RESEARCH ARTICLE

Ammonoid taxonomy with supervised and unsupervised machine learning algorithms

Floe Foxon¹

¹ Department of Data Management and Statistical Analysis, PinneyAssociates, Inc. – Pittsburgh, USA

Cite as: Foxon F (2021). Ammonoid taxonomy with supervised and unsupervised machine learning algorithms. *PaleorXiv ewkx9*, ver. 3, peer-reviewed by PCI Paleo. DOI: 10.31233/osf.io/ewkx9

Published: 1st October 2021

Recommender:

Kenneth De Baets

Reviewers:

Jérémie Bardin and an anonymous reviewer

Correspondence:

ffoxon@pinneyassociates.com

THIS ARTICLE WAS PEER-REVIEWED BY PCI PALEO

Read the Editor's recommendation and Referees' reports at [DOI:10.24072/pci.paleo.100010](https://doi.org/10.24072/pci.paleo.100010)

Abstract

Ammonoid identification is crucial to biostratigraphy, systematic palaeontology, and evolutionary biology, but may prove difficult when shell features and sutures are poorly preserved. This necessitates novel approaches to ammonoid taxonomy. This study aimed to determine the taxa of ammonoids by their conch geometry using supervised and unsupervised machine learning algorithms. Ammonoid measurement data (conch diameter, whorl height, whorl width, and umbilical width) were taken from the Paleobiology Database (PBDB). 11 species with ≥ 50 specimens each were identified providing $N = 781$ total unique specimens. Naive Bayes, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbours, Support Vector Machine, and Multilayer Perceptron classifiers were applied to the PBDB data with a 5×5 nested cross-validation approach to obtain unbiased generalization performance estimates across a grid search of algorithm parameters. All supervised classifiers achieved $\geq 70\%$ accuracy in identifying ammonoid species, with Naive Bayes demonstrating the least over-fitting. The unsupervised clustering algorithms K-Means, DBSCAN, OPTICS, Mean Shift, and Affinity Propagation achieved Normalized Mutual Information scores of ≥ 0.6 , with the centroid-based methods having most success. This presents a reasonably-accurate proof-of-concept approach to ammonoid classification which may assist identification in cases where more traditional methods are not feasible.

© 2021 Author(s)

This work is licensed under the Creative Commons Attribution 4.0 International License.

Introduction

Ammonoidea represent a morphologically diverse group of extinct cephalopods, ranging from the Early Devonian to the Cretaceous/Paleogene boundary (Ritterbush et al., 2014; Walker and Ward, 2002). Found in marine sedimentary rocks, ammonoids are crucial index fossils for biostratigraphy (House, 1985; Cox, 1995). Furthermore, the use of ammonoids in systematic palaeontology (Kennedy, 1984) and evolutionary biology (Monnet et al., 2015) is substantive.

Ammonoid taxonomy utilizes conch morphology such as coiling and aperture shape. Ornaments such as ribs (their direction, spacing, and type) may be used for classification (De Baets et al., 2013a,b), as well as keels, spines, nodes, and tubercles. Where the shell is eroded or broken, suture lines across the sediment-filled interior are highly diagnostic for ammonoid order (Wiedmann and Kullmann, 1980; Klug and Hoffmann, 2015).

Korn (2010) standardized a number of conch geometry parameters and proportions for taxonomic study of ammonoids, shown in Table 1. Individual conchs may be described by terms based on the values of these conch proportions, for example ammonoids with $0.35 \leq CWI \leq 0.6$ have 'discoidal' general conch shape.

Since ammonoids exhibit intraspecific variation (De Baets et al., 2015), it follows that each species has a typical range of conch proportions which are diagnostic of taxonomy. Hohenegger and Tatzreiter (1992) attempted multivariate classification of *Balatonites* by their morphological parameters with discriminant analysis. This work identified two well-separated groups in data from 19 reported species, suggesting these really represent only two distinct biological species. In contrast, Bardin et al. (2015) more recently demonstrated no overlapping morphology in *Dactylioceras* specimens, suggesting these are well classified. Taken together, these studies imply that classification algorithms are useful tools for assessing the validity of taxonomy in the literature.

The aim of this study is to determine the taxa of ammonoids by their conch proportions, and expand upon the work of Hohenegger and Tatzreiter (1992) and Bardin et al. (2015) by implementing a range of supervised and unsupervised machine learning algorithms not previously applied to ammonoid taxonomy. This presents novel methodological research and lays the groundwork for future methods in biostratigraphy, systematic palaeontology, and evolutionary biology.

Table 1. Ammonoid conch parameters and proportions.

