep learning is adaptive to the constraints of paleontological data, including incomplete or unresolved phylogenies and fossil morphotypes derived from extinct taxa. Novelty detection recognizes new morphologies and expands our knowledge of a clade's morphospace, by explicitly quantifying our uncertainty. The abundance of fossil pollen in geological samples also means that phylogenetic determinations can be made with populations, not just individual specimens.

Other fossil taxa—with phylogenetically informative morphology that can be imaged consistently—could be analyzed with this approach. Significant morphological convergence would confound the analysis, so the best candidates would be taxa for which the relationship between morphology and phylogeny are well established. This includes both plant and animal clades with existing image databases, including foraminifera (56), bivalves (57), and plant phytoliths (58).

Applied at a broad scale, phylogenetically informed machine-learning models will allow researchers to harness the evolutionary information within the fossil record. Morphology can be seamlessly incorporated into phylogenetic analyses, allowing us to identify taxon origins and extinctions that would otherwise go unrecognized.

Acknowledgments

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

Author Contributions

M.-E.A., S.K., and S.W.P. designed the research. M.E.A. conducted the analysis. S.W.P. and S.K. directed the research. S.W.P., M.E.A., and S.K. wrote the manuscript.

Preprints

A preprint of this article is published at https://doi.org/10.1101/2023.07.09.545296.

Data Availability

The Airyscan images used in this study were submitted to the Illinois Databank (https://doi.org/10.13012/B2IDB-8817604˙V1). The codebase is publicly available on GitHub (DOI: 10.5281/zenodo.8128443) and can be accessed using this link: https://github.com/paleopollen/Novel_Pollen_Phylogenetic_Placement.

References

1   LeCun Y, Bengio Y. 1995. Convolutional networks for images, speech, and time series. In: Arbib MA, editor. *The handbook of brain theory and neural networks*. Cambridge, MA: MIT Press. p. 255–2583361. https://doi.org/10.5555/303568.303704.

2   Krizhevsky A, Sutskever I, Hinton GE. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*. 60(6):84–90. https://doi.org/10.1145/3065386.

3   Cuthill JFH, Guttenberg N, Ledger S, Crowther R, Huertas B. 2019. Deep learning on butterfly phenotypes tests evolution's oldest mathematical model. *Sci Adv*. 5(8):eaaw4967.

4   Hendrycks D, Gimpel K. 2017. A baseline for detecting misclassified and out-of-distribution examples in neural networks. In: 5th International Conference on Learning Representations. April 24–26, 2017, Toulon, France.

5   Kong S, Ramanan D. 2021. Opengan: open-set recognition via open data generation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV 2021). Virtual Conference. October 11–17, 2021. https://doi.org/10.1109/ICCV48922.2021.00085813822.

6   Mander L, Punyasena SW. 2014. On the taxonomic resolution of pollen and spore records of earth's vegetation. *Int J Plant Sci*. 175(8):931–945.

7   Birks HJB, *et al.* 2016. Does pollen-assemblage richness reflect floristic richness? A review of recent developments and future challenges. *Rev Palaeobot Palynol*. 228:1–25.

8   Traverse A. 2007. *Paleopalynology*. Vol. 28. Berlin, Germany: Springer Science & Business Media.

9   Punyasena SW, Tcheng DK, Wesseln C, Mueller PG. 2012. Classifying black and white spruce pollen using layered machine learning. *New Phytol*. 196(3):937–944.

10 Romero IC, *et al.* 2020. Improving the taxonomy of fossil pollen using convolutional neural networks and superresolution microscopy. *Proc Natl Acad Sci U S A*. 117(45):28496–28505.

11 Unal I. 2017. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med*. 2017: 3762651. https://doi.org/10.1155/2017/3762651.

12 Jafarian M, Soroudi A, Keane A. 2020. Resilient identification of distribution network topology. *IEEE Trans Power Deliv*. 36(4): 2332–2342.

13 Felsenstein J. 1973. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Syst Biol*. 22(3):240–249.

14 Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125(1):1–15.

15 Parins-Fukuchi C. 2018. Use of continuous traits can improve morphological phylogenetics. *Syst Biol*. 67(2):328–339.

16 Punyasena SW, Urban MA, Adaïmé M-E, Romero I, Jaramillo C. 2023. Pollen of Podocarpus (Podocarpaceae): airyscan confocal superresolution images. University of Illinois Urbana-Champaign, Urbana, IL, USA. https://doi.org/10.13012/B2IDB-8817604_V1.

