Earth-Science Reviews 252 (2024) 104765

Contents lists available at ScienceDirect

Earth-Science Reviews

journal homepage: www.elsevier.com/locate/earscirev

Review Article

Artificial intelligence in paleontology

Congyu Yu[a,b,c,*], Fangbo Qin[d], Akinobu Watanabe[e,c,f], Weiqi Yao[g], Ying Li[d], Zichuan Qin[h,i], Yuming Liu[h], Haibing Wang[j], Qigao Jiangzuo[j,k], Allison Y. Hsiang[l], Chao Ma[a,b], Emily Rayfield[h], Michael J. Benton[h], Xing Xu[m,n,j]

[a] *State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation & Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China*

[b] *Key Laboratory of Deep-time Geography and Environment Reconstruction and Application of Ministry of Natural Resources, Chengdu University of Technology, Chengdu 610059, China*

[c] *Division of Paleontology, American Museum of Natural History, New York, NY 10024, USA*

[d] *Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*

[e] *Department of Anatomy, New York Institute of Technology College of Osteopathic Medicine, Old Westbury, NY 11568, USA*

[f] *Life Science Department, Natural History Museum, London SW7 5BD, UK*

[g] *Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China*

[h] *Palaeobiology Research Group, School of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK*

[i] *School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham B15 2TT, UK*

[j] *Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China*

[k] *School of Earth and Space Sciences, Peking University, Beijing 100087, China*

[l] *Department of Geological Sciences, Stockholm University, Svante Arrhenius väg 8, Stockholm 10691, Sweden*

[m] *Centre for Vertebrate Evolutionary Biology, Yunnan University, Kunming 650091, China*

[n] *Paleontological Museum of Liaoning, Shenyang Normal University, Liaoning Province, 253 North Huanghe Street, Shenyang 110034, China*

ARTICLE INFO

Keywords:
Paleontology
Fossil
Artificial intelligence
Machine learning
Deep learning
Classification
Segmentation
Prediction

ABSTRACT

The accumulation of large datasets and increasing data availability have led to the emergence of data-driven paleontological studies, which reveal an unprecedented picture of evolutionary history. However, the fast-growing quantity and complication of data modalities make data processing laborious and inconsistent, while also lacking clear benchmarks to evaluate data collection and generation, and the performances of different methods on similar tasks. Recently, artificial intelligence (AI) has become widely practiced across scientific disciplines, but not so much to date in paleontology where traditionally manual workflows have been more usual. In this study, we review >70 paleontological AI studies since the 1980s, covering major tasks including micro- and macrofossil classification, image segmentation, and prediction. These studies feature a wide range of techniques such as Knowledge-Based Systems (KBS), neural networks, transfer learning, and many other machine learning methods to automate a variety of paleontological research workflows. Here, we discuss their methods, datasets, and performance and compare them with more conventional AI studies. We attribute the recent increase in paleontological AI studies most to the lowering of the entry bar in training and deployment of AI models rather than innovations in fossil data compilation and methods. We also present recently developed AI implementations such as diffusion model content generation and Large Language Models (LLMs) that may interface with paleontological research in the future. Even though AI has not yet been a significant part of the paleontologist's toolkit, successful implementation of AI is growing and shows promise for paradigm-transformative effects on paleontological research in the years to come.

* Corresponding author at: State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation & Institute of Sedimentary Geology, Chengdu University of Technology, Chengdu 610059, China.

E-mail address: congyuyu@cdut.edu.cn (C. Yu).

https://doi.org/10.1016/j.earscirev.2024.104765
Received 6 August 2023; Received in revised form 6 February 2024; Accepted 25 March 2024
Available online 2 April 2024

1. Introduction

1.1. Data-driven earth sciences

Recently, artificial intelligence (AI) has shown fast-growing applications in a wide range of fields in earth sciences and elucidates their transformation towards data-driven studies (Bergen et al., 2019; Reichstein et al., 2019; Cheng et al., 2020; Sun et al., 2022). Global hydrology (Yao et al., 2023), weather forecasting (Bi et al., 2023; Zhang et al., 2023), seismology (Wu et al., 2019, Kong et al., 2019, Mousavi and Beroza, 2023), remote sensing (Pires de Lima and Marfurt, 2020), carbon cycle (Tao et al., 2023), and many other subfields in earth sciences have benefited from AI advances. However, challenges in data collection and processing, task complexity, and a lack of suitable models means that a large number of earth science fields, including paleontology, have primarily relied on traditional manual workflows.

Large-scale datasets and complex statistical methods have enabled deep-time high-resolution evolutionary models of both global and local biodiversity from marine and terrestrial faunas (Fan et al., 2020, Zhou et al., 2021, Dai et al., 2023), and also offer an indispensable opportunity to evaluate the causes and processes of series of bio-geological events. Reconstructing the co-evolution of life and environment through time is the key to understanding earth history. Specifically, the "Big Five" mass extinction events have been observed across the Ordovician–Silurian, late Devonian, Permian–Triassic, Triassic–Jurassic, and Cretaceous–Paleogene boundaries (Sepkoski, 1978, 1979, 1984; Benton, 1995; Payne and Finnegan, 2007), with profound perturbations in biogeochemical processes. While previous studies have taken a diversity of geochemical and modeling approaches to reveal evolution across time and space, the ability to integrate various approaches to provide comprehensive analysis and underlying mechanism is still limited. Recently, the community started to pay attention to data-driven studies that incorporate large-scale datasets and novel data analysis methods, e.g., deep learning. For sedimentary studies, Koeshidayatullah et al. (2020) developed deep learning-based method to automate image classification and object detection in carbonate petrographic images. Emmings et al. (2022) applied text mining coupled with multivariate statistical analysis to a large library of published sedimentary datasets from the Precambrian to the present to explore pyrite morphology and chemistry as paleo-redox proxies. In ocean chemistry studies, Mete et al. (2023) developed Gaussian Progress Regression machine learning models to accurately simulate the spatial and vertical distribution of trace elements (e.g., barium) in the sea.

However, the scarcity of earth science samples from certain temporal or spatial regions means that enormous gaps exist, and while these may be approximated by extrapolation, the risk is that such approaches masking fine-scale changes. To date, systematic statistical analysis has not yet been completed based on big data and there are few common datasets for model evaluation. Compared with other earth science fields, there are far fewer AI-based paleontological studies. In this review, we show the development of AI-based paleontological studies since the 1980s, covering distinctive tasks and groups of organisms, with a particular focus on the evolution of datasets and algorithms.

1.2. Data-driven paleontology

Since paleontology concerns ancient organisms, it relies based primarily on fossil materials. However, fossil records are extremely patchy fragments of total evolutionary history, and large-scale quantitative paleontological studies have only been feasible when relevant datasets are compiled (Sepkoski, 1978, 1979, 1984; Benton, 1995; O'Leary et al., 2013; Hsiang et al., 2019). A large portion of traditional paleontological studies focus on fossil morphology, including anatomical description and comparative studies among specimens, which can be considered as "fossil-driven" studies. On the other hand, "data-driven" paleontological studies can be characterized as studies that work with relatively large number of fossil specimens, and apply various analytical techniques to discover patterns from data. Ultimately, all paleontological studies are based on fossil specimens, thus the boundary between "fossil-driven" and "data-driven" cannot always be well defined, and we do not attempt to do so here. However, is has been observed that progress in paleontological data collection, application of advanced analytical methods, and increasing availability of data sharing and distribution have resulted in an increase in the number of paleontological studies that follow the "data-driven" paradigm (Sepkoski, 2013; Smith et al., 2023).

Early examples of "data-driven" paleontological studies can be traced back to Matthew (1926), who studied horse macroevolution by compiling the fossil records of horses and their early relatives, globally and systematically comparing morphological changes in teeth, limbs and skulls, temporal and spatial distribution, and tempo and mode of their evolution. Although Matthew (1926) did not carry out any quantitative analysis, such a systematic summary and comparison of horse evolution presented a primitive form of data-driven paleontological research. Building on examples such as these from horse evolution, Simpson (1944) was one of the first to present a comprehensive proposition to convert paleontological data into numerical, statistical problems to explore rates and modes of evolution. His examples were mainly from mammalian evolution, but he extended his ideas across models of macroevolution and biogeography of all fossil groups.

Later, with much more abundant fossil records, historical marine biodiversity became a focus for data-driven paleontological studies. A classic example is the series of works on Phanerozoic marine biodiversity and major extinction events by Sepkoski (1978, 1979, and 1984) and Raup and Sepkoski (1982). Quantitative phylogenetic developed much more rapidly since the 1960s. Numerical taxonomy (= phenetics) was proposed as a means to compare characteristics across taxa to quantitatively reconstruct their evolutionary history (Sneath and Sokal, 1962, 1973). Numerical taxonomy introduced a series of quantitative methods to eliminate influences from intuitive judgements and allow studies to be reproducible. Along the same lines, Hennig (1966) launched the study of cladistics at almost the same time. While both phenetics and cladistics use morphological characters as input data, the former clusters taxa based on overall morphological similarity while the latter clusters taxa based on evolutionary relationships (i.e., inferring monophyletic clades using synapomorphies, or shared derived character states). The use of phenetics has been superseded by cladistics, which is used in paleontological studies to infer phylogenies along with modern statistical phylogenetic frameworks (i.e., maximum likelihood and Bayesian methods).

Ancient DNA only covers a tiny sliver of evolutionary history, with the oldest sample sequenced dating only back to 2.4 Ma (Kjær et al., 2022), which is very young by geological standards. Nevertheless, the field of "paleo-bioinformatics", derived from "bioinformatics", is valuable in exploring details of human and late Quaternary to Holocene taxa on the basis of information theory and morphological data (Yu et al., 2021). In general, fossil morphology is the sole resource for a large portion of paleontological studies. The compilation of morphological data has made more quantitative studies feasible and especially through the introduction of morpho-space and other quantitative tools. Raup (1966) analyzed the shape of different theoretical shell forms by illustrating specimens in a space given by three geometric measurements and now is often seen as an early form of morpho-space (Budd, 2021). During the decades since these works, methodological advancements in quantitative characterization, visualization, and analysis of biological form have truly revolutionized morphological research, including paleontology. Chief among these advances is the Procrustean paradigm in geometric morphometrics (GM), which uses Cartesian coordinates to mathematically describe differences and changes in shape (Bookstein, 1992; Rohlf and Marcus, 1993; Slice, 2007; Adams et al., 2013). Coupled with non-destructive 3D imaging techniques in fossils, GM is common practice in quantitative comparative morphological studies and has been widely used in various groups of organisms such as dinosaurs

(Bhullar et al., 2012; Hanson et al., 2021; Choiniere et al., 2021) and mammals (Lungmus and Angielczyk, 2019; Goswami et al., 2022). These highly multivariate, or multidimensional, anatomical data have allowed paleontologists to investigate the tempo and mode of phenotypic evolution with much greater morphological fidelity than before.

We cannot list every kind of quantitative paleontological study here, and many more are included in textbooks and software (e.g., Hammer and Harper, 2001, 2008). While the amount, modalities, and scopes of fossil data have increased rapidly during the last six decades, and we have witnessed a variety of data-driven paleontological studies on diversified topics and organisms, paleontological AI applications remain scarce.

1.3. Classic AI models and tasks

Although ideas of machine-aided analyses had been suggested earlier, AI was first proposed as a scientific idea at a workshop held in Dartmouth College (New Hampshire, USA) in 1956, which was the groundbreaking landmark of this field (McCarthy et al., 2006). There have been ups and downs in the development of AI since 1956 following advances in hardware, algorithms, and mismatches between expectation and reality. There have been several reviews of the history of AI and its subfields such as deep learning to which we refer the reader for further details (LeCun et al., 2015; Jordan and Mitchell, 2015; Baraniuk et al., 2020; Bengio et al., 2021; Toosi et al., 2021; Xu et al., 2021).

During its growth, novel ideas in AI have frequently been proposed and rejected; for example, the Knowledge-Based System (KBS) was once popular but has now mostly been abandoned (Bell, 1985). A KBS, or Expert System, encompasses a knowledge base and an inference engine. It was the very first AI model used in paleontology (Beightol and Conrad, 1988; Swaby, 1990, 1992). In fossil KBS, there was usually a database of annotated subjects, such as fossil specimen images, and a set of predetermined discriminate rules. New specimens can be identified or classified based on given rules and the rules themselves may be modified and updated. Most of these KBSs are based on descriptive rules and may store a certain number of images for illustration reference.

Traditional handcrafted models including KBS have very limited adaptability, complexity, and optimality. As such, machine learning is the current mainstream AI theory, which allows an agent to optimize its performance on a certain task by learning from experience through provision of training data sets. The learning paradigms include supervised learning, semi-supervised learning, unsupervised learning, and reinforcement learning (Fig. 1). Typically, a machine learning model (e.g., Gaussian mixture model, support vector machine (SVM), neural network, random forest, etc.) is trained on a dataset, during which the model's output is evaluated at each iteration according to ground truths and the model's parameters are gradually tuned according to the learning objective.

Deep learning is a subfield of machine learning methods that started with Hinton et al. (2006), Hinton and Salakhutdinov (2006) and Bengio and LeCun (2007), but its core idea was presented much earlier (e.g., LeCun et al., 1989). The most popular deep learning architecture, deep neural network (DNN), arranges interconnected nodes in a layered structure that resembles neurons in the brain and can approximate any given function at reasonable cost (Hornik et al., 1989). With the availability of large training datasets and hardware, deep learning has boomed since 2012 (Krizhevsky et al., 2012; Goodfellow et al., 2014; LeCun et al., 2015). The multiple layered (up to >100) structure and representation learning allow a DNN to learn to perceive rich, complex, and hierarchical feature representations from unstructured data. For example, given an image, a DNN can extract the low-level features (color, texture, edge, etc.), the mid-level features (shape, parts, etc.) and high-level features (semantics, category, context, etc.). Compared to traditional machine learning techniques, deep learning has absolute superiority in description and generalization abilities, especially when large datasets with annotation are available for training.

