

Point by point responses to the reviewers:

David Lazarus (Referee)

david.lazarus@mfn.berlin

Received and published: 21 July 2020

General comments

Comment: As the authors note, human collection of routine occurrence data for radiolarians or other organisms is time consuming, requires rare, expensive expert workers, and suffers from inconsistencies between data collectors. Automated collection of occurrence data is likely to prove revolutionary to fields such as micropaleontology, where the vast numbers of specimens and species preserved in the fossil record means that the quantity and quality of the data can be expanded by orders of magnitude. This is a rapidly developing subject and papers are appearing in quick succession. Most work in micropaleontology however so far has been on the most intensely studied groups - pollen, planktonic foraminifera and coccolithophores. Only a very few studies have been done on radiolarians, despite their importance to carbon cycling, polar biogeochronology and evolution research. This ms is, so far as I know, the first attempt to essentially throw an AI system at an entire assemblage without a human pre-selection process of the input images. It is a very valuable contribution to see how well this works, and where bottlenecks or barriers arise.

Response: We thank the reviewer for all the constructive comments about the MS, the aim is to propose, on the long term, a network that would be able to identify most of the common species worldwide, and on longer timescale. Overall, to briefly sum up the corrections made : we improved the training set by adding more than 4000 images, and fine tuning the taxonomy according to the comments of the reviewer, and we trained a new network that uses this new training set.

Comment: One very important advantage of this as a fully automated system, including primary image acquisition, is that, having identified common categories and trained the system to identify them, future image acquisition can concentrate on the less common categories, resulting in a great improvement in the efficiency of the taxonomic specialist who should be examining and tagging only unknown/poorly documented-rare forms. They might wish to point this out more clearly in the ms.

Response: Indeed, focusing on rare species is difficult, as only few specimen can be found in the entire core. To this aim, images of rare species will be progressively added to the network when more samples / cores will be identified. Thank to this system, a taxonomist expert can focus on analyzing rare species, this is added to the conclusions.

Comment: They post their data and code, or pre-prints very openly - kudos. As Marchant et al. submitted gives the technical component of the AI system I will not comment further here on it. The ms, and the study itself, are well done and the ms is worthy of publication with only moderate revisions. Nonetheless there are several things that should be improved prior to acceptance.

Response: One of our goal that to make as much data and code open so that most people can use it quickly in their lab.

Critique

Citation

Comment: The ms does not cite some other prior attempts to automate radiolarian identification, should try to cite some of these since there are so few of them: e.g. Apostol, L. A., Ma I  rquez, E., Gasmen, P., and Solano, G. (2016); Ke  eli, A. S., Kaya, A., and Ke  eli, S. U. (2017).

Response: These missing references were added to the MS lines 64 to 70.

Images and Taxonomy

Comment: Stacked glassy images are used in this project as the basis for identification. These are arguably the best single image to use if a study is done with just one image per specimen (not actually a requirement for this type of work). However, stacked images often do not show important interior characters (for example in actinommids, pylonids and plagiacanthids, and in many instances in other families as well). These interior characters are important for species identification in many taxa and I found the images sometimes to be frustrating to interpret as an expert in the taxonomy of these forms. The standard method of imaging and presentation in research for this material is to show a manually focussed set of a few unstacked images in transmitted light. Possibly there is a way to use the raw images better? Also, precisely because such stacked images are uncommon in the field it is not clear how easy it will be for community members to add to their open image database, despite the laudable call for contributions. Very few researchers for that matter have access to the, for micropaleontologic standards, complex, expensive equipment used by the authors in this study, tho perhaps only manually generated images are needed as input.

Response: We are aware of the issue that the internal features are usually not visible on our specimens. This is why some grouping of species, or taxa identified at a higher rank (e.g. pylonoid spp) were created to prevent “too optimistic” identification. As pointed out by the reviewer, when a single image can be used, as in this study, stacked images are the best choice, for routine and unsupervised analyses. As the stacking occurs directly on the FOV, and not on the vignettes, it is not easy, to directly work on unstacked vignette in our workflow. Moreover, using several images (at least one focused on the centre of the shell, and one on the outside), for the identification of a single specimen is not something we can achieve at the moment. Regarding the use of unstacked image with our database, very encouraging results were produced (see last paragraph of our discussion). To address this issue properly, even if stacking is a more and more used technique in micropalaeontology, we encourage scientist with a large set of unstacked image to share them, in order to integrate them in our network, or create another specific network dedicated to unstacked images.