17 Jaramillo CA, *et al.* 2014. Palynological record of the last 20 Million years in Panama. In: Stevens WD, Montiel OM, Raven P, editors.

Paleobotany and biogeography: a festschrift for Alan Graham in His 80th Year. St. Louis: Missouri Botanical Garden Press. p. 134–253.

18 Martínez C, *et al.* 2020. Neogene precipitation, vegetation, and elevation history of the central andean plateau. *Sci Adv*. 6(35): eaaz4724.

19 Carvalho MR, *et al.* 2021. Extinction at the end-cretaceous and the origin of modern neotropical rainforests. *Science*. 372(6537): 63–68.

20 Tropicos.org. Missouri botanical garden; 2022 [accessed 2022 Jun 20]. https://www.tropicos.org.

21 Leslie AB, *et al.* 2018. An overview of extant conifer evolution from the perspective of the fossil record. *Am J Bot*. 105(9): 1531–1544.

22 Hill RS, Carpenter RJ. 1991. Evolution of Acmopyle and Dacrycarpus (Podocarpaceae) foliage as inferred from macrofossils in south-eastern Australia. *Aust Syst Bot*. 4(3):449–479.

23 Zuiderveld K. 1994. Contrast limited adaptive histogram equalization. In: *Graphic Gems IV*. San Diego (CA): Academic Press Professional, Inc. p. 474–485.

24 Kong S, Punyasena S, Fowlkes C. 2016. Spatially aware dictionary learning and coding for fossil pollen identification. In: Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW 2016). Las Vegas, NV, June 26–July 01, 2016. https://doi.org/10.1109/CVPRW.2016.165.

25 Xie S, Girshick R, Dollár P, Tu Z, He K. 2017. Aggregated residual transformations for deep neural networks. In: 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). Honolulu, HI, July 21–26, 2017. https://doi.org/10.1109/CVPR.2017.634.

26 Deng J. 2009. ImageNet: a large-scale hierarchical image database. In: CVPR: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami Beach, FL, June 20–25, 2009. https://doi.org/10.1109/CVPR.2009.5206848.

27 He K, Girshick R, Dollar P. 2019. Rethinking imagenet pre-training. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV 2019). Seoul, South Korea, October 27–November 02, 2019. https://doi.org/10.1109/ICCV.2019.00502.

28 Chen Y-T, Shi J, Ye Z, *et al.* 2022. Multimodal object detection via probabilistic ensembling. In: European Conference on Computer Vision (ECCV). Tel Aviv, Israel, October 23–27, 2022. https://doi.org/10.1007/978-3-031-20077-9_9.

29 Nair V, Hinton GE. 2010. Rectified linear units improve restricted Boltzmann machines. In: 27thInternational Conference on Machine Learning (ICML 2010). Haifa, Israel, June 21–24, 2010. p. 807–814.

30 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 15(1):1929–1958.

31 Kingma DP, Ba J. 2014. Adam: a method for stochastic optimization, arXiv, arXiv:1412.6980, preprint: not peer reviewed.

32 Höhna S, *et al.* 2016. Revbayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol*. 65(4):726–736.

33 Tribble CM, *et al.* 2022. Revgadgets: an R package for visualizing Bayesian phylogenetic analyses from RevBayes. *Methods Ecol Evol*. 13(2):314–323.

34 Kozlov AM, Zhang J, Yilmaz P, Glöckner FO, Stamatakis A. 2016. Phylogeny-aware identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Res*. 44(11):5022–5033.

35 Van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J Mach Learn Res*. 9(11):2579–2605.

36 Blomberg SP, Garland T Jr, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*. 57(4):717–745.

37 Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proc Natl Acad Sci U S A*. 111(29):E2957–E2966.

38 Zhang C, Stadler T, Klopfstein S, Heath TA, Ronquist F. 2016. Total-evidence dating under the fossilized birth–death process. *Syst Biol*. 65(2):228–249.

39 Quiroga MP, Mathiasen P, Iglesias A, Mill RR, Premoli AC. 2016. Molecular and fossil evidence disentangle the biogeographical history of Podocarpus, a key genus in plant geography. *J Biogeogr*. 43(2):372–383.