Parameter	Abbreviation	Equation
Largest conch diameter in ammonoid cross-section	dm_1	-
Largest whorl width in ammonoid cross-section	ww_1	-
Largest whorl height in ammonoid cross-section	wh_1	-
Umbilical width	uw	-
Conch width index	CWI	$\frac{ww_1}{dm_1}$
Whorl width index	WWI	$\frac{ww_1}{wh_1}$
Umbilical width index	UWI	$\frac{uw}{dm_1}$

Methods

Data

Ammonoid data were sourced from the Paleobiology Database (PBDB; Clapham et al., 2020), a non-governmental, non-profit public paleontological database supported by the US National Science Found-

dition. Data were downloaded from the PBDB on 10 November, 2020, using the taxon/taxa name 'Ammonoidea', which provided $N = 19,576$ unique specimens. Specimens with missing diameter, whorl height, whorl width, and umbilical width were excluded from analyses. To reduce over- or underfitting models, species with fewer than 50 specimens in the dataset were also removed. The remaining data consisted of $N = 781$ specimens from 11 species. Sources for these specimens are described in Table 2.

Table 2. Data sources.

Species	Source	N	Period	Region
<i>Badouxia canadensis</i>	Longridge et al. (2006)	148	Early Jurassic	British Columbia, Canada
<i>Owenites koeneni</i>	Shigeta and Nguyen (2014)	44	Early Triassic	Northeast Vietnam
	Kummel and Steele (1962)	16	Early Triassic	Nevada, United States
	Kuenzi (1965)	12	Early Triassic	Washington, United States
	Shevyrev (1995)	8	Early Triassic	Northwest Caucasus
	Brayard and Bucher (2008)	3	Early Triassic	Northwest Guangxi, South China
	Shevyrev (1990)	1	Early Triassic	West-Central Tajikistan
	Popov (1962)	1	Early Triassic	North Caucasus
	Brayard et al. (2013)	1	Early Triassic	Utah, United States
<i>Cladiscites crassestriatus</i>	Gamsjäger (1982)	76	Late Triassic	Austria and Indonesia
<i>Arkanites relictus</i>	Manger and Saunders (1980)	75	Late Carboniferous	Arkansas, United States
<i>Retites semiretia</i>	Manger and Saunders (1980)	65	Late Carboniferous	Arkansas, United States
<i>Karakaschiceras attenuatus</i>	Aguirre-Urreta (1998)	62	Early Cretaceous	West-Central Argentina
<i>Cladiscites beyrichi</i>	Gamsjäger (1982)	60	Late Triassic	Austria and Indonesia
<i>Meekoceras gracilitatis</i>	Jenks et al. (2010)	44	Early Triassic	Nevada, United States
	Tozer (1961)	7	Early Triassic	Queen Elizabeth Islands, Arctic Archipelago
	Brühwiler et al. (2012)	3	Early Triassic	North Pakistan
	Shevyrev (1995)	1	Early Triassic	Northwest Caucasus
	Dagis and Ermakova (1990)	1	Early Triassic	Siberia, Russia
<i>Paralegoceras sundaicum</i>	De Marez-Oyens (1933)	53	Early Permian	Timor
<i>Andidiscus behrendseni</i>	Hillebrandt (2006)	50	Early Jurassic	Argentina and Chile
<i>Eoamalteus multicostatus</i>	Hillebrandt (2006)	50	Early Jurassic	Argentina

Analyses

For each ammonoid, the conch parameters dm_1 , ww_1 , wh_1 , and uw were used to find the conch proportions CWI , WWI , and UWI as defined in Table 1. All specimens were then plotted to observe apparent within-population variability in three-dimensional proportion space.

Supervised algorithms

Classification was achieved with supervised machine learning algorithms. The target variable was the specimen species, and the fit data were the conch proportions.

First, a range of supervised algorithms were compared with a 5×5 nested cross-validation approach to determine which is most suitable for ammonoid taxonomy. Nested cross-validation was implemented to avoid potential bias in performance evaluation due to over-fitting in model selection (Cawley and Talbot, 2010). This way, estimates for the unbiased generalization performance of each classifier were obtained through test and train accuracies averaged across the outer folds. For the inner folds, a grid search was utilized to select the model parameters which resulted in the best test accuracy. The classifiers implemented, as well as the range of parameters used in cross-validation, are summarized in Table 3.

Table 3. Classification models for cross-validation.

Classifier	Parameters
Naive Bayes	Variance smoothing: 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1 , 10^2 , 10^3 , 10^4
Decision Tree	Depth: 1, 2, 3, ..., 49, 50 Impurity criterion: gini, entropy Splitter: best, random
Random Forest	Depth: 1, 2, 3, ..., 49, 50 Impurity criterion: gini, entropy Number of estimators: 100, 200
Gradient Boosting	Depth: 1, 2, 3, ..., 49, 50 Number of estimators: 100, 200 Learning rate: 0.1, 0.5, 1.0 Loss function: deviance Split quality criterion: Friedman MSE
K-Nearest Neighbours	K: 1, 2, 3, ..., 49, 50 Algorithm: ball tree, KD tree, brute-force search
Support Vector Machine	Regularization parameter: 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1 , 10^2 , 10^3 , 10^4 Kernel: linear, polynomial (order: 2, 3), radial basis function, sigmoid
Multilayer Perceptron	Regularization parameter: 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1 , 10^2 , 10^3 , 10^4 Solver: lbfgs, SGD, adam

Next, the model with the best test accuracy (and with similar train accuracy to prevent over-fitting) was implemented with a five-fold cross-validation approach to identify how performance varied for each species.