Classic AI tasks can be categorized roughly into classification, segmentation, prediction, etc. Classification sorts inputs into different categories according to a given classification scheme (supervised learning) or inherent patterns in data (unsupervised learning). The results can either be binary or multi-class. There is a whole suite of methods working on classification task such as KBS, SVM, and neural networks. Many well-studied AI datasets, such as the MNIST and ImageNet, were established for training classification models. Since most paleontological studies work with the phylogeny or taxonomy of extinct organisms, classification has received the most attention among all paleontological

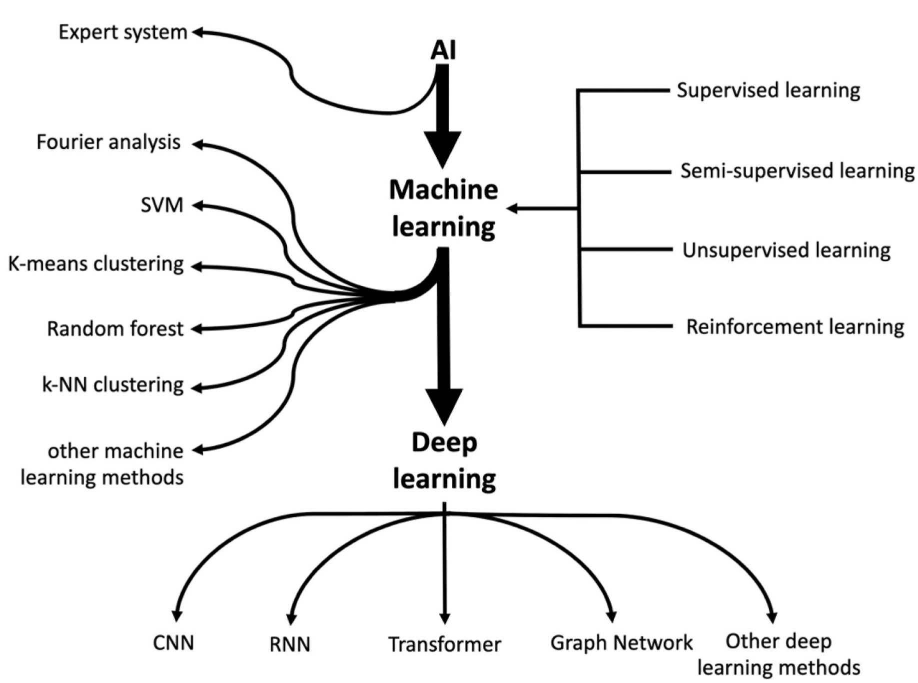

Fig. 1. A simplified structure of AI. Abbreviations: CNN, convolutional neural network; k-NN, k-nearest neighbors; RNN, recurrent neural network; SVM, support vector machine.

AI tasks.

Segmentation converts an image into partitions that represent the positions and shapes of certain objects. Traditional segmentation techniques used thresholding to distinguish pixels based on their color values, and more complicated methods such as region growing were developed to work with more complex scenarios. However, methods based on linear interpolation are not capable of segmenting even relatively simple images. Another major pathway towards automated segmentation is edge detection, which is usually based on human-designed operators, such as the first-order Canny operator (Canny, 1986) and the second order Laplacian of Gaussian (LoG) operator (Basu, 2002). Because most of these operators have barely been applied in paleontological studies, we do not provide a more comprehensive introduction to these algorithms here. To our knowledge the only current paleontological example using an edge detection algorithm is FossilMorph, a program for fossil image statistical analysis and classification developed by Zheng et al. (2022).

Prediction is also a major task in AI applications, namely, to estimate the probability or continuous value of a certain output according to the intrinsic laws learned from training datasets. A typical example of AI prediction in earth sciences is weather forecasting. By integrating historic patterns and data from surrounding regions, AI can make short-term prediction on future local weather with reasonable accuracy (Reichstein et al., 2019; Bi et al., 2023; Zhang et al., 2023). Although a substantial number of paleontological studies focus on prediction of certain aspects of extinct organisms (e.g., their behavior, soft tissue morphology, and ecology), AI has only been applied in a limited way to such studies (Anemone et al., 2011; Fan et al., 2021; Xu et al., 2022).

2. Paleontological AI study development

In this part, we review paleontological AI applications from the early 1980s to 2023. Studies are categorized by their tasks, including classification (micro- and macrofossils), segmentation, and prediction (Fig. 2A). Although the organisms covered here are from distantly related groups, the central idea and methods used in research are usually interchangeable. For each section, we make efforts to arrange reviewed studies chronologically to show the rise and fall of different AI methods and advancements in fossil data compilation. We further divide Section 2.1 on microfossil classification into three sub-sections because microfossils have dominated past paleontological AI studies and they represent the most complete history of AI applications in paleontology.

2.1. Microfossil classification

Microfossils have always been a subject of intensive paleontological study due to their large quantity (resulting from high preservation potential) and crucial roles in evolution research, sedimentary geology, paleoclimate, and many other fields. In contrast to macrofossils, microfossils generally have a size <1 mm that can be barely seen by the naked eye. Microfossils are diverse tiny organisms including but not limited to foraminifera, conodonts, algae such as coccoliths, plant pollens, and fragments from other organisms (e.g., ichthyoliths). The recognition, identification, and classification of microfossils is normally tedious, and thus many researchers have proposed methods to automate different aspects in a traditional handcrafted research workflow, such as sampling, imaging, measurement, identification, and classification. The most fundamental challenge is microfossil classification, including classification from similarly sized particles and into different species. Many early attempts applied Fourier analysis to extract patterns from outline shapes and other components of shape that are diagnostic for species identification (Healy-Williams, 1983, 1984) or to identify fossils (Belyea and Thunell, 1984; Burke et al., 1987; Garratt, 1992; Garratt and Swan, 1992). Thresholding based on image pixel gray values was also used in outline extraction (segmentation) from fossil images (Hills, 1988). However, these early methods can hardly work under realistic scenarios, as they often require extensive human effort in data preparation or can only be applied to a specific step in the workflow.

2.1.1. Knowledge based systems

Automated specimen identification and classification of microfossils using KBSs was an early and widespread application of AI (Brough and Alexander, 1986, Beightol and Conrad, 1988, Riedel, 1989, Swaby, 1990, 1992, Athersuch et al., 1994, Liu et al., 1994 and Yu et al., 1996). Sometimes the size of a KBS can be fairly large, as shown by Swaby (1990, 1992) who built visual identification expert systems (VIDE) for

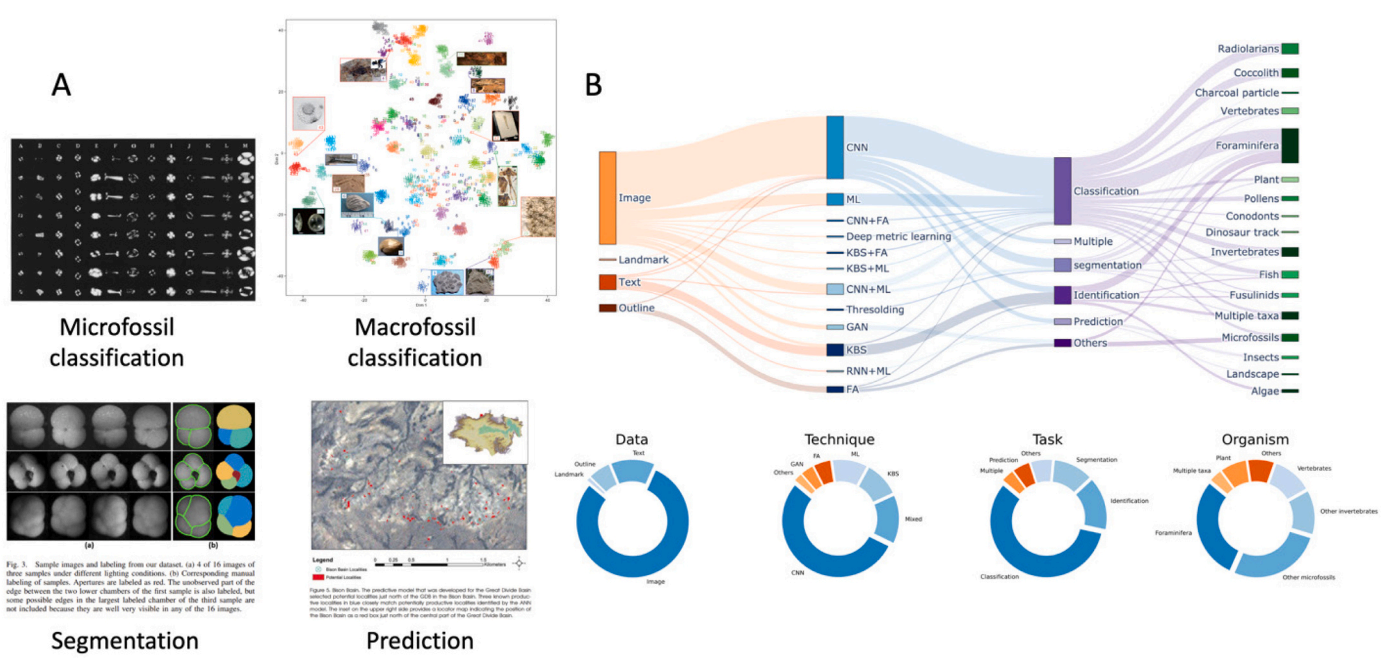

Fig. 2. A. Examples of paleontological AI tasks, images are modified from Dollfus and Beaufort (1999) Microfossil classification, Liu et al. (2023) Macrofossil classification, Ge et al. (2017) segmentation, and Anemone et al. (2011) prediction. B. proportions of taxa, methods, input data, and task in paleontological AI studies. Abbreviation: CNN, convolutional neural network; FA, Fourier analysis; GM, geometric morphometric; KBS, knowledge-based system; ML, machine learning.

fossil foraminifera and conodont identification, integrating 3500 images, 100 attribute-value tables, and over 10,000 lines of text information. Liu et al. (1994) constructed a dual-step identification system that embedded a KBS and an image analysis subsystem, which extracted graphic information from foraminifera chambers and suture images in addition to previous shape analysis based on Fourier analysis and edge detection. The system was built in CLASSIC, a knowledge-based system shell that provides the knowledge base scheme and inference engine. Most KBSs were built up according to a given taxonomic hierarchy and anatomical terminologies, resembling a traditional paleontological key, such as a dichotomous key, in using diagnostic characters for classification. However, the scope of those studies was usually incomplete for more general identification; for example, there were only 30 identifiable species of modern planktonic foraminifera (out of a total of ~50) with 25 discriminate rules in another CLASSIC based KBS prototype by Yu et al. (1996). Further, limited computing performance in the 1990s meant that KBSs often required considerable time in identification (e.g., 3–4 min/specimen by Yu et al., 1996).

Most KBS studies in paleontology were conducted from the late 1980s to mid-1990s, and a few studies continued to explore such application in the 2000s. For example, Kaya et al. (2013) built an expert system to classify pollen images of *Onopordum*, an extant plant in the family Asteraceae. The community has witnessed similar rise and fall of KBS in conventional AI studies (Davis, 1982; Bell, 1985). As a predesigned system, KBS is commonly restricted by prior knowledge from experts that can incorporate personal biases and, although robust in processing recorded inputs, KBS cannot process unfamiliar objects. It thus has limited intelligence (though its definition is ambiguous). The capability of a given system may be improved by incrementally adding "knowledge", but ultimately KBS has to be built on well-organized databases and a confidently agreed taxonomy, both of which are not practical for extinct and extant organisms. KBS cannot "learn" or "discover" hidden knowledge, thus its popularity has gradually declined as other probability-based methods (e.g., random forest) perform better in prediction and require less effort in design and maintenance. However, the modern knowledge-graph (Chen et al., 2020) may be seen as a successor to KBS by incorporating knowledge from a given field or even a very broad range (e.g., Wikipedia), although the reasoning part largely remains neglected.

2.1.2. (Convolutional) neural network

A first attempt to deploy CNN in image processing was conducted by Fukushima (1980), and its paleontological use can be traced back to *Système de Reconnaissance Automatique de Coccolithes* (SYRACO, or SYRACO1) by Dollfus and Beaufort (1996). This closely followed the application of neural networks to identification of extant marine phytoplankton; Boddy et al. (1994) were first to use back-propagation to update neuron status in the network, a commonly used training method. Coincidently, this was when one of the last paleontological KBS studies was conducted (Yu et al., 1996). SYRACO used two-dimensional fast Fourier transformation and a two-layered neural network to identify coccoliths; a detailed description of the system is given by Dollfus (1997). Its updated version SYRACO2 was introduced by Dollfus and Beaufort (1999), which used a deeper 5-layered neural network and back-propagation algorithm. On the basis of larger training datasets including 2000 images from 13 coccolith species and non-coccolith objects, SYRACO2 showed overwhelmingly better performance than its predecessor. SYRACO2 can make 40 classifications per second using a 200 MHz processor and has reached a mean recognition efficiency of 86%, while for SYRACO1 this level was 49%. Dollfus and Beaufort (1999) called SYRACO2 a fat neural network because of the significantly increased number of parameters (~800,000). Beaufort and Dollfus (2004) added parallel neural networks and dynamic view to improve the performance in object detection and classification accuracy. The SYRACO dynamic version added five motor modules for data argumentation by correcting translation, rotation, dilatation, contrast, and symmetry in fossil images. However, as the analysis was run on sedimentary samples that contain numerous coccolith-sized objects, false positives resulting from the inclusion of similar objects was a significant problem for SYRACO. The dynamic version was probably not an ideal update as CNN itself is space invariant, so four of the five modules were theoretically unnecessary. Nevertheless, the appearance and continuous iteration of SYRACO indicate that the deficiencies of KBS in automated paleontological workflow had been realized, while the application of CNN continued to unfold in both paleontological studies and other fields. The SYRACO system has been actively used in analyzing fossil coccoliths from marine sediments since its first deployment (Beaufort et al., 2001, 2022), confirming this as one of the most successful and long-lasting applications among paleontological AI studies.