Comment: The images are only ca 300 pixels in size. This is a bit marginal. While sufficient for most features, taxonomically important small features can get lost (bladed vs cylindrical spines, etc). Computation costs increase with image size but if possible I would suggest nearly doubling the image resolution for future work, particularly if the database used is going to be promoted as a standard for future contributions.

Response: We agree that some features cannot be unambiguously resolved using the 256x256 pixels image size. We did run some tests using bigger vignette images size, but on our server when images are larger than 320 px in size, the server crashes (even with 26 go of allocated RAM space). To test if bigger images can produce significantly better results, the database was reduced to about 13k images by removing some images of classes that contain more than 300 images. This way, we could increase the images size to 384 px without crash. However, results were not better, and in fact, worse, as the overall accuracy drop to 87 %. This could be due to small images being resized, generating pixellisation that was learn as a feature and creating confusion between classes. Other testing were conducting by adding images to the database and resizing the dataset to 256 px to compensate for the increased RAM needed. Results were better than the original one presented in the first version of the manuscript were achieved, and are integrated in the revision of the manuscript to show as updated data as possible. Moreover, it has to be noted that small features that can be taxonomically important (such a bladed vs cylindrical spines) are likely to be difficultly learned by the network while the images quality, and intraspecific morphological variability would play a most important role, before the network can focus on such small features. We added in the manuscript lines 168- 172 a short discussion on this point.

Comment: The cited 17K image database is significantly smaller in numbers of species images. There are non radiolarian broad categories like 'spicule' - 8 categories, ca 8K images, plus supraspecific spp group categories - nearly 20 named, ca 4K images. The number of tagged species images is thus well under 10K. This distinction should be noted clearly in the ms.

Response: To address this issue, and as new images were generated since the original submission of this manuscript, the database was recently increased up to about 22k images. Numbers were corrected in the text (lines 235 to 243), and database composition is now detailed. The database now contains 116 classes corresponding to species or groups of two to three species and containing 11,126 images, 7 classes corresponding to genera and containing 1,932 images, and 1 corresponding to family and containing 677 images. The 8 non-radiolarians classes contain 8,011 images.

Comment: For many taxa images have multiple image types, e.g. *Acrosphaera spinosa* specimens largely covered with some sort of milky bubble (preparation or image stacking artefact?) - 20 of 60 images. This problem is fairly common, seen in many folders. This particular image problem does not appear in routine sample preparation of similar materials and should be explained, as also the effect on identification accuracy.

Response: This issue is due to the glu' s viscosity that might prevents air bubbles to escape from perforated type shells, common in collosphaeridae. Although not ideal, images of specimens containing bubbles, but still recognizable were kept into the database to integrate this variability into the neural network and prevent misidentification with class "bubble" .

Comment: There are many, mostly minor issues with the taxonomic image database, tho some of these do not appear in the ms itself due to dropping rare image categories in the analysis.

Response: As the database is open-access and available online, it can be corrected when errors are reported. It was recently corrected for some images mis-identification, and classes names, also according to errors reporting below. We accept any suggestion to improve the taxonomical framework of the database.

Comment: There are really too many incorrect names being used in the ms and supporting image database. While not a problem for testing (the name tags as such have no effect on the system) these should be corrected. I have not gone through all 100 folders but based on a sampling of these I note the following. 'Strichocorys spp' - no such genus, content appears to be *Phormostichoartus pitomorphus*, *Theocalyptra davisiana* - correct name (since 30 years) is *Cycladophora davisiana*, *Calyptra cervus* instead of intended? name *Corocalyptra cervus*, tho recommended name is *Eucecryphalus cervus* - and see also below, *Pseudodictyophimus gracilipes*, *kilmari*, not *Dictyophimus* (again, since decades) - which importantly changes the family assignment.

Response: We acknowledge our mistakes : numerous names were corrected and new images and species were added to the database. The corrected names / taxonomy is now visible on the online database (autoradio.cerege.fr), on the downloadable file, and in the manuscript / figures. For instance, "Strichocorys spp" (error with an extra r) was corrected to "Phormostichoartus pitomorphus" , *Theocalyptra davisiana* was corrected to *Cycladophora davisiana*, *Calyptra cervus* was corrected to *Corocalyptra cervus*, *Dictyophimus gracilipes* to *Pseudodictyophimus gracilipes*, and other mistakes after a careful check on Lazarus et al., on WoRMS, and mikrotax.org, for a consistent and updated taxonomy.