Unsupervised algorithms

In addition to the supervised classification algorithms, a range of unsupervised clustering algorithms were implemented. Because cross-validation is not a clearly defined concept for unsupervised algorithms, cross-validation was not implemented for these methods. The following describes the unsupervised algorithms used and how model selection was approached.

A K-Means clustering model was selected in the usual way with the ‘elbow’ method by selecting the value of k for which a reduction in the sum of squared distances of samples to their nearest cluster center is diminished (the ‘elbow’ of an inertia against k curve).

Similarly, a DBSCAN clustering model was selected with an ‘elbow’ method. First, a value for the minimum number of neighbours around a point to define a core point was chosen heuristically as $m = \lfloor \ln(N) \rfloor$. A nearest neighbours algorithm was then implemented to calculate the average distance between each point and its m nearest neighbours. The optimal value of epsilon was then selected as the distance which corresponded to the elbow on a distance against data point number curve (with the data points sorted by distance).

Building on the DBSCAN model, an OPTICS clustering model was implemented using a DBSCAN cluster method and the same values of epsilon and m .

The final two unsupervised models implemented were Mean Shift and Affinity Propagation clustering. ‘Elbow’ methods are not well defined for model selection with these two algorithms, therefore grid searches were used to select the models with the highest Caliński-Harabasz index or Variance Ratio Criterion (Caliński and Harabasz, 1974), which identifies the model with the most dense and well separated clusters. This internal validation metric is naive to the ground truth labels, and so keeps the model selection process unsupervised.

For the Mean Shift model, a range of bandwidths with quantile parameters 0.01, 0.02, . . . 0.15 were searched. The upper limit of 0.15 was selected because preliminary analyses showed models with quantiles > 0.15 predicted only two clusters. Pretending not to know the true number of species, a brief visual inspection of the ammonoid data suggests at least three clusters exist, so it is reasonable to exclude models predicting fewer clusters.

Similarly, the range of preference values $-0.394, -0.344, \dots -0.044$ were searched for the Affinity Propagation model. The lower limit of -0.394 and upper limit of -0.044 were selected because preliminary analyses showed models with preference < -0.394 predicted only one cluster, and models with positive preference predicted 781 clusters. Again, visual inspection of the data finds these numbers of clusters to be unrealistic, so it is reasonable to exclude such models.

After model selection, external validation was performed on all unsupervised methods by computing Fowlkes-Mallows scores (Fowlkes and Mallows, 1983), which compare the classes predicted by clustering to the actual classes via the geometric mean of the pairwise precision and recall. Normalized Mutual Information (NMI) scores (Strehl and Ghosh, 2003) were also computed to compare these models with extant literature, however this metric is not adjusted against chance.

All analyses were conducted in Python version 3.7.6 with the packages NumPy version 1.18.5, Matplotlib version 3.2.2, pandas version 1.0.5, and scikit-learn version 0.23.1. Code is provided in the [Supplementary information](#).

Results

The final analysis data are summarized in [Table 4](#). While some species share similar values on average for individual conch proportions, fewer have similar average values for all three conch proportions. Exceptions include *{Andidiscus behrendseni, Eoamalthus multicosatus, and Badouxia canadensis}*, and *{Arkanites relictus and Retites semiretia}*. Standard deviations on mean conch proportions are generally small relative to those proportions, suggesting relatively narrow spreads, which favour a classification approach.

Table 4. Final analysis data summary.

Species	N	Average <i>CWI</i> (SD)	Average <i>WWI</i> (SD)	Average <i>UWI</i> (SD)
<i>Badouxia canadensis</i>	148	0.30 (0.03)	0.89 (0.09)	0.41 (0.05)
<i>Owenites koeneni</i>	86	0.34 (0.07)	0.7 (0.1)	0.15 (0.07)
<i>Cladiscites crassestriatus</i>	76	0.45 (0.05)	0.8 (0.1)	0.00003 (0.00002)
<i>Arkanites relictus</i>	75	0.65 (0.07)	1.9 (0.3)	0.40 (0.06)
<i>Retites semiretia</i>	65	0.58 (0.04)	1.9 (0.3)	0.47 (0.05)
<i>Karakaschiceras attenuatus</i>	62	0.27 (0.03)	0.52 (0.08)	0.14 (0.03)
<i>Cladiscites beyrichi</i>	60	0.58 (0.05)	1.0 (0.1)	0.00003 (0.00001)
<i>Meekoceras gracilitatis</i>	56	0.23 (0.02)	0.46 (0.05)	0.16 (0.04)
<i>Paralegoceras sundaicum</i>	53	0.53 (0.07)	1.5 (0.3)	0.41 (0.09)
<i>Andidiscus behrendseni</i>	50	0.3 (0.1)	0.8 (0.3)	0.33 (0.06)
<i>Eoamalthus multicosatus</i>	50	0.25 (0.03)	0.73 (0.07)	0.40 (0.02)