Schiebel et al. (2003) and Bollmann et al. (2004) gave a brief review on automated sedimentary particle (including foraminifera) analysis although there were only a few automated micropaleontological studies at that time. They explained the need to build microscope systems to automatically acquire fossil images in the field and pointed out the differences between structural/statistical techniques and Artificial Neural Network (ANN) in automated image analysis. The first category requires given knowledge from experts (e.g., KBS), but the tailored features and inherent inflexibility make those techniques incapable of working with new taxa or messy sediments. At that time, ANN was more likely to be a promising direction according to the results from pollen (France et al., 2004) and coccoliths (Dollfus and Beaufort, 1996, 1999; Beaufort and Dollfus, 2004). Schiebel et al. (2003) and Bollmann et al. (2004) developed the Computer Guided Nannofossil Identification System (COGNIS) for the segmentation, preprocessing, and classification of microfossil images. The training dataset included 979 images covering 14 Holocene coccolith species at a resolution of 48 × 48 pixels and a validation dataset (named classification dataset in original study) of 715 images from the same 14 species. Their 5-layer CNN reached a recognition rate of 75% on average, but the error rates varied significantly across species from 3% to 88%. The low resolution in this study seems to have been a compromise between computational cost and sampling, and the 48 × 48 pixels resolution came from down-sampling of the original image. COGNIS-light was developed at the same time to classify only a particular foraminifera species, *Florisphaera profundal*, from all other biotic or abiotic particles. The training dataset had 1000 images of *F. profunda* and 1000 images of other particles, but such a training strategy resulted in strong false positives, likely due to the strongly biased samples.

A key methodological innovation in CNN-based microfossil classification is transfer learning which uses a pre-trained model or part of it on new tasks. Depending on the relevance between the source and target domains, it saves training cost and takes advantage of relevant information from pre-training datasets. Zhong et al. (2017) summarized previous work on foraminifera classification and used pre-trained classic networks (ResNet-50, VGG-16, and Inception V3) in paleontological AI studies for the first time, indicating the versatility of neural networks and large datasets. Keçeli et al. (2018) used de novo trained and pre-trained CNNs to classify radiolarians. Interestingly, the pre-trained VGG-16 models performed much better, but it is uncertain whether this was caused by neural network depth or pre-training. The *Endless Forams* dataset was created by Hsiang et al. (2019), including 34,640 planktonic foraminifera images covering 35 extant species. They showed that transfer learning based on classical datasets such as the ImageNet (Deng et al., 2009, including 14,197,122 images indexed by 21,841 synsets) performed well on the identification of foraminifera, achieving an accuracy of 87.4% using VGG-16. As Zhong et al. (2017) and Mitra et al. (2019) suggested, many of these AI classification models only performed a role as "proofs of concept" that CNNs can be used in paleontological studies, but are far from demonstrations of practical usefulness.

Liu and Song (2020) collected a large thin section image dataset of 30,815 images from 18 fossil groups (mostly invertebrates) and four minerals or sedimentary structures. With pre-trained models on ImageNet, their classification results showed high accuracies of >90%, and the more balanced dataset prevented overfitting in biotic structures. Following this, Liu et al. (2023) presented the Fossil Image Dataset (FID) with 415,339 images from 50 fossil clades (including various invertebrates, vertebrates, plants, microfossils, and trace fossils), and an online model (www.ai-fossil.com) is also available for fossil image identification. They showed that certain clades were more difficult to identify than others and proportions of complete to fragmentary fossil hugely influence the rate of correct identification. Models pre-trained on ImageNet performed well regardless of the huge differences between ImageNet and fossil datasets. Moreover, activation mapping on fossil images may provide previously overlooked information about taxonomy and character evolution.

In summary, from the late 1990s to early 2000s, CNN created significant progress in paleontological AI studies and resulted in two successful applications, SYRACO and COGNIS. CNN based models require much less effort in system construction and also run much faster than KBS but do demand significant front-loaded time input to generate high-quality training datasets. Pre-trained models can partly relieve the burden from limited training fossil data.

2.1.3. Other machine learning methods

While CNN-based models began to thrive in paleontological AI studies, other machine learning methods still played essential roles, especially for small scale data. Many studies combined traditional machine learning methods and neural network in fossil related tasks. Marmo and Amodio (2006) and Marmo et al. (2006) used k-Nearest Neighbor (k−NN) and a three-layered perceptron classifier to automatically classify chamber arrangements in foraminifera. However, both the training (207 and 200 images) and testing datasets (70 and 80 images) comprised too few images and only five classes, and the extremely high accuracy (97.1%) from the perceptron classifier is likely a result of over-fitting. Wong (2011) developed Microfossil Quest, an interactive system for microfossil search, identification, and education. This system integrated KBS and various machine learning methods including k-NN, K-means clustering, ANN, etc. Almost 4000 specimens were included in the database and many of the identifications were collected using crowdsourcing; however, the testing experiment only used 238 specimens and achieved accuracy at the genus and species level of 81% and 47%, respectively. The concept of crowdsourcing is now widely used for general AI dataset annotation (e.g., drawing a bounding box of a dog from the background), but paleontological studies often require more specific knowledge that the public is unlikely to have. Keçeli et al. (2017) manually segmented images of radiolarians and used AlexNet (Krizhevsky et al., 2012) to extract patterns from these manually processed features (e.g., eccentricity and circularity). They then used SVM, k-NN, Adaboost, and Random Forest to classify radiolarian species. The results showed that SVM generally performed the best among all classification methods. However, such a complicated workflow counteracts the idea of automation by AI. Since SVM had been applied to radiolarian classification previously by Apostol et al. (2016) to achieve equivalent performance and CNNs without manually processed features have been applied to radiolarians with high accuracies (Carlsson et al., 2023), it may be unnecessary to apply CNN to extract data from manually processed features for subsequent classification.

Karaderi et al. (2021) applied deep metric learning for the first time to the *Endless Forams* dataset. Deep metric learning involved learning the distance function between objects in high-dimensional metric space, with distance indicating the similarity in category or characteristic. By comparing with other published benchmarks on the same dataset, Karaderi et al. (2021) found that deep metric learning exceeded other methods in identifying planktic foraminifera species, reaching an accuracy of 92%. However, the validity of metric learning and its natural deficiencies should be more carefully evaluated in comparison to the performances of different learning strategies (Musgrave et al., 2020).

Pollen and other palynomorphs have commonly been explored using AI methods as specimens are abundant, similar to marine microfossil studies. However, most palynological AI studies have focused on extant taxa, as reviewed by Treloar et al. (2004), Li et al. (2004), Zhang et al. (2004), Holt et al. (2011), and Daood (2018). KBS, traditional machine learning methods, CNN, and transfer learning (de Geus et al., 2019) have been applied in pollen image localization, recognition, identification, and classification. There are only a few AI studies on fossil pollen taxa and coverage has been restricted. Kong et al. (2016) proposed an unsupervised learning method to select representative feature patches from pollen images, then used these image patches as the dictionary as the basis of a sparse coding model. They performed SVM to classify pollen images into three selected species and reached an accuracy of 86.13%. Bourel et al. (2020) introduced CNN and decision trees in pollen recognition and claimed to be able to identify both fossil and modern pollens to the genus level, sometimes even species level. This study focused on three families (Amaranthaceae, Poaceae, and Cyperaceae) including 1698 pollen grains, in which 223 are attributed to the damage dataset and 97 are fossils. Although the integration of multi-CNNs and comparably large training datasets allowed successful identification of damaged pollen grains, the sampling strategy from only three families, a fairly short geological timespan, and geographically restricted localities raised doubts about the generalizability of such a study.

It is commonly acknowledged that CNN has outperformed many traditional machine learning methods in imaging related tasks since AlexNet (Krizhevsky et al., 2012), but this view may not hold when training data is limited or the tasks can be clearly delineated. Xu et al. (2020) applied combinational machine learning including scale-invariant feature transform (SIFT), K-means, and SVM to automatically recognize microfossils from images, and compared the performances between these traditional machine learning methods and classical CNN models. Traditional machine learning methods were overwhelmingly better than CNNs in this case because of the existence of abiotic rock images. Because CNNs are usually very large models with millions to billions of parameters that need to be tuned during training, small datasets cannot fulfill the request to optimize all parameters.

2.2. Macrofossil classification

Macrofossils are large enough for direct observation by the naked eye and are thus much fewer in number compared to microfossils. Their morphology is also more likely to be affected during preservation. Incompleteness and deformation of macrofossil specimens make it more challenging to construct a proper training dataset for automated classification. The classification of macrofossils often requires identification of diagnostic features, and they need larger data volumes to accommodate their morphological features. For example, the size of a dinosaur skeleton 3D scan is much larger than that of foraminifera 2D image. All these challenges have resulted in only a few examples of macrofossil classification studies, but there have been attempts.

Based on a dataset from Huang et al. (2023), including 16 fusulinid (large Paleozoic foraminifera) genera with 150 images of each, Hou et al. (2023) proposed a triple-base model using differently augmented images in the original, gray, and skeleton view (OGS) to improve identification performance. The OGS triple-base strategy showed generally better performance in almost all CNN architectures tested than using only one or two image types, and activation mapping indicates different hot-zones in different image types, indicating the complexity of characteristic features under different scenarios.

Insects occupy a large portion of biodiversity in both the modern and ancient biosphere, and the huge numbers of species and their widespread impacts in terrestrial ecosystems, as well the shortage of trained entomologists, have led to strong calls to identify and classify insects. Most entomological AI studies focused on extant species classification, of which Martineau et al. (2017) reviewed 44 studies for automated image acquisition, feature extraction, classification, and datasets. De Cesaro Júnior and Rieder (2020) surveyed automated identification of extant insects with a primary focus on machine learning methods. Among the 33 studies examined, 63% used CNNs based methods, while 29% used handcrafted features in the same tasks. Few studies introduced more complicated methods than CNNs, such as the attention mechanism. Contrary to foraminifera AI studies, large scale datasets have been better presented in extant insect studies. The largest dataset was proposed by Liu et al. (2019), the *PestNET* (https://www.pestnet.org/), which incorporated >80,000 images with over 580,000 pests classified into 16 classes (not necessarily species). *PestNET* not only utilized traditional CNNs for classification, but also introduced the attention mechanism in feature extraction and reached mean average precision of 75.46% in multi-class detection. There have been several other surveys and perspectives about AI applications in entomology (Valan et al., 2019; Høye et al., 2021; Kasinathan et al., 2021; Amarathunga et al., 2021). Since most insect-related machine learning studies worked with extant species, current AI-based entomological studies are probably not limited by the sparsity of data, but by the lack of well annotated datasets, incongruence in taxonomy, and currently overlooked links with molecular methods such as DNA barcoding. There are several insect-based AI studies working on prediction the ecological role of fossil insects, which we will discuss in Section 2.4.

Lallensack et al. (2022) discriminated ornithischian and theropod footprints using VGG-16 and a dataset with >1000 footprint outline silhouettes. The models performed better than human experts on the testing dataset, but sampling bias and information loss from 3D to 2D was a major problem. Wills et al. (2023) compiled a theropod dinosaur tooth dataset including 1702 teeth and applied several machine learning methods to infer the morphotype of isolated teeth from Middle Jurassic localities. The results indicated that even isolated teeth bear enough information for classification to the level of family (e.g., Therizinosauroidea and Troodontidae).

da Conceição et al. (2023) developed PaleoWood, a machine learning based classifier for Paleozoic gymnosperm woods. They sampled 62 genera of Paleozoic gymnosperm woods and 16 morphological characters for training and validation, and their models were based on logistic regression (LR), linear discriminant analysis (LDA), and k-NN. Convergence played a crucial role in influencing the classification results as the model cannot distinguish it from homology. Such a classification system is more like a re-run of KBS than more recent models.

The scale of macrofossil classification studies is limited comparing to microfossil classification, and the studied taxa are restricted, too. But the existence of large-scale extant organism datasets and scarcity of trained expert do indicate the possibility of more automated workflow aided by AI.

2.3. Segmentation

The applications of various imaging techniques in paleontology have resulted in rapid growth in the available data, covering both micro and macrofossils, and their modalities include optical micrographs, electron micrographs, tomographic images, and many others. Traditional paleontological imaging segmentation is largely manually processed due to inconsistencies in mineral composition rendering automatic segmentation challenging, thus requiring tremendous efforts especially with high-resolution imaging techniques. While AI-based models have been widely applied in medical tasks covering almost all kinds of data modalities and tissues (Litjens et al., 2017; Shen et al., 2017), segmentation models were only developed recently for both 2D and 3D fossil images including foraminifera (Ge et al., 2017), alvarezsaurian dinosaur skeletal histological thin section (Qin et al., 2022a), and protoceratopsian dinosaur CT scan (Yu et al., 2022). These segmentation models are all based on CNN.