Comment: There are also some typos such as *Stylotractus* instead of *Stylatractus* universus. Lastly there is at least one instance of possible oversplitting - i. e. *Dictyocoryne trun- catum* vs *Euchitonia triangulum*. What is the difference?

Response: Typos were corrected (all species of *Stylatractus*, *Stichocorys*, and others). With regards to oversplitting, *Dictyocoryne truncatum* was fused in the new CNN with *E. triangulum*, and even with *D. profunda* as both might be synonyms (as suggested by Boltovskoy) and exhibit a large inter and intra specific morphological variability (even more when specimens appear broken).

Comment: There are also a certain number of specimens in the individual taxon folders that are not con-specific. A brief sampling yields:

image name given correct name total images in category 14875 *Cyrtopera langucula* *Artopilium undulatum* 6

Response: the problematic image was moved to a new folder "Artopilium undulatum" .

11233 *Anthocyrtdium ophirensense* unknown but pores much too large to be conspecific 20

Response: This image was removed from the database as not identifiable.

01235 *Zygocircus productus* *Zygocircus piscicaudatus* 67

Response: As we see a lot of morphological gradation between both species in our specimens, the class is now called *Zygocircus piscicaudatus productus* and contain both species. A new class was created and contains *Zygocircus capulosus* specimens found in Miocene sample.

13329 *Amphisphaera* sp. B appears conspecific w. some specimens labeled as *Drup-*

patractus irregularis 21/28

Response: The whole bi-spicular actinommidae group was updated and corrected.

mult. *Calyptra cervus* multiple *Cycladophora* species including [u1486-...] *Cy- cladophora cabrilloensis*

Response: A class *Cycladophora cabrilloensis* was added.

Comment: The large majority of the identifications (so far as the image quality allows) appear to be correct as monospecific classes, even if sometimes the name for the class are not. An attempt to provide a standard set of names (provisional of course as taxonomy is always being revised) was given by a group of taxonomists in Lazarus et al. 2015. I suggest for the sake of data standardization that they use these, or at least provide a taxonomic appendix where they explain any variant usage.

Response: Classes names were corrected as a standardization effort by using Lazarus et al., 2015 and WoRMS.

Performance measurements

Comment: The time has come, the Walrus said, To talk of many things: Of shoes and ships and sealing-wax, Of cabbages and kings (Lewis Carroll, Through the Looking Glass, 1871). Radiolarians are to a highly unusual degree morphologically diverse. There are at least a dozen topologically highly distinct Baupläne alone in living/late Neogene assemblages. Distinctions between these broad morphologic groups are trivially easy. This is quite in contrast to most other clades of organisms, and in particular to planktonic foraminifera, which were used by this research group to initially develop their algorithms and work flows. Planktonic foraminifera are morphologically very conservative (once described to me by a foram specialist- as 'basically just popcorn'), and can be considered for an imaging system as a single broader category for analysis purposes. For radiolarians, it is really not very informative to know that the system can distinguish between forms as radically different as, well, cabbages and kings. The true test of performance is its capacity to accurately identify and distinguish between species within topologic/taxonomically similar groupings, such as within radiolarian families/subfamilies. This is not only for purely taxonomic reasons, but also for the utility of a system in applied research. Radiolarians encode ecologic-environmental information almost entirely at the level of individual species. Attempts to use genus or family level taxa as proxies in applied paleoenvironmental research have yielded almost no useful signals. Geologic age in the Cenozoic is partially recoverable at the genus level, but at a resolution so poor as to make it uninteresting for actual use.

Response: First of all, we have to consider that the neural network that we trained, and overall, the whole workflow is a compromise between distinguishing as many species as possible, and try to keep a good accuracy for each class which mostly depend on the growing number of images in each of them. The more images will be progressively added to the database, the more accurate will be the identification, and the more we will be able to go to the specific level.