The distribution of ammonoids in three-dimensional proportion space is shown in [Figure 1](#). Intra-species clustering is readily apparent for most species, especially *Badouxia canadensis*, *Cladiscites beyrichi*, *Meekoceras gracilitatis*, and *Karakaschiceras attenuatus*, which display the least intra-cluster distance. Within-population variability is considerably greater in the species *Arkanites relictus*, *Paralegoceras sundaicum*, and *Andidiscus behrendseni*.

Unbiased generalization performance estimates for the supervised models are shown in [Table 5](#).

Figure 1. Within-population variability of ammonoid conch proportions. Descriptive terms for proportion ranges are sourced from Korn (2010).

Due to its high test accuracy and similar train accuracy, the Support Vector Machine was selected as the best supervised model (see Discussion). Table 6 contains a detailed classification report for the cross-validated Support Vector Machine, grid searched over the same parameters as specified in Table 3.

Finally, Table 7 contains the scores for the unsupervised models, whose corresponding clusterings are shown in Figure 2.

Discussion

This research aimed to determine the taxa of ammonoids by well-defined geometric conch parameters using machine learning algorithms. Supervised models obtained at least 70% accuracy in ammonoid classification. Unsupervised models obtained Fowlkes-Mallows scores of at least 0.5 and Normalized Mutual Information scores of at least 0.6. The latter models predicted between five and 15 species when 11 were present. This proof-of-concept approach to ammonoid classification is novel in its application and may in future provide additional insight for biostratigraphy, systematic palaeontology, and evolutionary biology. A discussion of the models implemented follows.

Supervised models

The supervised models' average test accuracies ranged from 0.704 for the Decision Tree to 0.781 for the Support Vector Machine. As a sanity test, these results can be compared to a very simple classifier. On this imbalanced dataset, a naive classifier approach which simply predicts the majority class would achieve an accuracy of $\frac{148}{781} \approx 0.190$. Relative to this baseline classifier, all the models presented here

Table 5. Supervised model results.

Classifier	Test Accuracy (SD)	Train Accuracy (SD)
Support Vector Machine	0.781 (0.032)	0.844 (0.021)
Gradient Boosting	0.755 (0.039)	0.912 (0.045)
Random Forest	0.749 (0.045)	0.900 (0.067)
Naive Bayes	0.745 (0.034)	0.778 (0.012)
K-Nearest Neighbours	0.743 (0.040)	0.939 (0.074)
Multilayer Perceptron	0.737 (0.035)	0.798 (0.008)
Decision Tree	0.704 (0.038)	0.940 (0.064)

Table 6. Detailed classification report for support vector machine model.

Species	F1-Score	Precision	Recall
<i>Cladiscites crassestriatus</i>	0.96	0.96	0.96
<i>Cladiscites beyrichi</i>	0.95	0.95	0.95
<i>Owenites koeneni</i>	0.90	0.85	0.96
<i>Badouxia canadensis</i>	0.83	0.74	0.94
<i>Eoamalteus multicostatus</i>	0.74	0.88	0.64
<i>Meekoceras gracilitatis</i>	0.74	0.65	0.87
<i>Arkanites relictus</i>	0.72	0.67	0.78
<i>Karakaschiceras attenuatus</i>	0.70	0.88	0.58
<i>Retites semiretia</i>	0.65	0.55	0.80
<i>Paralegoceras sundaicum</i>	0.62	0.89	0.47
<i>Andidiscus behrendseni</i>	0.33	0.67	0.22

are far superior. The relative performance of these models however is less clear. The dispersions of the individual outer fold test accuracies as measured by the standard deviation on each model were somewhat large relative to the differences in accuracies, therefore for this particular set of models, the generalized test accuracies did not differ largely.

Another result to consider when evaluating which model performed best is the average train accuracy. The largest absolute differences between test and train accuracies were observed for the Decision Tree and K-Nearest Neighbours classifiers. This suggests significant over-fitting in these models, which may lead to poor out-of-sample performance. Compared to single tree models, tree ensemble methods generally have less over-fitting (Shalev-Shwartz and Ben-David, 2014), and this is true of the Random Forest and Gradient Boosting models presented in this study, however the reduction is not particularly large. Conversely, the smallest differences between test and train accuracies were observed in the Naive Bayes and Support Vector Machine models, with the downside to the latter approach being its high memory intensiveness.