Hou et al. (2020, 2021) published ADMorph (Archives of Digital Morphology, http://www.admorph.ivpp.ac.cn), an open dataset of vertebrate fossils for deep learning studies, and tested the segmentation performance of various classical DNNs such as U-net, PointNet, and VoxNet on that dataset. These two studies show that paleontological data can be prepared and processed in the workflow resembling other kinds of images, and that deep learning can significantly save processing time in imagery tasks such as segmentation.

2.4. Prediction

While AI applications to fossil insect classification are limited (Section 2.2), there have been few paleo-ecological studies using AI to predict ancient mimicry behaviors. Fan et al. (2021) studied plant mimesis in both extant and extinct insects, which showed similar results consistent with the deep origin of biological mimesis. Based on the same neural network, Xu et al. (2022) further analyzed mimicry and insect camouflage from mid-Cretaceous Kachin ambers. In these two studies, the Siamese Network was firstly pre-trained on the Totally-Looks-Like dataset (TLL, Rosenfeld et al., 2018), which comprises 6016 image pairs that look similar to the naked eye but may come from totally unrelated objects. The model was then fine-tuned using specifically constructed mimic insect-plant pair datasets. Dissimilarities are measured between plant-insect pairs to quantify mimesis behavior. However, both studies were confined by fine-tuning of the datasets as they could not carry out exhaustive sampling over all plant-insect mimesis behaviors. Nicholson et al. (2015) suggested that there had been a fast expansion in our knowledge of extinct insect diversity and taxonomy since 1994. Established fossil insect databases, although not as comprehensive as extant insect databases, are ample resources for data-driven entomological studies, particularly AI-based studies.

Anemone et al. (2011) made attempts to use AI for fossil exploration prediction. A four-layered ANN was trained to explore the connection between landscapes and fossil preservation, and the results from Bison Basin, Wyoming, USA seemed to be encouraging. Kopperud et al. (2019) combined long short-term memory (LSTM) recurrent neural network and manually labeled text dataset to predict the historical occurrence of Bryozoa, a group of marine invertebrates. Martín-Perea et al. (2020) demonstrated AI-based identification of fossiliferous levels in both archaeological and paleontological sites, with testing results from two Late Miocene paleontological sites in Madrid, Spain. Multiple methods including SVM and RF were used to quantify the spatial distribution of fossils and these provided information about the faunal assemblage and directions for excavation.

We cannot discuss every paleontological AI study in detail. Fig. 3 illustrates major progress in the development of paleontological and mainstream AI. An introductory list is presented in supplementary material 1 with publication dates, study organisms, methods, and tasks. Fig. 2 illustrates the number of reviewed studied regarding their studied organism, methods, tasks, and input data modality.

3. Results

Here we reviewed paleontological AI studies primarily by highlighting their tasks (classification, segmentation, and prediction). While the earliest paleontological AI study can be traced back to the early 1980s, and relevant discussions appeared even earlier (Sneath, 1979), there was <1 study per year on average before 2000 (Fig. 3). The number of studies and data scale only significantly increased recently (Fig. 4). There have been remarkable changes in the methods used, while the studied organisms, tasks, and data modalities have remained stable through the last four decades.

Among studies surveyed here, roughly 1/3 focused on foraminifera, 1/3 on other microfossils, and the rest on other organisms including plants, insects, dinosaurs, etc. (Fig. 2B). A possible explanation for such preference is data availability. Microfossils (including foraminifera) can often be completely preserved in marine sediments and suffer less

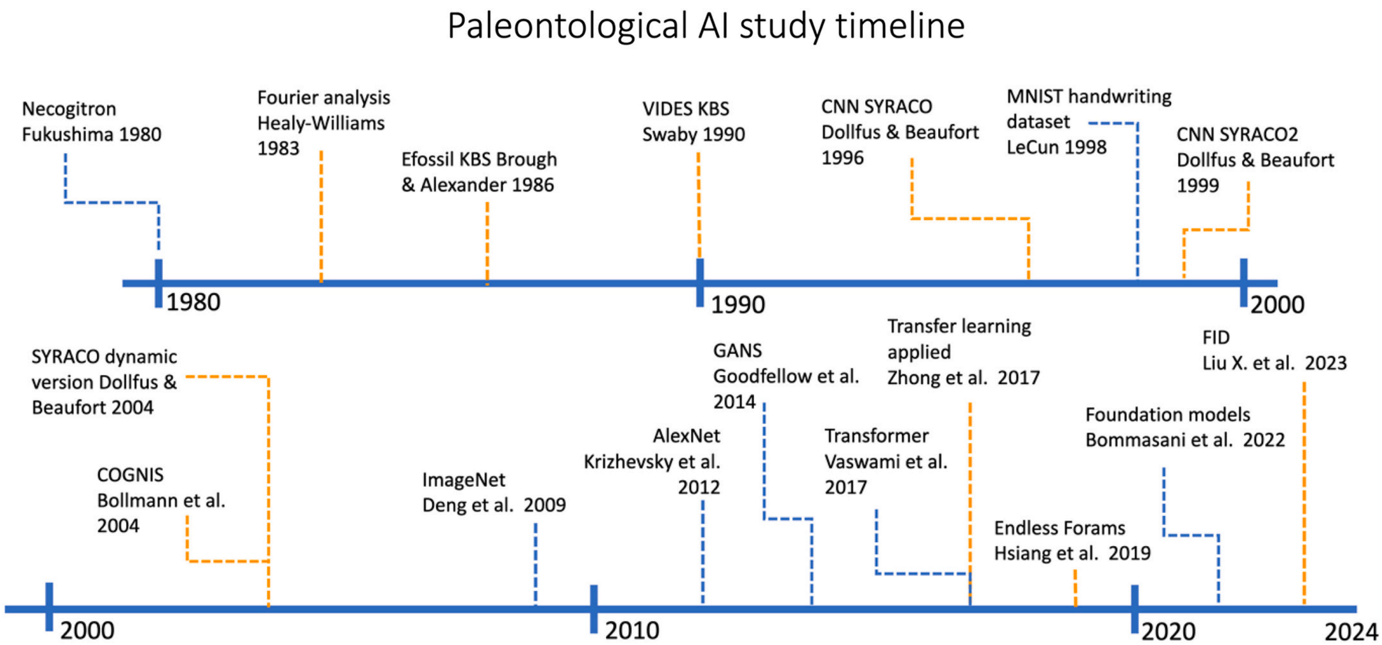

Fig. 3. Major progress in paleontological AI study development. Orange dash line shows paleontological AI study progress and blue dash line shows AI progress in computer sciences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

deformation compared to macrofossils. Further, many microfossils have been key indicators for paleoclimate, paleogeography, stratigraphy, and many other research fields. The need to process large quantities of specimens has quickly driven the development and deployment of microfossil AI applications. Most practical AI applications (e.g., SYRACO and COGNIS) and open datasets (e.g., Endless Forams and FID) have focused on microfossils to date. Studies on macrofossils are notably fewer, with most working with CT scans of individual structures/elements and histological thin sections, which are not strongly relevant to gross fossil morphology or taxonomic classification.

While microfossils have been the most popular organisms in paleontological AI studies since the 1980s, studies on other small-sized fossils like insects and plant pollen are rare despite their large quantities of available fossils. Various reviews have summarized AI applications on extant insects and plants (Treloar et al., 2004; Li et al., 2004; Zhang et al., 2004; Holt et al., 2011; Martineau et al., 2017; Daood, 2018, and De Cesaro Júnior and Rieder, 2020), showing that most studies worked with extant species probably due to better annotated datasets and request from agriculture and industry. There is also interoperability between AI identification and classification for extant and extinct taxa. Rani et al. (2022) reviewed advances in automatic (extant) microorganism recognition, among 100 studies from 1995 to 2021 and, although only a single paleontological study (Mitra et al., 2019) was mentioned, they showed similar trends in the development of methods and data scales.

3.1. Methods

During the 1980s to mid-1990s, the first generation of fossil AI was based on KBS, resembling the evolution of AI models at the same time or slightly earlier. But KBS, normally as a handcrafted rule-based system, has internal deficiencies in both system building and maintenance, and was rapidly replaced by CNN in general AI applications and more specific fossil studies. CNN has proven to be effective in imaging and a variety of other tasks. Although the first attempt building CNN-based AI model was by Fukushima (1980), its first application in paleontological studies occurred much later (Dollfus and Beaufort, 1996).

CNN has been the most common model in current paleontological AI studies. It is the sole model used in more than half of the studies surveyed here, even without counting those incorporating multiple models (Fig. 2B). With the maturation of open-source machine learning frameworks such as TensorFlow and PyTorch, building neural network models is becoming easier. Rapid advancements in graphics processing units (GPUs) and availability of cloud computing services like Google Colab allow almost anyone to access unprecedented computational power. There are also plenty of ready-made models for training or fine-tuning. Most de novo designed CNN models in paleontological AI studies were shallow, seldom exceeding five layers. Meanwhile, AlexNet (Krizhevsky et al., 2012) had five convolutional layers and three fully connected layers, and Deep Residual Network (ResNet) can reach hundreds to thousands of layers in depth (He et al., 2016). What we can conclude is that there is a probably a 10-year gap in model design between proposed paleontological AI models and classic works. Part of this is that the amount of data needed to train very deep networks is not available. Hsiang et al. (2019) tested VGG-16, Inception V3, and DenseNet-121 on the Endless Forams datasets, which significantly outnumbered most paleontological datasets surveyed here, but the shallowest VGG-16 had the best performance. Many researchers have admitted that the paleontological AI models they developed were only "proof-of-concept", which refers to both models and data coverage. At the moment, we have witnessed success from much more complicated models on complicated tasks, and there is now hardly any need to validate the feasibility of CNN or many other models in paleontological AI studies.

Transfer learning has also been widely used to overcome the computational cost at training stage. Many studies chose to use models pre-trained on large image datasets such as ImageNet (Deng et al., 2009; Zhong et al., 2017; Keçeli et al., 2018; Mitra et al., 2019; Hsiang et al., 2019; Marchant et al., 2020). Classic networks including VGG-16, ResNet50, and U-net are often preferred as the models or the backbones of larger models. There have been very few specifically designed models for paleontological AI studies (e.g., Marchant et al., 2020 created a custom CNN, Base-Cyclic, that adapts to image size; (Qin et al., 2022a) designed a dual-resolution network to distinguish primary and secondary osteons in dinosaur histological thin sections). Although customized models can performed better than off-the-shelf models, the additional costs of model design, training, and fine-tuning can not be ignored.

While CNN is the most favorable model in paleontological AI studies,

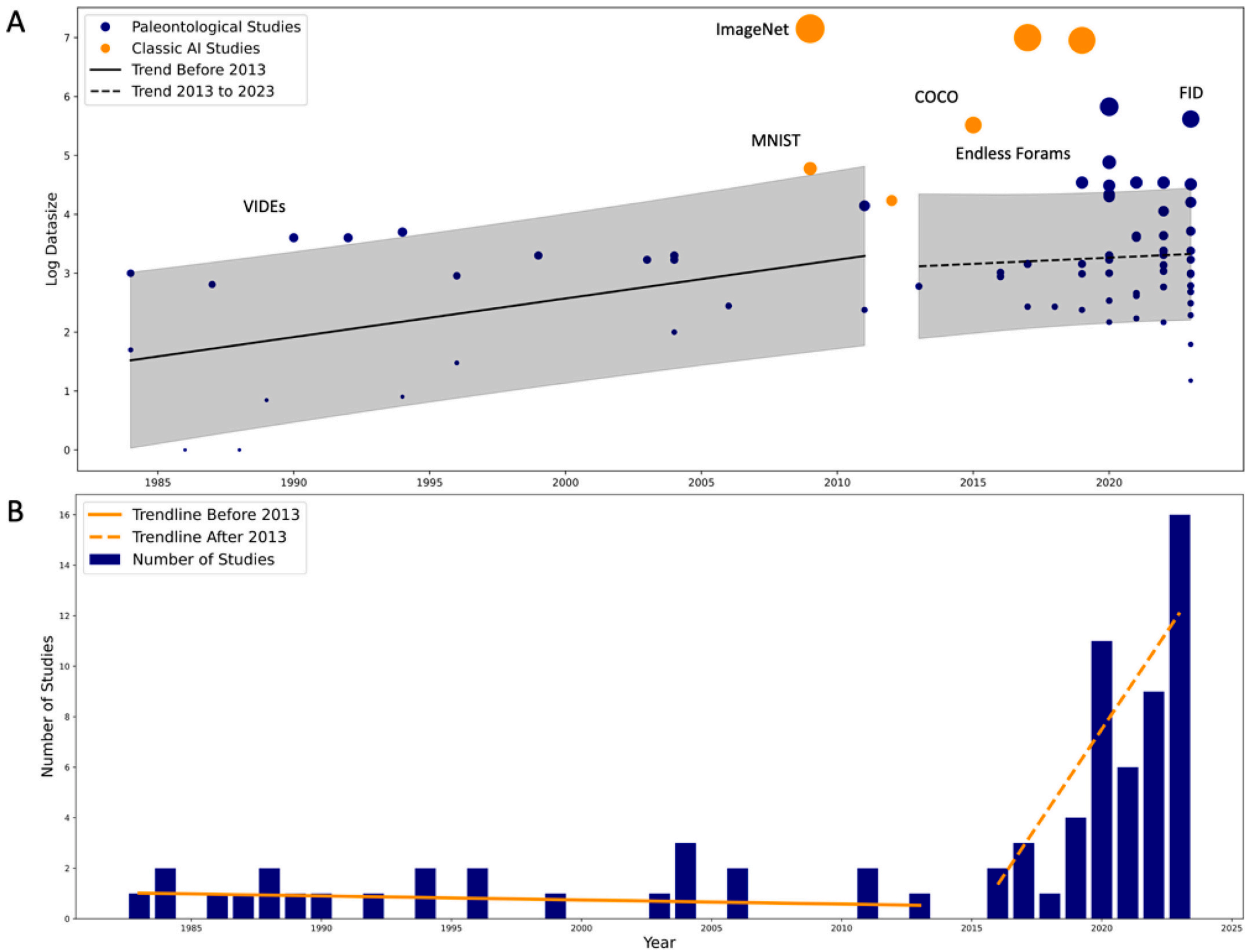

Fig. 4. A. Data size (log10) of paleontological and commonly used AI datasets from 1983 to present, with trendlines showing the data size pattern before and after 2013, gray areas indicate 75% confidence interval. B. number of studies of paleontological studies from 1983 to present, with trendlines showing the number of paleontological AI studies before and after 2013.

traditional machine learning methods including SVM, k-NN, k-means, RF are still actively employed. Xu et al. (2020) criticized the overuse of neural networks in fossil image recognition. In spite of their advantages, CNN-based models usually need to be trained on large and balanced datasets, and the results are often fragile in interpretability. Paleontological data are intrinsically sparse and fragmentary; as a result, CNN may be outperformed by other machine learning or even more concise methods under many scenarios.