40 Boldeanu M, Cucu H, Burileanu C, Muarmureanu L. 2021. Automatic pollen classification using convolutional neural networks. In: 2021 44th International Conference on Telecommunications and Signal Processing (TSP). Virtual Conference, July 26–28, 2021. https://doi.org/10.1109/TSP52935.2021.9522626.

41 Sevillano V, Aznarte JL. 2018. Improving classification of pollen grain images of the POLEN23E dataset through three different applications of deep learning convolutional neural networks. *PLoS One*. 13(9):e0201807.

42 Woutersen A, *et al.* 2023. The evolutionary history of the Central Asian steppe-desert taxon Nitraria (Nitrariaceae) as revealed by integration of fossil pollen morphology and molecular data. *Bot J Linn Soc*. 202(2):195–214.

43 Biffin E, Conran JG, Lowe AJ. 2011. Podocarp evolution: a molecular phylogenetic perspective. In: Turner BL. Cernusak LA, editors. *Ecology of the Podocarpaceae in tropical forests, Smithsonian Contributions to Botany*. Washington (DC): Smithsonian Institution Scholarly Press. p. 1–2095.

44 Leslie AB, *et al.* 2012. Hemisphere-scale differences in conifer evolutionary dynamics. *Proc Natl Acad Sci U S A*. 109(40): 16217–16221.

45 Pocknall DT. 1981. Pollen morphology of the New Zealand species of *Dacrydium selander*, *Podocarpus l'heritier*, and *Dacrycarpus endlicher* (podocarpaceae). *N Z J Bot*. 19(1):67–95.

46 Hooghiemstra H, Wijninga VM, Cleef AM. 2006. The paleobotanical record of Colombia: implications for biogeography and Biodiversity1. *Ann Missouri Bot Gard*. 93(2):297–325.

47 Morley RJ. 2011. Dispersal and paleoecology of tropical podocarps. In: *Ecology of the Podocarpaceae in tropical forests, Smithsonian contributions to botany*. Washington (DC): Smithsonian Institution Scholarly Press. p. 21–4195.

48 Chen W-Y, Liu Y-C, Kira Z, Wang Y-CF, Huang J-B. 2019. A closer look at few-shot classification. In: 7th International Conference on Learning Representations (ICLR 2019). New Orleans, LA, May 6–9, 2019.

49 Wang Y, Yao Q, Kwok JT, Ni LM. 2020. Generalizing from a few examples: a survey on few-shot learning. *ACM Comput Surv*. 53(3): 1–34.

50 Greenwood DR, Hill CR, Conran JG. 2013. Prumnopitys anglica sp. nov.(Podocarpaceae) from the Eocene of England. *Taxon*. 62(3): 565–580.

51 Wrońska-Pilarek D, *et al.* 2020. Pollen morphology and variability of *Abies alba* Mill. Genotypes from South-Western Poland. *Forests*. 11(11):1125.

52 Weber M, Ulrich S. 2017. Paldat 3.0—second revision of the database, including a free online publication tool. *Grana*. 56(4): 257–262.

53 Astolfi G, *et al.* 2020. Pollen73s: an image dataset for pollen grains classification. *Ecol Inform*. 60:101165.

54 Punt W, Hoen PP, Blackmore S, Nilsson S, Thomas AL. 2007. Glossary of pollen and spore terminology. *Rev Palaeobot Palynol.* 143(1–2):1–81.

55 Romero IC, Urban MA, Punyasena SW. 2020. Airyscan superresolution microscopy: a high-throughput alternative to electron microscopy for the visualization and analysis of fossil pollen. *Rev Palaeobot Palynol.* 276:104192.

56 Marchant R, Tetard M, Pratiwi A, Adebayo M, de Garidel-Thoron T. 2020. Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *J Micropalaeontol.* 39(2):183–202.

57 Edie SM, Collins KS, Jablonski D. 2023. High-throughput micro-ct scanning and deep learning segmentation workflow for analyses of shelly invertebrates and their fossils: examples from marine Bivalvia. *Front Ecol Evol.* 11:1127756.

58 Hovsková K, Neustupa J, Pokornỳ P, Pokorná A. 2022. Phylogenetic, ecological and intraindividual variability patterns in grass phytolith shape. *Ann Bot.* 129(3):303–314.