Based on the above considerations, the Naive Bayes and Support Vector Machine approaches appear to be the most appropriate classifiers for ammonoid taxonomy. Detailed classification results for the Support Vector Machine model suggest that species of the *Cladiscites* genus are most well classified. Despite their data being sourced from multiple publications which differ by region, stratigraphic level, preservation, and authorship, *Owenites koeneni* and *Meekoceras gracilitatis* were among the highest scored species in this model. *Andidiscus behrendseni* was by far the least well-classified species, with precision comparable to the other species but with very low recall, meaning that most specimens identified as *Andidiscus behrendseni* by the model were in fact *Andidiscus behrendseni*, but many *Andidiscus behrendseni* specimens in the data set were not identified.

Table 7. Unsupervised model results.

Method	Predicted number of species (clusters)	Fowlkes-Mallows Score	NMI Score
(Actual)	11	-	-
Mean Shift	9	0.552	0.635
K-Means	5	0.545	0.607
DBSCAN	9	0.522	0.626
OPTICS	9	0.515	0.621
Affinity Propagation	15	0.503	0.614

For these particular fit data, the neural network approach of the Multilayer Perceptron did not offer a performance benefit over other classifiers. The Multilayer Perceptron is however a more dated neural network (Hastie et al., 2009), and a more recent paradigm such as the convolutional neural network popularized by Krizhevsky et al. (2012), which improves upon the Multilayer Perceptron by utilizing regularization, is likely to improve vastly upon the present models for richer sources of data, for example images of ammonoids which include additional features such as ornamentation and suture lines. However, these models are beyond the scope of the present pilot study and these data are unavailable in the PBDB.

The accuracies of the models demonstrated in this study are highly comparable to related studies using other animal properties as features in supervised classifiers. For example, Gunasekaran and Revathy (2010) classified animal species using acoustics with up to 70.3% overall accuracy, Manohar et al. (2016) classified animal species using images with up to 79.54% overall accuracy, and Atanbori et al. (2016) classified birds using motion with up to 66% accuracy. Most relevantly, Meister et al. (2011) achieved similar accuracy ($\geq 73.7\%$) classifying *Fuciniceras* specimens with data from Fourier transforms of ribs.

Supervised models operate on the assumption that the labels (here species) in the training data are true. Therefore, given a robust taxonomy, supervised models such as those in the present study may enable researchers to identify specimens in an unambiguous and objective way. With larger datasets and more features beyond conch proportions, the methods demonstrated here may provide a rigorous quantitative determination of ammonoid taxa.

Unsupervised models

Overall, Fowlkes-Mallows and NMI scores were relatively low across all unsupervised methods. The highest scores were achieved by the Mean Shift and K-Means algorithms respectively. In a dataset consisting of 11 distinct species, the K-Means model estimated five clusters (species) were present, whereas the Mean Shift algorithm predicted nine. Considering the number of clusters, Fowlkes-Mallows score, and NMI score, the most successful unsupervised model applied to the ammonoid data was the Mean Shift algorithm.

While the DBSCAN and OPTICS algorithms also predicted nine species, their scores were amongst the lowest, and a visual inspection of the clusters makes it clear that both algorithms consistently overestimate the range of conch proportions which define individual species (clusters). This can be seen in their subplots in Figure 2, which are dominated in all areas of proportion space by single colours. In contrast, the distribution of colours exhibited by the K-Means, Mean Shift, and Affinity Propagation algorithms resemble the actual distribution much more closely.

Visually, Affinity Propagation had the most success in distinguishing between species such as *Eomaltheus multicostratus* and *Badouxia canadensis*, but was oversensitive to intraspecific variation in species such as *Owenites koeneni*.

Figure 2. Clustering of ammonoid conch proportions from supervised machine learning methods. All subplots have the same axes as Figure 1; The first subplot is identical to Figure 1. The colour maps between subplots are independent.

The results of this study suggest centroid-based supervised clustering algorithms such as K-Means and Mean Shift are best suited to ammonoid taxonomy, while density-based approaches may be less appropriate.

The NMI scores achieved by these models are comparable to extant unsupervised classification literature with similar objectives and number of classes, for example [Clink and Klinck \(2021\)](#) classified 10 primates using acoustics with NMI scores between 0.55 and 0.73, and [Saryan et al. \(2020\)](#) achieved similar scores when classifying 10 plant taxa.

In contrast to supervised models, unsupervised models operate on the assumption that the labels (here species) in the data are unknown. These methods then allow for the construction of a robust and quantitative taxonomy. When applied to a settled taxonomy, unsupervised models quantify the congruence between species definitions and taxonomic attributions of specimens, and their actual morphological clustering. Unsupervised models such as those in the present study, improved with a greater range of

ammonoid features, should be implemented extensively by researchers to suggest improvements upon the current taxonomy.

Strengths

This study's application of supervised and unsupervised methods in machine learning to ammonoid fossils is novel, and the use of nested cross-validation and both internal and external validation metrics is rigorous.