3.2. Tasks and data modalities

About two-thirds of paleontological AI tasks are multi-class classification, as taxonomic identification plays a fundamental role in paleontology (Fig. 2B). The established taxonomy offers a solid basis for training supervised learning models, thus in principle we can construct a model that can accommodate all taxa of interest. However, the problem is that the amount of training data that is necessary to train such a model is largely unfeasible to collect. Other typical tasks include (image) segmentation and prediction; while they are both classic AI applications, the number surveyed here is significant fewer than classification task. Other common tasks in mainstream AI studies, such as text processing, image inpainting, and feature engineering, remain untouched in the paleontology field.

Around 80% of studies surveyed here worked on image data, and several had outlines or silhouettes that are directly derived from images (Fig. 2B). This is not surprising since photographic images are the most direct and inexpensive method to record fossil morphology, especially for microfossils. However, the morphology of macrofossils cannot be easily captured by such low-resolution 2D images. A detailed 3D image that is enough for morphology description, for example a high-resolution CT scan, is too large for current AI model training. There are also too few of them available in general to adequately train models. In practice, high-resolution macrofossil images are often used for illustration, measurement, morphology description reference, and geometric morphometric data collection, but not directly in analysis. Large 3D and high-resolution 2D images need to be compressed for subsequent studies. There have been no been paleontological AI studies working with those compressed data modalities from original images, and the compression process itself – i.e., reducing size without losing information – may be a target for AI learning.

3.3. Data scale

Large amounts of data are necessary for AI model training. Datasets comprising up to $\sim 10^3$ images were constructed in the 1980s (Swaby, 1990, Fig. 4). Later CNN-based models were often trained on hundreds to thousands of images. The construction of large-scale open datasets, Endless Forams for example, was a major advance. Several fossil image datasets have reached ~10^4 to 10^5 images, but many recent studies continue to use small datasets with fewer than ~10^3 images (Fig. 4). Several fossil AI dataset have equivalent or even larger scale than classical deep learning datasets such as the MINST handwritten digit dataset (http://yann.lecun.com/exdb/mnist/, 60,000 training and 10,000 testing images in 28 × 28 pixels) created in 1998, but still are two to three orders smaller than other commonly used baseline datasets (for example, ImageNet comprises 14 million images across >20,000 categories). There is a roughly 20-year lag (from MNIST 1998 to Endless Forams 2019) between mainstream and paleontological AI studies in the term of data scaling. The recent boom in the number of paleontological AI studies is more likely the result of a continuously lowering bar in training and deploying AI models, at least from the perspective of data scale.

Most datasets surveyed here were manually collected and annotated, but there are also exceptions that use web crawling and published literature (Liu et al., 2023) for collection and crowdsourcing for annotation (Wong, 2011), which are both common practice in preparing large-scale AI training datasets. However, paleontology is a specialist field that needs expert knowledge and long-term practice in order to generate training sets, making automated and crowdsourced dataset creation a more difficult prospect. There have been no unsupervised learning examples in relevant studies, and relying on automatically crawled data or crowdsourcing can result in potentially misleading bias during model training.

4. Perspectives

In this section, we discuss the various aspects of paleontological studies that might benefit from established AI achievements in the near future, and how recent progress including Transformer, one/few-shot learning, auto-content generation, and Large Language Models will interact with paleontology.

4.1. Automated workflow

Automated species classification has always been a focus in paleontological AI studies, and such an idea was repeatedly proposed by Gaston and O'Neill (2004), MacLeod et al. (2010) and many others. Now there is prototype system developed based on large datasets and deep learning have been developed techniques (Liu et al., 2023). From KBS to CNN, most (if not all) paleontological AI studies surveyed here are within the range of supervised learning, meaning that there needs to be a solid labeled dataset for model training, with clear taxonomy for fossil classification. Taxonomy and phylogeny of extinct organisms is based on their morphology, more specifically synapomorphies in the context of cladistics. However, the definition of most morphological characters is fundamentally qualitative and subjective, as is the interpretation of homologies – not at the analysis level, but at the character definition and coding level. There is good reasoning and evidence for many characters, but that does not change the fact that it is a qualitative/subjective determination by human, and different researchers can and do disagree about different hypotheses of characters/homologies. Linear classifiers on top of hand-engineered features are only able to make simple partitions of the output space, indicating more complex models are obligatory for automated species classification.

De Garidel-Thoron et al. (2020) presented microfossil sorter (MiSo) system that can automatically pick microfossils from other coarse sedimentary fractions and process up to ~8000 samples/day. Richmond et al. (2022) developed a system for foraminifera manipulation, sorting, imaging, and classification called Forabot. Their training set used the Endless Forams dataset (Hsiang et al., 2019) and resulted in a system that can process approximately 27 specimen per hour. While these systems remain prototypes with limited practical applications, the combination of hardware and software opens the possibility of fully automated data collection and analysis.

To reduce bias, artificial (morphological) characters/features may become a vital component of systematic paleontology in the near future. Eventually, semi-supervised or even unsupervised learning could facilitate the discovery of synapomorphies and diagnostic features. The application of transfer learning in paleontology since 2017 has indicated the feasibility of automated character construction as models pre-trained on common objects showed appreciable performances on fossil datasets. Activation maps, which show the regions of attention that the machine uses to make its determinations, have been applied to fossil image classification (Hou et al., 2023; Liu et al., 2023) and suggest that machines can discover previously hidden patterns in taxonomy.

Even though automated classification may seem unrealistic at this moment for paleontologists, AI can certainly help to automate current paleontological workflows. The implementation of AI to quantitative phenotypic data has shown exciting potential for high-throughput phenotypic research (He et al., 2024). Studies have used machine learning algorithms for classifying existing shape data into clusters (e.g., Soda et al., 2017; Courtenay et al., 2019; Arriaza et al., 2023). While classification of data is a valuable application of AI, the greatest gains will likely come from using AI to automate the collection of quantitative phenotypic data. Collection of morphometric measurements, especially landmark-based geometric morphometric data is largely done manually and often a rate limiting step in current morphological studies. (Semi-) automated landmarking tools exist, including TINA Geometric Morphometrics Tool (Bromiley et al., 2014), auto3DGM (Boyer et al., 2015), ALPACA (Porto et al., 2021) implemented in the open software SlicerMorph (Rolfe et al., 2021), and others (e.g., Aneja et al., 2015). Alternatively, users can employ landmark-free methods that permit visualization and analysis of shape variation based on 3D mesh models, such as Generalized Procrustes Surface Analysis (Pomidor et al., 2016), spherical harmonics (Shen et al., 2009; Dalmasso et al., 2022), non-rigid surface registration (Snyders et al., 2014; Claes et al., 2018), and Morphological Variation Quantifier (morphVQ; Thomas et al., 2023). Although these techniques empower users to collect rich phenotypic data rapidly, AI has the potential to further automate and enhance these efforts.

To date, machine learning applications for morphological data have largely proceeded on clinical data and human faces. For rapid phenotyping and analysis of diverse biological systems, both supervised and unsupervised machine learning approaches could become indispensable. Currently, ML-morph (Porto and Voje, 2020) and tools implemented in the open-source Cytomine software (Vandaele et al., 2018) provide a supervised approach for automatically placing landmarks on 2-D images. These studies examined the performance of these methods on sets of different biological structures, ranging from *Drosophila* wings to entire bodies of fish, exhibiting patterns in shape variations that largely mirror that of manually landmarked datasets. Because these are supervised machine learning approaches, training datasets are required that consist of corresponding image and landmark data for dozens, if not hundreds, of specimens. Wöber et al. (2022) used CNN on 2-D images of fishes to allow major anatomical variation to be identified without a priori decisions on morphological features of interest. These authors report that the morphological features picked up by CNN are similar to the manually collected landmark data in distinguishing between fish populations and that it was able to identify distinct groups that match population clusters based on genetic rather than landmark data (Wöber et al., 2022). Existing machine learning tools show much promise for automated landmarking with high accuracy and precision under supervised algorithm, whereas unsupervised algorithms could lead to new discoveries about phenotypic differences and transformations. Extensions of these methods to 3-D image data (e.g., μCT data) will be paramount for our efforts to characterize phenotypes of taxa across the tree of life.

Despite enormous potential in the machine learning approach for

GM, ongoing utilization and future advances in these techniques involve important considerations. First, image acquisition standardization will be critical for both supervised and unsupervised methods because the accuracy of phenotypic characterization will be strongly influenced by image properties and quality. One way to reduce these effects is to expand the training dataset to include images in variable settings and orientations, although this strategy come with obvious costs. Extending the implementation of machine learning tools to 3D surface data may allow consideration of differences from orientation. Secondly, as with any landmark data, consideration of the landmark scheme that suits the biological question of interest is crucial. What one gains in speed and automation may be lost in interpretability. Thirdly, investigators should ensure replicability by reporting on any parameters used for the AI procedure and making any training datasets openly available. Next, these AI tools should allow for landmark positions to be checked and adjusted manually, ideally within the same program (e.g., Bromiley et al., 2014; Vandaele et al., 2018). This allows an investigator to assess the quality of the landmark placement and revise as needed. Finally, advancements in machine learning techniques, as well as non-AI automated landmark approaches, that allow for accurate phenotyping of broad, comparative sampling or developmental series should be made. Thus far, these methods have been applied to fairly restricted datasets, mostly within a genus.

4.2. Datasets

Data availability is another significant obstacle for future paleontological AI studies. Approaches for characterizing morphology often discard a substantial amount of data during the process. A typical CT-scan of a dinosaur skull may result in a 3D model of several gigabytes, but after encoding in a morphological character matrix, the overall morphology turns into tens to hundreds of bytes, a compression of approximately 10^6. Such data reduction likely results in significant loss of meaningful and useful morphological information. Moreover, as noted above, the coding process is often subjective, being based on an understanding of anatomy and evolution. Several studies (e.g., Allmon et al., 2018) suggested that paleontology is going to embrace Big Data. However, Big Data should usually meet the requirements of the four Vs: volume, variety, velocity, and veracity. Paleontological data have deficiencies in at least the volume and collecting velocity terms. Among recent data-driven studies, only a few have worked with data in a non-handcrafted manner (e.g., Fan et al., 2020), while most of them, are still within the frame of traditional handcrafted workflows (though with larger datasets than studies before, e.g., character matrices containing thousands of characters). Although transfer learning has been applied in paleontological AI studies since Ge et al. (2017) and data augmentation has begun to be applied by (Hou et al., 2023; Ferreira-Chacua and Koeshidayatullah, 2023), there lacks systematic evaluation and solution for the influence from data sparseness and imbalance, which may benefit from similar studies in earth sciences (Koeshidayatullah, 2022)(Dawson et al., 2023).

Current paleontological AI models have only been developed for a limited range of input data modalities. However, paleontological studies generally involve a wide range of data modalities and associated metadata including fossil morphology, geological age, paleo-environments of localities, isotopic composition, small or even macro molecular remnants, and many others. Fossil data collection is not limited to standard images and text, but rather encompasses images from various advanced imaging techniques, the occurrence information of fossil localities, morphological measurements, morphological character matrices, phylogenetic trees, and contents from published literatures, most of which remained untouched by AI models but will surely be subjected to a data-driven AI studies in the near future. In the next decade or two, we may expect that many recently proposed methods in AI will be adopted to paleontology and there will be larger, better labeled, better organized, better aligned, and consistently maintained datasets for both model training and further studies.

4.3. Models

Currently only basic CNNs and several primitive variants have been applied to fossil data. Although a simple CNN architecture can largely outperform many other methods, the increase in neural network depth can rapidly approximate extremely complicated functions that are challenging to describe using handcrafted features. Deep learning technology is still developing rapidly. The progress mainly lies on three aspects: base model, task head, and learning paradigm.