Limitations

The primary limitation of this study is the small sample size. This may result in overfitting if the algorithm is too complex, or underfitting if the algorithm is too simple (Mehryar et al., 2018). Either way, a larger number of ammonoid specimens for training would result in better generalization (Mehryar et al., 2018). In this case, the sample size was limited by the availability of data in the PBDB, therefore alternate data sources are required for better generalization. Under similar circumstances of limited data availability, machine learning approaches with fewer data have been published in recent months (Shen et al., 2020; Jiang et al., 2020). Relatedly, the number of species used is very limited. This is also a consequence of the available PBDB data and the inclusion criteria of a species having at least 50 specimens. Future studies should seek to replicate these findings with richer data sources in both sample size and species counts.

In addition, the conch parameters used as fit data in these models are convenient to obtain and quantify but represent only a small subset of possible features. Future models may utilise the presence and characteristics of shell ornamentation and suture lines for greater statistical power. These were not directly available in the PBDB, though studies in the literature have explored quantitative analyses of suture lines (Manship, 2004; Wegerer et al., 2018).

A possible source of error in supervised models is the assumption that the species attributed to each specimen is itself suitable. It is not possible for all authors in the literature from which data are drawn to conduct their own statistical evaluations of specimens for verification. Indeed, some literature describe qualitative differences between their material and that from other localities. For example, multiple sources were used to provide *Owenites koeneni* and *Meekoceras gracilitatis* data, and these sources differed in region, stratigraphic level, preservation, and authorship which may compromise their utility as a single homogeneous population related by conch proportions. A model trained on unreliable data produces unreliable results. This emphasises a need for future supervised models to be particularly critical of data sources. No such review was conducted for this pilot study (however, in these models *Owenites koeneni* and *Meekoceras gracilitatis* were classified as well as other species with data from singular sources). This also highlights the role of unsupervised models in future research.

Furthermore, in the present study no distinction is made between mature and juvenile specimens, nor males and females. However, the range of conch proportion values across ontogeny and sexual dimorphism may be unique to individual species anyway, and therefore diagnostic of ammonoid taxonomy.

Acknowledgements

The author would like to acknowledge Dr Natasha Vitek (New York) for their inspirational presentation at the Washakie Museum's Paleontology & Archeology Symposium in July 2019, as well as Dr Dieter Korn (Berlin), Kenneth De Baets (Erlangen), Jérémie Bardin (Paris), and an anonymous reviewer for helpful comments on the manuscript.

Additional information

Competing interests

The author declares he has no personal or financial conflict of interest relating to the content of this study.

Data availability

Data are available online at: osf.io/j3q67/

Supplementary information

Code and data are provided as supplementary information at: osf.io/j3q67/

References

- Aguirre-Urreta M (1998). The ammonites *Karakaschiceras* and *Neohoploceras* (Valanginian Neocomitidae) from the Neuquen Basin, West-Central Argentina. *Journal of Paleontology* 72, 39–59.
- Atanbori J, Duan W, Murray J, Appiah K, and Dickinson P (2016). Automatic classification of flying bird species using computer vision techniques. *Pattern Recognition Letters* 81, 53–62. doi: 10.1016/j.patrec.2015.08.015.
- Bardin J, Rouget I, Benzaggagh M, Fürsich FT, and Cecca F (2015). Lower Toarcian (Jurassic) ammonites of the South Riffian ridges (Morocco): systematics and biostratigraphy. *Journal of Systematic Palaeontology* 13, 471–501. doi: 10.1080/14772019.2014.937204.
- Brayard A and Bucher H (2008). *Smithian (Early Triassic) ammonoid faunas from northwestern Guangxi (South China): taxonomy and biochronology*. Vol. 55. Fossils and Strata. Chichester, UK: Wiley-Blackwell. isbn: 978-1-4051-8666-7.
- Brayard A, Bylund K, Jenks J, Stephen D, Olivier N, Escarguel G, Fara E, and Vennin E (2013). Smithian ammonoid faunas from Utah: implications for Early Triassic biostratigraphy, correlation and basinal paleogeography. *Swiss Journal of Palaeontology* 132, 141–219. doi: 10.1007/s13358-013-0058-y.
- Brühwiler T, Bucher H, Ware D, Schneebeli-Hermann E, Hochuli P, Roohi G, Rehman K, and Yaseen A (2012). Smithian (Early Triassic) ammonoids from the Salt Range, Pakistan. *Special Papers in Palaeontology* 88, 1–114.
- Caliński T and Harabasz J (1974). A dendrite method for cluster analysis. *Communications in Statistics* 3, 1–27. doi: 10.1080/03610927408827101.
- Cawley G and Talbot N (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research* 11, 2079–2107.
- Clapham M, Uhen M, Mashall C, Park L, Hopkins M, Lockwood R, Sessa J, Cohen P, Holroyd P, Pálffy J, Hendy A, Lloyd G, Kröger B, Miller A, Erwin D, Crampton J, Purvis A, Holland S, McElwain J, McClennen M, Zaffos A, Peters S, Czaplewski J, Heim N, Syverson V, and Jenkins J (2020). *The Paleobiology Database*. url: <https://paleobiodb.org/>.
- Clink D and Klinck H (2021). Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods in Ecology and Evolution* 12, 328–341. doi: 10.1111/2041-210X.13520.