1) The base model acts as the backbone of a deep neural network, which determines the feature representation ability. The early AlexNet has a simple structure, i.e., only five convolutional layers (Krizhevsky et al., 2012). ResNet expands the network depth up to 200 layers by introducing residual unit (He et al., 2016) that significantly improves the feature extraction ability and training convergence. Beyond convolutional neural networks, Vaswani et al. (2017) introduced Transformers, which use a self-attention mechanism that weights different parts of inputs based on their significance. The novel attention-mechanism-based Vision Transformer (ViT) and multi-layer perceptron based MLP-Mixer, which achieve promising performance when the pre-training dataset is large enough, have also been proposed (Dosovitskiy et al., 2020; Tolstikhin et al., 2021). And now has been applied in micropaleontological studies (Ferreira-Chacua and Koeshidayatullah, 2023). Specific base models are often required in specific tasks. For example, the stacked Hourglass Network was proposed specifically for keypoint detection (Newell et al., 2016), and dynamic graph CNN was proposed specifically for point cloud encoding (Wang et al., 2019). Given the uniqueness of paleontological data, the design of specific base model is of considerable importance.
2) The task head is used to process the features offered by the base model and then infer the output. The classification task head is very simple, and can be realized by simply pooling the feature map and feeding the resulting vector into a classifier. The task head design becomes more challenging when describing the object details using fine-grained outputs (Qin et al., 2022b). In many cases, the deep learning model has multiple output branches to fulfill multi-task inference (Kim and Park, 2022). Multi-task prediction and fine-grained outputs are also required in many paleontological tasks.
3) The learning paradigm refers to how the model is trained. The most widely used paradigm is supervised learning on a large annotated dataset. The design of the training loss function significantly influences the training outcome (Wang et al., 2020a)). Since data collection and annotation are expensive and time-consuming, especially the case in paleontology, data-driven supervised learning can hardly be feasible. Therefore, a new learning paradigm is urgently needed to overcome the data sparsity problem. Many researchers have been working on *few-shot learning*, aiming to enable the AI model to learn rapidly from only a few known examples, just as humans learn (Wang et al., 2020b). Few-shot learning has been explored in many scenarios, such as CT medical image diagnosis and microscopic vision measurement (Chen et al., 2021 & Qin, 2023). Further, semi-supervised learning uses both labeled and unlabeled data for learning, which can also relieve the data sparsity problem and exploit huge amounts of raw data (Van Engelen and Hoos, 2020). In the future, we can expect more novel, specially designed AI models and high efficiency but low-cost learning paradigms for paleontological data processing.

4.4. New techniques

There have been many innovative and successful methods that are totally unfamiliar to paleontology, such as Generative Adversarial

Networks (GANs, Goodfellow et al., 2014), diffusion models (Sohl-Dickstein et al., 2015), and Large Language Models (LLMs, or Foundation Models, Bommasani et al., 2022).

GANs usually have two neural networks, a generator and a discriminator, that compete with each other to produce outputs that are as close to "realistic" samples as possible, which means they can be trained in both supervised and unsupervised manner. One of the major applications of GANs is content synthesis and manipulation that can possibly be used for complementing missing paleontological data, which is similar to image inpainting (Guillemot and Le Meur, 2014; Elharrouss et al., 2020). GANs has been applied to generate realistic petrographic datasets for data augmentation (Ferreira et al., 2022), and also foraminifera images (Ferreira-Chacua and Koeshidayatullah, 2023) to facilitate the creation of high-fidelity annotated datasets. Hou et al. (2023) applied GANs to apply super-resolution to microfossil images. GANs belong to a large family called generative models, to which diffusion models (which outperform GANs in image synthesis) also belong (Sohl-Dickstein et al., 2015; Ho et al., 2020). Diffusion models are trained to learn the structure of training datasets in the latent space, similar to the particle diffusion process in thermodynamics. They are then able to reverse the process to conduct tasks such as denoising images. The recently developed content generation systems DALL-E (https://openai.com/product/dall-e-2) and Stable Diffusion (https://stability.ai/stable-diffusion) have shown astonishing performance in text-to-image generation using diffusion models, and can potentially be used for more complicated and customized content such as videos.

Most (vertebrate) skeletal fossils are incomplete and soft tissues are rarely preserved in fossils. As such, the original organisms have to be reconstructed based on anatomical structures, environmental constraints, and inferred morphology from closely related taxa. Such reconstruction is necessarily subjective. Further, along with the reconstruction of fossil organisms themselves, the reconstruction of phylogenetic relationships is often impeded by missing links in evolution resulting from the sparsity of the fossil records. As a similar task, image inpainting has been facilitated by AI during the last several decades (Guillemot and Le Meur, 2014; Elharrouss et al., 2020).

AI-based generative models allow researchers to hypothesize and even reconstruct what has been lost from fossil preservation on the basis of more comprehensive scope and reproducibility.

Since the proposal of the attention-based transformer by Vaswani et al. (2017), its parallelization has led to large pre-trained models, such as BERT (Bidirectional Encoder Representations from Transformers, Devlin et al., 2018) and GPT iterations (Generative Pre-Training Transformer, Radford et al., 2018 & 2019, Brown et al., 2020; OpenAI, 2023), which are also called Large Language Models (LLMs). These unified models have revolutionized the way we process and generate natural language by integrating multi-modal data (e.g., text, images, sequences, etc.) to accomplish various tasks (Radford et al., 2021). Models such as Bootstrapping Language-Image Pre-training (BLIP, Li et al., 2022) and Once for All (OFA, Wang et al., 2022) showed that comparable performance can be reached in not only classic machine learning datasets, but also unseen complex tasks and domains (e.g., building a website from scratch and understanding a funny joke). LLMs often have more than one billion parameters, dwarfing the complexity of current paleontological AI models. Due to their unprecedented large sizes and power, LLMs, or Foundation Models, and their concomitant training costs and emergent properties, have hugely impacted AI development and applications (Bommasani et al., 2022). There have been rapidly increasing LLMs-based studies across a variety of science field, including chemistry (Boiko et al., 2023), math (Romera-Paredes et al., 2023), medical sciences (Thirunavukarasu et al., 2023), and earth sciences (Lin et al., 2023). They can facilitate knowledge sharing among scientists by providing summary of lengthy research literatures, translating text across languages, generating hypothesis based on given scenarios, and even reveal previously hidden knowledge.

The combination of multi-modal data by foundation models requires capabilities for integrating information from different sources, such as fossil records, geological data, written descriptions, and extant organisms, to provide a more comprehensive understanding of ancient life. AI can be deeply involved in current laborious tasks in paleontology, including comparative fossil description, microfossil classification, fossil imaging data processing, and morphological character encoding, to assist paleontologists in conducting large-scale data-driven studies. Although none of those models has yet been tested on paleontological tasks, their future impact on paleontology may be greater than all other AI techniques we discussed above. However, disinformation generated by AI and the copyright of training data are two major challenges for AI development in both paleontology and other fields.

CRediT authorship contribution statement

Congyu Yu: Conceptualization, Investigation, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Fangbo Qin:** Investigation, Writing – original draft, Writing – review & editing. **Akinobu Watanabe:** Investigation, Writing – original draft, Writing – review & editing. **Weiqi Yao:** Investigation, Writing – review & editing. **Ying Li:** Investigation, Writing – review & editing. **Zichuan Qin:** Investigation, Writing – review & editing. **Yuming Liu:** Investigation. **Haibing Wang:** Investigation, Writing – review & editing. **Qigao Jiangzuo:** Investigation, Writing – review & editing. **Allison Y. Hsiang:** Investigation, Writing – review & editing. **Chao Ma:** Project administration, Writing – review & editing. **Emily Rayfield:** Project administration, Writing – review & editing. **Michael J. Benton:** Project administration, Writing – review & editing. **Xing Xu:** Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Congyu Yu reports financial support was provided by Deep-time Digital Earth (DDE) Big Science Program. Haibing Wang reports financial support was provided by the Youth Innovation Promotion Association, Chinese Academy of Sciences. Allison Y. Hsiang reports financial support was provided by Swedish Research Council (Vetenskapsrådet) Starting Researcher Grant. Chao Ma reports financial support was provided by Deep-time Digital Earth (DDE) Big Science Program. Xing Xu reports financial support was provided by National Natural Science Foundation of China. Xing Xu reports financial support was provided by Yunnan Revitalization Talent Support Program.

Data availability

All data used in this research are described in this article.

Acknowledgments

We thank Mark A. Norell and Jin Meng from the American Museum of Natural History for their valuable discussion at an early stage of this project, and Marc Andre Conard for discussion. This work is funded by the National Natural Science Foundation of China (Grant No. 42288201, 42272017); the Youth Innovation Promotion Association, CAS (2021068); Yunnan Revitalization Talent Support Program (202305AB350006); Swedish Research Council (Vetenskapsrådet) Starting Researcher Grant (ÄR-NT 2020-03515); Chengdu University of Technology Zhufeng Starting Grant (10912-KYQD2023-09966). This work is a contribution to the Deep-time Digital Earth (DDE) Big Science Program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.

org/10.1016/j.earscirev.2024.104765.