- Cox BM (1995). *Ammonites: Fossil Focus Guide*. Nottingham, UK: British Geological Survey. isbn: 978-0-85272-257-2.
- Dagis A and Ermakova S (1990). Early Olenekian ammonoids of Siberia. *Trudy Akademiyi Nauk SSSR* 737, 1–112.
- De Baets K, Bert D, Hoffmann R, Monnet C, Yacobucci M, and Klug C (2015). Ammonoid intraspecific variability. In: *Ammonoid Paleobiology: From anatomy to ecology*. Ed. by Klug C, Korn D, De Baets K, Kruta I, and Mapes R. Vol. 43. Topics in Geobiology. Dordrecht: Springer, pp. 359–426. isbn: 978-94-017-9629-3.
- De Baets K, Klug C, Korn D, Bartels C, and Poschmann M (2013a). Emsian Ammonoidea and the age of the Hunsrück Slate (Rhenish Mountains, Western Germany). *Palaeontographica Abteilung A* 299, 1–113. doi: 10.1127/pala/299/2013/1.
- De Baets K, Klug C, and Monnet C (2013b). Intraspecific variability through ontogeny in early ammonoids. *Paleobiology* 39, 75–94.
- De Marez-Oyens F (1933). On *Paralegoceras sundaicum* Haniel and related forms. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Amsterdam* 36, 88–98.
- Fowlkes E and Mallows C (1983). A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association* 78, 553–569. doi: 10.2307/2288117.
- Gamsjäger B (1982). *Systematik und Phylogenie der obertriadischen Cladiscitidae Zittel, 1884 (Ammonoidea)*. 1st ed. Österreichische akademie der wissenschaften, mathematisch-naturwissenschaftliche klasse denkschriften. Springer. isbn: 978-3-211-86499-9.
- Gunasekaran S and Revathy K (2010). Content-based classification and retrieval of wild animal sounds using feature selection algorithm. *2010 Second International Conference on Machine Learning and Computing*, 272–275. doi: 10.1109/ICMLC.2010.11.
- Hastie T, Tibshirani R, and Friedman J (2009). *The elements of statistical learning: Data mining, inference, and prediction*. 2nd ed. Springer Series in Statistics. Springer. isbn: 978-0-387-84858-7.
- Hillebrandt A (2006). Ammoniten aus dem Pliensbachium (Carixium und Domerium) von Südamerika. *Revue de Paléobiologie* 25, 1–403.
- Hohenegger J and Tatzreiter F (1992). Morphometric methods in determination of ammonite species, exemplified through *Balatonites* shells (Middle Triassic). *Journal of Paleontology* 66, 801–816.
- House M (1985). The ammonoid time-scale and ammonoid evolution. *Geological Society, London, Memoirs* 10, 273–283. doi: 10.1144/GSL.MEM.1985.010.01.21.
- Jenks J, Brayard A, Brühwiler T, and Bucher H (2010). New Smithian (Early Triassic) ammonoids from Crittenden Springs, Elko County, Nevada: Implications for taxonomy, biostratigraphy and biogeography. *New Mexico Museum of Natural History and Science Bulletin* 48, 1–41.
- Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, Shi J, Dai J, Cai J, Zhang T, Wu Z, He G, and Huang Y (2020). Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua* 63, 537–551. doi: 10.32604/cmc.2020.010691.
- Kennedy WJ (1984). Systematic palaeontology and stratigraphic distribution of the ammonite faunas of the French Coniacian. *Special Papers in Palaeontology* 31, 1–160.
- Klug C and Hoffmann R (2015). Ammonoid septa and sutures. In: *Ammonoid Paleobiology: From anatomy to ecology*. Ed. by Klug C, Korn D, De Baets K, Kruta I, and Mapes RH. Vol. 43. Dordrecht: Springer, pp. 45–90. isbn: 978-94-017-9630-9.
- Korn D (2010). A key for the description of Palaeozoic ammonoids. *Fossil Record* 13, 5–12. doi: 10.1002/mmng.200900008.
- Krizhevsky A, Sutskever I, and Hinton G (2012). ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th international conference on neural information processing systems - volume 1*. Red Hook, NY, USA: Curran Associates Inc., pp. 1097–1105.

- Kuenzi W (1965). Early Triassic (Scythian) ammonoids from Northeastern Washington. *Journal of Paleontology* 39, 365–378.
- Kummel B and Steele G (1962). Ammonites from the *Meekoceras gracilitatus* zone at Crittenden Spring, Elko County, Nevada. *Journal of Paleontology* 36, 638–703.
- Longridge L, Smith P, and Tipper H (2006). The Early Jurassic ammonite *Badouxia* from British Columbia, Canada. *Palaeontology* 49, 795–816. doi: 10.1111/j.1475-4983.2006.00562.x.
- Manger W and Saunders W (1980). Lower Pennsylvanian (Morrowan) ammonoids from the North American midcontinent. *The Paleontological Society Memoir* 10, 1–56.
- Manohar N, Sharath Kumar Y, and Kumar G (2016). Supervised and unsupervised learning in animal classification. *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 156–161. doi: 10.1109/ICACCI.2016.7732040.
- Manship L (2004). Pattern matching: Classification of ammonitic sutures using GIS. *Palaeontologia Electronica* 7, 7.2.5A.
- Mehryar M, Rostamizadeh A, and Talwalkar A (2018). *Foundations of machine learning*. 2nd ed. Adaptive Computation and Machine Learning. Cambridge, USA: The MIT Press. isbn: 978-0-262-03940-6.
- Meister C, Dommergues J, Dommergues C, Lachkar N, and Hariri K (2011). Les ammonites du Pliensbachien du jebel Bou Rharraf (Haut Atlas oriental, Maroc). *Geobios* 44, 1–60. doi: 10.1016/j.geobios.2010.06.006.
- Monnet C, Brayard A, and Brosse M (2015). Evolutionary trends of Triassic ammonoids. In: *Ammonoid Paleobiology: From macroevolution to paleogeography*. Ed. by Klug C, Korn D, De Baets K, Kruta I, and Mapes RH. Vol. 43. Dordrecht: Springer, pp. 25–50. isbn: 978-94-017-9633-0.
- Popov Y (1962). Nekotorya rannetriasovye ammonoidei Severnogo Kavkaza. *Paleontologicheskii Zhurnal* 3, 40–46.
- Ritterbush K, Hoffmann R, Lukeneder A, and De Baets K (2014). Pelagic palaeoecology: the importance of recent constraints on ammonoid palaeobiology and life history. *Journal of Zoology* 292, 229–241. doi: <https://doi.org/10.1111/jzo.12118>.
- Saryan P, Gupta S, and Gowda V (2020). Species complex delimitations in the genus *Hedychium*: A machine learning approach for cluster discovery. *Applications in Plant Sciences* 8, e11377. doi: 10.1002/aps.11377.
- Shalev-Shwartz S and Ben-David S (2014). *Understanding machine learning: From theory to algorithms*. Cambridge, UK: Cambridge University Press. isbn: 978-1-107-05713-5.
- Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, Quan S, Zhang F, Sun R, Qian L, Ge W, Liu W, Liang S, Chen H, Zhang Y, Li J, Xu J, He Z, Chen B, Wang J, Yan H, Zheng Y, Wang D, Zhu J, Kong Z, Kang Z, Liang X, Ding X, Ruan G, Xiang N, Cai X, Gao H, Li L, Li S, Xiao Q, Lu T, Zhu Y, Liu H, Chen H, and Guo T (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182, 59–72. doi: 10.1016/j.cell.2020.05.032.
- Shevyrev A (1990). Ammonoidei i khronostratigrafiya Triasa. *Akademiya Nauk SSSR, Trudy Paleontologicheskogo Instituta* 241, 1–179.
- Shevyrev A (1995). Triassic ammonoids of the northwest Caucasus. *Rossiyskaya Akademiya Nauk, Trudy Paleontologicheskogo Instituta* 264, 1–174.
- Shigeta Y and Nguyen H (2014). Cephalopods. In: *Olenekian (Early Triassic) stratigraphy and fossil assemblages in northeastern Vietnam*. Ed. by Shigeta Y, Komatsu T, Maekawa T, and Dang HT. Vol. 45. National Museum of Nature and Science Monographs. Tokyo, Japan: National Museum of Nature and Science, pp. 65–167. isbn: 978-4-87803-038-3.
- Strehl A and Ghosh J (2003). Cluster ensembles — a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research* 3, 583–617. doi: 10.1162/153244303321897735.

- Tozer E (1961). *Triassic stratigraphy and faunas, Queen Elizabeth Islands, Arctic Archipelago*. Vol. 316. Geological Survey of Canada Memoir. Ottawa, Canada: Ottawa Dept. of Mines and Technical Surveys.
- Walker C and Ward D (2002). *Fossils*. 2nd ed. Smithsonian Handbooks. London, UK: Dorling Kindersley, Inc. isbn: 978-0-86318-921-0.
- Wegerer M, De Baets K, and Korn D (2018). Quantitative analysis of suture lines in Carboniferous ammonoids. *Fossil Record* 21, 223–236. doi: 10.5194/fr-21-223-2018.
- Wiedmann J and Kullmann J (1980). Ammonoid sutures in ontogeny and phylogeny. In: *The Ammonoidea*. Ed. by House M and Senior J. Vol. 18. Systematics Association Special Volume. London and New York: Academic Press, pp. 215–255.