References

Adams, D.C., Rohlf, F.J., Slice, D.E., 2013. A field comes of age: geometric morphometrics in the 21st century. Hystrix Ital. J. Mammal. 24, 7–14.
Allmon, W.D., Dietl, G.P., Hendricks, J.R., Ross, R.M., 2018. Bridging the two fossil records: Paleontology's "big data" future resides in museum collections. In: Rosenberg, G.D., Clary, R.M. (Eds.), Museums at the Forefront of the History and Philosophy of Geology: History Made, History in the Making, vol. 535 0. Geological Society of America.
Amarathunga, D.C., Grundy, J., Parry, H., Dorin, A., 2021. Methods of insect image capture and classification: a Systematic literature review. Smart Agric. Technol. 1, 100023.
Aneja, D., Vora, S.R., Camci, E.D., Shapiro, L.G., Cox, T.C., 2015. Automated detection of 3D landmarks for the elimination of non-biological variation in geometric morphometric analyses. In: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, pp. 78–83. https://doi.org/10.1109/CBMS.2015.86.
Anemone, R., Emerson, C., Conroy, G., 2011. Finding fossils in new ways: an artificial neural network approach to predicting the location of productive fossil localities. Evol. Anthropol. 20, 169–180.
Apostol, L.A., Márquez, E., Gasmen, P., Solano, G., 2016. RadSS: A radiolarian classifier using support vector machines. In: 2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA), pp. 1–6. https://doi.org/10.1109/IISA.2016.7785347.
Arriaza, M.C., et al., 2023. An evaluation of landmark-based methods to explore tooth score morphology: a case study on felids and hyenids. Appl. Sci. 13, 3864.
Athersuch, J., Banner, F., Higgins, A., Howarth, R., Swaby, P., 1994. The application of expert systems to the identification and use of microfossils in the petroleum industry. Math. Geol. 26, 483–489.
Baraniuk, R., Donoho, D., Gavish, M., 2020. The science of deep learning. PNAS 117, 30029–30032.
Basu, M., 2002. Gaussian-based edge-detection methods-a survey. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 32, 252–260.
Beaufort, L., Dollfus, D., 2004. Automatic recognition of coccoliths by dynamical neural networks. Mar. Micropaleontol. 51, 57–73.
Beaufort, L., de Garidel-Thoron, T., Mix, A.C., Pisias, N.G., 2001. ENSO-like forcing on oceanic primary production during the late Pleistocene. Science 293, 2440–2444.
Beaufort, L., et al., 2022. Cyclic evolution of phytoplankton forced by changes in tropical seasonality. Nature 601, 79–84.
Beightol, D.S., Conrad, M.A., 1988. Expert systems identify fossils and manage large paleontological databases. Geobyte; (United States) 3:1.
Bell, M.Z., 1985. Why expert systems fail. J. Oper. Res. Soc. 36, 613–619.
Belyea, P.R., Thunell, R.C., 1984. Fourier shape snalysis and planktonic foraminiferal evolution: the Neogloboquadrina-Pulleniatina lineages. J. Paleontol. 58, 1026–1040.
Bengio, Y., LeCun, Y., 2007. Scaling Learning Algorithms towards AI.
Bengio, Y., Lecun, Y., Hinton, G., 2021. Deep learning for AI. Commun. ACM 64, 58–65.
Benton, M.J., 1995. Diversification and extinction in the history of life. Science 268, 52–58.
Bergen, K.J., Johnson, P.A., de Hoop, M.V., Beroza, G.C., 2019. Machine learning for data-driven discovery in solid Earth geoscience. Science 363, eaau0323.
Bhullar, B.-A.S., et al., 2012. Birds have paedomorphic dinosaur skulls. Nature 487, 223–226.
Bi, K., et al., 2023. Accurate medium-range global weather forecasting with 3D neural networks. Nature 1–6. https://doi.org/10.1038/s41586-023-06185-3.
Boddy, L., Morris, C.W., Wilkins, M.F., Tarran, G.A., Burkill, P.H., 1994. Neural network analysis of flow cytometric data for 40 marine phytoplankton species. Cytometry 15, 283–293.
Boiko, D.A., MacKnight, R., Kling, B., Gomes, G, 2023. Autonomous chemical research with large language models. Nature 624, 570–578.
Bollmann, J., et al., 2004. Automated particle analysis: Calcareous microfossils. In: Francus, P. (Ed.), Image Analysis, Sediments and Paleoenvironments. Springer, Netherlands, pp. 229–252. https://doi.org/10.1007/1-4020-2122-4_12.
Bommasani, R., et al., 2022. On the opportunities and risks of foundation models. Preprint at. https://doi.org/10.48550/arXiv.2108.07258.
Bookstein, F.L., 1992. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press. https://doi.org/10.1017/CBO9780511573064.
Bourel, B., et al., 2020. Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains. Comput. Geosci. 140, 104498.
Boyer, D.M., Winchester, J.M., Glynn, C., Puente, J., 2015. Detailed anatomical orientations for certain types of morphometric measurements can be determined automatically with geometric algorithms. Anat. Rec. 298, 1816–1823.
Bromiley, P.A., Schunke, A.C., Ragheb, H., Thacker, N.A., Tautz, D., 2014. Semi-automatic landmark point annotation for geometric morphometrics. Front. Zool. 11, 61.
Brough, D.R., Alexander, I.F., 1986. The Fossil expert system. Expert. Syst. 3, 76–83.
Brown, T., et al., 2020. Language models are few-shot learners. Adv. Neural Inf. Proces. Syst. 33, 1877–1901.
Budd, G.E., 2021. Morphospace. Curr. Biol. 31, R1181–R1185.
Burke, C.D., Full, W.E., Gernant, R.E., 1987. Recognition of fossil fresh water ostracodes: Fourier shape analysis. Lethaia 20, 307–314.
Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. 679–698.
Carlsson, V., et al., 2023. Convolutional neural network application on a new middle Eocene radiolarian dataset. Mar. Micropaleontol. 183, 102268.
Chen, X., Jia, S., Xiang, Y., 2020. A review: Knowledge reasoning over knowledge graph. Expert Syst. Appl. 141, 112948.
Chen, X., Yao, L., Zhou, T., Dong, J., Zhang, Y., 2021. Momentum contrastive learning for few-shot COVID-19 diagnosis from chest CT images. Pattern Recogn. 113, 107826.
Cheng, Q., Oberhänsli, R., Zhao, M., 2020. A new international initiative for facilitating data-driven Earth science transformation. https://doi.org/10.1144/SP499-2019-158.
Choiniere, J.N., et al., 2021. Evolution of vision and hearing modalities in theropod dinosaurs. Science 372, 610–613.
Claes, P., et al., 2018. Genome-wide mapping of global-to-local genetic effects on human facial shape. Nat. Genet. 50, 414–423.
Courtenay, L.A., et al., 2019. Combining machine learning algorithms and geometric morphometrics: a study of carnivore tooth marks. Palaeogeogr. Palaeoclimatol. Palaeoecol. 522 (28–39), 1.
da Conceição, D.M., Esperança Júnior, M.G.F., Iannuzzi, R., Recamonde-Mendoza, M., Malta, G.B.B.O., 2023. PaleoWood: a machine learning approach for determining the affinity of Paleozoic gymnosperm woods. J. S. Am. Earth Sci. 121, 104125.
Dai, X., et al., 2023. A Mesozoic fossil lagerstätte from 250.8 million years ago shows a modern-type marine ecosystem. Science 379, 567–572.
Dalmasso, G., et al., 2022. 4D reconstruction of murine developmental trajectories using spherical harmonics. Dev. Cell 57, 2140–2150.e5.
Daood, A., 2018. Pollen Grains Recognition Based on Computer Vision Methods. Florida Institute of Technology.
Davis, R., 1982. Expert systems: where are we? And where do we go from here? AI Mag. 3, 3.
Dawson, H.L., Dubrule, O., John, C.M., 2023. Impact of dataset size and convolutional neural network achitecture on transfer learning for carbonate rock classification. Comput. Geosci. 171, 105284.
De Cesaro Júnior, T., Rieder, R., 2020. Automatic identification of insects from digital images: a survey. Comput. Electron. Agric. 178, 105784.
De Garidel-Thoron, T., Marchant, R., Tetard, M., Adebayo, M., Gally, Y., . Automated recognition and picking of foraminifera using the MiSo (microfossil sorter) prototype. https://meetingorganizer.copernicus.org/EGU2020/EGU2020-18067.html. https://doi.org/10.5194/egusphere-egu2020-18067.
de Geus, A.R., Barcelos, C.A.Z., Batista, M.A., da Silva, S.F., 2019. Large-scale pollen recognition with deep learning. In: 2019 27th European Signal Processing Conference (EUSIPCO), pp. 1–5. https://doi.org/10.23919/EUSIPCO.2019.8902735.
Deng, J., et al., 2009. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
Dollfus, D., 1997. Reconnaissance de formes naturelles par des réseaux de neurones artificiels: Application au nannoplancton calcaire. Aix-Marseille 3.
Dollfus, D., Beaufort, L., 1996. Automatic pattern recognition of calcareous nanoplankton. In: Proceedings of the Conference on Neural Networks and their Applications (NEURAP 96), pp. 306–311.
Dollfus, D., Beaufort, L., 1999. Fat neural network for recognition of position-normalised objects. Neural Netw. 12, 553–560.
Dosovitskiy, A., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.
Elharrouss, O., Almaadeed, N., Al-Maadeed, S., Akbari, Y., 2020. Image inpainting: a review. Neural. Process. Lett. 51, 2007–2028.
Emmings, J.F., et al., 2022. Pyrite mega-analysis reveals modes of anoxia through geological time. Sci. Adv. 8, eabj5687.
Fan, J., et al., 2020. A high-resolution summary of Cambrian to early Triassic marine invertebrate biodiversity. Science 367, 272–277.
Fan, L., Xu, C., Jarzembowski, E.A., Cui, X., 2021. Quantifying plant mimesis in fossil insects using deep learning. Hist. Biol. 34, 907–916.
Ferreira, I., Ochoa, L., Koeshidayatullah, A., 2022. On the generation of realistic synthetic petrographic datasets using a style-based GAN. Sci. Rep. 12, 12845.
Ferreira-Chacua, I., Koeshidayatullah, A., 2023. ForamViT-GAN: Exploring New Paradigms in Deep Learning for Micropaleontological Image Analysis. Preprint at. https://doi.org/10.48550/arXiv.2304.04291.
France, I., Duller, A.W.G., Duller, G.A.T., 2004. Software aspects of automated recognition of particles: The example of pollen. In: Image Analysis, Sediments and Paleoenvironments, pp. 253–272.
Fukushima, K., 1980. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybern. 36, 193–202.
Garratt, J.A., 1992. Morphological Data from Coccolith Images Using Fourier Power Spectra. Kingston University.
Garratt, J., Swan, A., 1992. Morphological data from coccolith images. in 11–34.
Gaston, K.J., O'Neill, M.A., 2004. Automated species identification: why not? Philos. Trans. Royal Soc. Lond. B 359, 655–667.
Ge, Q., et al., 2017. Coarse-to-fine foraminifera image segmentation through 3D and deep features. In: In 2017 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, pp. 1–8. https://doi.org/10.1109/SSCI.2017.8280982.
Goodfellow, I.J., et al., 2014. Generative adversarial networks. arXiv:1406.2661 [cs, stat].
Goswami, A., et al., 2022. Attenuated evolution of mammals through the Cenozoic. Science 378, 377–383.

Guillemot, C., Le Meur, O., 2014. Image inpainting: overview and recent advances. IEEE Signal Process. Mag. 31, 127–144.
Hammer, Ø., Harper, D.A., 2001. PAST: paleontological statistics software package for education and data analysis. Palaeontol. Electron. 4, 1.
Hammer, Ø., Harper, D.A., 2008. Paleontological Data Analysis. John Wiley & Sons.
Hanson, M., Hoffman, E.A., Norell, M.A., Bhullar, B.-A.S., 2021. The early origin of a birdlike inner ear and the evolution of dinosaurian movement and vocalization. Science 372, 601–609.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. Preprint at. https://doi.org/10.48550/arXiv.1603.05027.
He, Y., et al., . Challenges and opportunities in applying AI to evolutionary morphology. Preprint at. https://ecoevorxiv.org/repository/view/6525.
Healy-Williams, N., 1983. Fourier shape analysis of *Globorotalia truncatulinoides* from late Quaternary sediments in the southern Indian Ocean. Mar. Micropaleontol. 8, 1–15.
Healy-Williams, N., 1984. Quantitative image analysis: Application to planktonic foraminiferal paleoecology and evolution. Geobios 17, 425–432.
Hennig, W., 1966. Phylogenetic Systematics. University of Illinois Press.
Hills, S.J., 1988. Outline extraction of microfossils in reflected light images. Comput. Geosci. 14, 481–488.
Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.
Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A fast learning algorithm for deep belief nets. Neural Comput. 18, 1527–1554.
Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Proces. Syst. 33, 6840–6851.
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. Rev. Palaeobot. Palynol. 167, 175–183.
Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Netw. 2, 359–366.
Hou, Y., et al., 2020. ADMorph: a 3D digital microfossil morphology dataset for deep learning. IEEE Access 8, 148744–148756.
Hou, Y., Canul-Ku, M., Cui, X., Hasimoto-Beltran, R., Zhu, M., 2021. Semantic segmentation of vertebrate microfossils from computed tomography data using a deep learning approach. J. Micropalaeontol. 40, 163–173.
Hou, C., et al., 2023. Fossil image identification using deep learning ensembles of data augmented multiviews. Preprint at. https://doi.org/10.48550/arXiv.2302.08062.
Hou, Y., Canul-Ku, M., Cui, X., Zhu, M., 2023. Super-resolution of vertebrate microfossil computed tomography images based on deep learning. X-ray Spectrometry 2023, 1.
Høye, T.T., et al., 2021. Deep learning and computer vision will transform entomology. Proc. Natl. Acad. Sci. 118, e2002545117.
Hsiang, A.Y., et al., 2019. Endless forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks. Paleoceanogr. Paleoclimatol. 34, 1157–1177.
Huang, H., et al., 2023. An image dataset of fusulinid foraminifera generated with the aid of deep learning. Geosci. Data J. 11, 46–56.
Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science 349, 255–260.
Karaderi, T., Burghardt, T., Hsiang, A.Y., Ramaer, J., Schmidt, D.N., 2021. Visual microfossil identification via Deep Metric Learning. In: International Conference on Pattern Recognition and Artificial Intelligence. Springer International Publishing, Cham, pp. 34–46.
Kasinathan, T., Singaraju, D., Uyyala, S.R., 2021. Insect classification and detection in field crops using modern machine learning techniques. Inf. Process. Agric. 8, 446–457.
Kaya, Y., Pınar, S.M., Erez, M.E., Fidan, M., 2013. An expert classification system of pollen of Onopordum using a rough set approach. Rev. Palaeobot. Palynol. 189, 50–56.
Keçeli, A.S., Kaya, A., Keçeli, S.U., 2017. Classification of radiolarian images with hand-crafted and deep features. Comput. Geosci. 109, 67–74.
Keçeli, A.S., Keçeli, S.U., Kaya, A., 2018. Classification of radiolarian fossil images with deep learning methods. In: 2018 26th Signal Processing and Communications Applications Conference (SIU). IEEE, pp. 1–4. https://doi.org/10.1109/SIU.2018.8404460.
Kim, Y.H., Park, K.R., 2022. MTS-CNN: Multi-task semantic segmentation-convolutional neural network for detecting crops and weeds. Comput. Electron. Agric. 199, 107146.
Kjær, K.H., et al., 2022. A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. Nature 612, 283–291.
Koeshidayatullah, A., 2022. Optimizing image-based deep learning for energy geoscience via an effortless end-to-end approach. J. Petrol. Sci. Eng. 215, 110681.
Koeshidayatullah, A., Morsilli, M., Lehrmann, D.J., Al-Ramadan, K., Payne, J.L., 2020. Fully automated carbonate petrography using deep convolutional neural networks. Mar. Pet. Geol. 122, 104687.
Kong, S., Punyasena, S., Fowlkes, C., 2016. Spatially aware dictionary learning and coding for fossil pollen identification. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, pp. 1305–1314. https://doi.org/10.1109/CVPRW.2016.165.
Kong, Q., et al., 2019. Machine learning in seismology: Turning data into insights. Seismol. Res. Lett. 90, 3–14.
Kopperud, B.T., Lidgard, S., Liow, L.H., 2019. Text-mined fossil biodiversity dynamics using machine learning. Proc. R. Soc. B Biol. Sci. 286, 20190022.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25. Curran Associates, Inc.
Lallensack, J.N., Romilio, A., Falkingham, P.L., 2022. A machine learning approach for the discrimination of theropod and ornithischian dinosaur tracks. J. R. Soc. Interface 19, 20220588.
LeCun, Y., et al., 1989. Backpropagation applied to handwritten zip code recognition. Neural Comput. 1, 541–551.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Li, P., Treloar, W.J., Flenley, J.R., Empson, L., 2004. Towards automation of palynology 2: the use of texture measures and neural network analysis for automated identification of optical images of pollen grains. J. Quat. Sci. 19, 755–762.
Li, J., Li, D., Xiong, C., Hoi, S., 2022. BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation. Preprint at. https://doi.org/10.48550/arXiv.2201.12086.
Lin, Z., et al., 2023. GeoGalactica: A Scientific Large Language Model in Geoscience. Preprint at. https://doi.org/10.48550/arXiv.2401.00434.
Litjens, G., et al., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.
Liu, X., Song, H., 2020. Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional neural networks. Sediment. Geol. 410, 105790.
Liu, S., Thonnat, M., Berthod, M., 1994. Automatic classification of planktonic foraminifera by a knowledge-based system. In: Proceedings of the Tenth Conference on Artificial Intelligence for Applications, pp. 358–364. https://doi.org/10.1109/CAIA.1994.323653.
Liu, L., et al., 2019. PestNet: an end-to-end deep learning approach for large-scale multi-class pest detection and classification. IEEE Access 7, 45301–45312.
Liu, X., et al., 2023. Automatic taxonomic identification based on the Fossil Image Dataset (>415,000 images) and deep convolutional neural networks. Paleobiology 49, 1–22.
Lungmus, J.K., Angielczyk, K.D., 2019. Antiquity of forelimb ecomorphological diversity in the mammalian stem lineage (Synapsida). Proc. Natl. Acad. Sci. 116, 6903–6907.
MacLeod, N., Benfield, M., Culverhouse, P., 2010. Time to automate identification. Nature 467, 154–155.
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, 183–202.
Marmo, R., Amodio, S., 2006. A neural network for classification of chambers arrangement in Foraminifera. In: Di Gesú, V., Masulli, F., Petrosino, A. (Eds.), Fuzzy Logic and Applications. Springer, pp. 271–278. https://doi.org/10.1007/10983652_33.
Marmo, R., Amodio, S., Cantoni, V., 2006. Microfossils shape classification using a set of width values. In: 18th International Conference on Pattern Recognition (ICPR'06), vol. 1, pp. 691–694.
Martineau, M., et al., 2017. A survey on image-based insect classification. Pattern Recogn. 65, 273–284.
Martín-Perea, D.M., Courtenay, L.A., Domingo, M.S., Morales, J., 2020. Application of artificially intelligent systems for the identification of discrete fossiliferous levels. PeerJ 8, e8767.
Matthew, W.D., 1926. The evolution of the horse: a record and its interpretation. Q. Rev. Biol. 1, 139–185.
McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E., 2006. A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. AI Mag. 27, 12.
Mete, Ö., et al., 2023. Barium in seawater: dissolved distribution, relationship to silicon, and barite saturation state determined using machine learning. Earth Syst. Sci. Data Discuss. 1–42 https://doi.org/10.5194/essd-2023-67.
Mitra, R., et al., 2019. Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance. Mar. Micropaleontol. 147, 16–24.
Mousavi, S.M., Beroza, G.C., 2023. Machine learning in earthquake seismology. Annu. Rev. Earth Planet. Sci. 51, 105–129.
Musgrave, K., Belongie, S., Lim, S.-N., . A metric learning reality check. Preprint at. http://arxiv.org/abs/2003.08505.
Newell, A., Yang, K., Deng, J., 2016. Stacked hourglass networks for human pose estimation. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII 14. Springer, pp. 483–499.
Nicholson, D.B., Mayhew, P.J., Ross, A.J., 2015. Changes to the fossil record of insects through fifteen years of discovery. PLoS One 10, e0128554.
O'Leary, M.A., et al., 2013. The placental mammal ancestor and the post–K-Pg radiation of placentals. Science 339, 662–667.
OpenAI, 2023. GPT-4 Technical report. Preprint at. https://doi.org/10.48550/arXiv.2303.08774.
Payne, J.L., Finnegan, S., 2007. The effect of geographic range on extinction risk during background and mass extinction. Proc. Natl. Acad. Sci. 104, 10506–10511.
Pires de Lima, R., Marfurt, K., 2020. Convolutional neural network for remote-sensing scene classification: transfer learning analysis. Remote Sens. 12, 86.
Pomidor, B.J., Makedonska, J., Slice, D.E., 2016. A landmark-free method for three-dimensional shape analysis. PLoS One 11, e0150368.
Porto, A., Voje, K.L., 2020. ML-morph: a fast, accurate and general approach for automated detection and landmarking of biological structures in images. Methods Ecol. Evol. 11, 500–512.
Porto, A., Rolfe, S., Maga, A.M., 2021. ALPACA: a fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures. Methods Ecol. Evol. 12, 2129–2144.
Qin, F., et al., 2023. Automated hooking of biomedical microelectrode guided by intelligent microscopic vision. IEEE/ASME Trans. Mechatron. 28, 2786–2798.

Qin, Z., Qin, F., Li, Y., Yu, C., 2022a. Intelligent objective osteon segmentation based on deep learning. Front. Earth Sci. 10.

Qin, F., Xu, D., Hannaford, B., Hao, T., 2022b. Object-agnostic vision measurement framework based on one-shot learning and behavior tree. IEEE Trans. Cybernet. 53, 5202–5215.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., 2018. Improving language understanding by generative pre-training. OpenAI paper at. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

Radford, A., et al., 2019. Language Models are Unsupervised Multitask Learners. Papers with Code. paper at. https://paperswithcode.com/paper/language-models-are-unsupervised-multitask.

Radford, A., et al., 2021. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning. PMLR, pp. 8748–8763.

Rani, P., Kotwal, S., Manhas, J., Sharma, V., Sharma, S., 2022. Machine learning and deep learning based computational approaches in automatic microorganisms image recognition: methodologies, challenges, and developments. Arch. Comput. Methods Eng. 29, 1801–1837.

Raup, D.M., 1966. Geometric analysis of shell coiling: general problems. J. Paleontol. 1178–1190.

Raup, D.M., Sepkoski, J.J., 1982. Mass extinctions in the marine fossil record. Science 215, 1501–1503.

Reichstein, M., et al., 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204.

Richmond, T., et al., 2022. Forabot: Automated planktic Foraminifera isolation and imaging. Geochem. Geophys. Geosyst. 23, e2022GC010689.

Riedel, R., 1989. Identify: a Prolog program to help identify fossils. Comput. Geosci. 15, 809–823.

Rohlf, F., Marcus, L.F., 1993. A revolution morphometrics. Trends Ecol. Evol. 8, 129–132.

Rolfe, S., et al., 2021. SlicerMorph: an open and extensible platform to retrieve, visualize and analyse 3D morphology. Methods Ecol. Evol. 12, 1816–1825.

Romera-Paredes, B., et al., 2023. Mathematical discoveries from program search with large language models. Nature 1–3. https://doi.org/10.1038/s41586-023-06924-6.

Rosenfeld, A., Solbach, M.D., Tsotsos, J.K., 2018. Totally looks like-how humans compare, compared to machines. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1961–1964.

Schiebel, R., et al., 2003. Automated analysis of planktic foraminifers part III: Neural network classification. In: EGS-AGU-EUG Joint Assembly, p. 12531.

Sepkoski, J.J., 1978. A kinetic model of Phanerozoic taxonomic diversity I. Analysis of marine orders. Paleobiology 4, 223–251.

Sepkoski, J.J., 1979. A kinetic model of Phanerozoic taxonomic diversity II. Early Phanerozoic families and multiple equilibria. Paleobiology 5, 222–251.

Sepkoski, J.J., 1984. A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. Paleobiology 10, 246–267.

Sepkoski, D., 2013. Towards "a natural history of data": evolving practices and epistemologies of data in paleontology, 1800–2000. J. Hist. Biol. 46, 401–444.

Shen, L., Farid, H., McPeek, M.A., 2009. Modeling three-dimensional morphological structures using spherical harmonics. Evolution 63, 1003–1016.

Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19, 221–248.

Simpson, G.G., 1944. Tempo and Mode in Evolution. Columbia University Press, New York.

Slice, D.E., 2007. Geometric morphometrics. Annu. Rev. Anthropol. 36, 261–281.

Smith, J.A., et al., 2023. Increasing the equitability of data citation in paleontology: capacity building for the big data future. Paleobiology 1–12. https://doi.org/10.1017/pab.2023.33.

Sneath, P.H.A., 1979. Numerical taxonomy and automated identification: some implications for geology. Comput. Geosci. 5, 41–46.

Sneath, P.H.A., Sokal, R.R., 1962. Numerical taxonomy. Nature 193, 855–860.

Sneath, P.H.A., Sokal, R.R., 1973. Numerical Taxonomy. The Principles and Practice of Numerical Classification. W.H. Freeman, New York.

Snyders, J., Claes, P., Vandermeulen, D., Suetens, P., 2014. Development and Comparison of Non-rigid Surface Registration Algorithms and Extensions. Technical Report KUL/ESAT/PSI/1401. KU Leuven, ESAT, Leuven, Belgium.

Soda, K.J., Slice, D.E., Naylor, G.J.P., 2017. Artificial neural networks and geometric morphometric methods as a means for classification: a case-study using teeth from *Carcharhinus* sp. (Carcharhinidae). J. Morphol. 278, 131–141.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. PMLR, pp. 2256–2265.

Sun, Z., et al., 2022. A review of earth artificial intelligence. Comput. Geosci. 159, 105034.

Swaby, P.A., 1990. Integrating artificial intelligence and graphics in a tool for microfossil identification for use in the petroleum industry. In: Proceedings of the 2nd Annual Conference on Innovative Applications of Artificial Intelligence, Washington, pp. 203–218.

Swaby, P., 1992. VIDES: an expert system for visually identifying microfossils. IEEE Expert 7, 36–42.

Tao, F., et al., 2023. Microbial carbon use efficiency promotes global soil carbon storage. Nature 618, 981–985. https://doi.org/10.1038/s41586-023-06042-3.

Thirunavukarasu, A.J., et al., 2023. Large language models in medicine. Nat. Med. 1–11 https://doi.org/10.1038/s41591-023-02448-8.

Thomas, O.O., et al., 2023. Automated morphological phenotyping using learned shape descriptors and functional maps: a novel approach to geometric morphometrics. PLoS Comput. Biol. 19, e1009061.

Tolstikhin, I.O., et al., 2021. Mlp-mixer: an all-mlp architecture for vision. Adv. Neural Inf. Proces. Syst. 34, 24261–24272.

Toosi, A., Bottino, A., Saboury, B., Siegel, E., Rahmim, A., 2021. A brief history of AI: How to prevent another winter (a critical review). arXiv:2109.01517 [cs].

Treloar, W.J., Taylor, G.E., Flenley, J.R., 2004. Towards automation of palynology 1: analysis of pollen shape and ornamentation using simple geometric measures, derived from scanning electron microscope images. J. Quat. Sci. 19, 745–754.

Valan, M., Makonyi, K., Maki, A., Vondráček, D., Ronquist, F., 2019. Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. Syst. Biol. 68, 876–895.

Van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109, 373–440.

Vandaele, R., et al., 2018. Landmark detection in 2D bioimages for geometric morphometrics: a multi-resolution tree-based approach. Sci. Rep. 8, 538.

Vaswani, A., et al., 2017. Attention is all you need. Adv. Neural Inf. Proces. Syst. 30.

Wang, Y., et al., 2019. Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. 38, 1–12.

Wang, Q., Ma, Y., Zhao, K., Tian, Y., 2020a. A comprehensive survey of loss functions in machine learning. Ann. Data Sci. 1–26.

Wang, Y.Q., Yao, Q., Kwok, J.T., Ni, L.M., 2020b. Generalizing from a few examples: a survey on few-shot learning. ACM Comput. Surv. 53, 1–34.

Wang, P., et al., 2022. OFA: Unifying architectures, tasks, and modalities through a simple sequence-to-sequence learning framework. Preprint at. https://doi.org/10.48550/arXiv.2202.03052.

Wills, S., Underwood, C.J., Barrett, P.M., 2023. Machine learning confirms new records of maniraptoran theropods in Middle Jurassic UK microvertebrate faunas. Pap. Palaeontol. 9, e1487.

Wöber, W., et al., 2022. Investigating shape variation using generalized Procrustes analysis and machine learning. Appl. Sci. 12, 3158.

Wong, C.M., 2011. Human-Based Computation for Microfossil Identification. University of Alberta.

Wu, X., Liang, L., Shi, Y., Fomel, S., 2019. FaultSeg3D: using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. Geophysics 84. IM35–IM45.

Xu, Y.X., Dai, Z., Wang, J., Li, Y., Wang, H., 2020. Automatic recognition of Palaeobios images under microscope based on machine learning. IEEE Access 8, 172972–172981.

Xu, Y., et al., 2021. Artificial intelligence: a powerful paradigm for scientific research. Innovation 2, 100179.

Xu, C., et al., 2022. Widespread mimicry and camouflage among mid-cretaceous insects. Gondwana Res. 101, 94–102.

Yao, F., et al., 2023. Satellites reveal widespread decline in global lake water storage. Science 380, 743–749.

Yu, S., Saint-Marc, P., Thonnat, M., Berthod, M., 1996. Feasibility study of automatic identification of planktic foraminifera by computer vision. J. Foraminifer. Res. 26, 113–123.

Yu, C., Jiangzuo, Q., Tschopp, E., Wang, H., Norell, M., 2021. Information in morphological characters. Ecol. Evol. 11, 11689–11699.

Yu, C., Qin, F., Li, Y., Qin, Z., Norell, M., 2022. CT segmentation of dinosaur fossils by deep learning. Front. Earth Sci. 9, 805271.

Zhang, Y., Fountain, D.W., Hodgson, R.M., Flenley, J.R., Gunetileke, S., 2004. Towards automation of palynology 3: pollen pattern recognition using Gabor transforms and digital moments. J. Quat. Sci. 19, 763–768.

Zhang, Y.C., et al., 2023. Skilful nowcasting of extreme precipitation with NowcastNet. Nature 619, 526–532. https://doi.org/10.1038/s41586-023-06184-4.

Zheng, W.-T., et al., 2022. FossilMorph: a new tool for fossil automatic measurement and statistical analysis. Palaeoworld. https://doi.org/10.1016/j.palwor.2022.11.001.

Zhong, B., Ge, Q., Kanakiya, B., Marchitto, R.M.T., Lobaton, E., 2017. A comparative study of image classification algorithms for Foraminifera identification. In: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–8. https://doi.org/10.1109/SSCI.2017.8285164.

Zhou, Z., Zhu, R., Wang, M., 2021. Spatiotemporal evolution of the Jehol Biota: Responses to the North China cration destruction in the Early Cretaceous. PNAS 118.