Comment: The ms therefore needs to clearly separate the performance of the system in distinguishing between morphologically/taxonomically similar forms and the ability to distinguish extremely dissimilar forms. I suggest in both the statistical analyses of output and in the creation/ organisation of the figures, and the image database for download, that a clustering to higher categories is done, e.g. radiolarian orders and families, plus 'other' for non radiolarian categories such as particles, diatoms or background. The results for pairings of related taxa such as *Lophophaena hispida* and *Peromelissa phalacra*, both within the lophophaenid subfamily of Plagiacanthidae (performance values ca 70%) suggest that the accuracy for

applied uses may be significantly lower than the current bulk statistics suggest. It is also important to report separately performance in identifying radiolarian species vs identifying broad categories such as 'spp.', diatoms, particles etc.

Response: To address this issue, and get a better idea of the system performance to distinguish between similar and dissimilar forms, the accuracy was also computed within each family (see new Fig. 4). This accuracy is computed by taking the accuracy and number of test specimens for each class into account. We can see that this accuracy is about 91% for each family (e.g. 85% for the plagiacanthidae). A short discussion was added lines 287 to 299. Regarding the clustering to a higher categories, as suggested, Fig. 4 was emended to add the accuracy at the family level. The website for the online catalogue (autoradio.cerege.fr) also shows an order / family / genus / species organization. However, the downloadable database was left with the original ranking, as it is directly used for the training step, which cannot handle a complex ranking system (all class need to share the same rank level).

Comment: Lastly, when errors Are made, the nature of these is significant. It is hard to understand how this system could mis-identify a *Pseudodictyophimus gracilipes* [incorrectly named *Dictyophimus gracilipes* in the ms] with *Hexacontium* spp. - these taxa are in different orders, and have [to a taxonomic expert] fully different morphologies. Some sort of statistic giving not just the error rate, but the type of error - misclassified as to species in same family, different family in order, or different order should be given.

Response: We implemented in the training of the CNN (Marchant et al, in press) a workflow to extract the closest images based on the CNN vector loading, to identify likely misclassified images in the training data set. This is useful for checking bad manual identification, to groundchek the neural network training. However, as a figure is generated for each image tagged as " misclassified ", considering the size of the database (currently more than 21K images), the sum of all the generated figures is heavy and was not integrated in supplementary material. Although it might be confusing to understand why a *P. gracilipes* which is a nasselaria, might be confused with a *Hexacontium* spp, which is a spumellaria, it is usually not a question of how much are two species related, but how similar are 2 black and white tiny images showing strange forms. From a far point of view, *P. gracilipes* is basically a somewhat rounded form exhibiting 3 to 4 spiny extension, depending on the point of view, which is also the case for *Hexacontium* spp when some spines are broken. Although not taxonomically closely related. If the *P. gracilipes* class only show 3 to 4 spined specimens, and *Hexacontium* 5 to 6 spined forms, depending on the view, a broken specimen of *Hexacontium* spp exhibiting 3 to 4 spines is likely to be classified in *P. gracilipes*. This is why including a large intraspecific morphological variability, with slightly broken specimens, bubble, and so on is important, as the majority of shells are not perfectly preserved in sediment.

Completeness of taxonomic/morphologic coverage

Comment: The authors have clearly made an effort to look at the entire assemblage of radiolarians, which is perhaps the most distinctive, laudable and novel aspect of this study. It must however be noted that the number of actual examined species is less than 80, while radiolarian diversity in tropics-subtropics is ca 500 (just in the sediments, not counting those, relatively few, species in the plankton that do not preserve).

Response: We add new images (about 5000) regarding rare species that were poorly images before, and new species not present in the samples previously investigated, were added to the database, as more samples were processed since the initial submission. The number of radiolarian classes is now up to 124 (with 101 classes with more than 10 images corresponding to the minimal number of images used in the training of the CNN).

Comment: Thus only about 15% of the diversity in these assemblages has been incorporated into the study. Adding a larger percentage of the species is a stated goal of their project and quite correct. However, as more species are added, the number of closely related species pairings will increase, in relation to comparisons between distantly related forms. This is likely to have a negative effect on the performance of the system, as accuracy in similar pairings appears to be fairly low at present. This may not affect using the system for classic assemblage based proxies of paleoenvironmental conditions as these can be based on relatively few, selected species or even species groups. There is not much demand however at the for this type of work, as it has been largely displaced by geochemical methods. Possibly having a cheap system to generate the data will revive it, but I am somewhat skeptical. Biostratigraphy however remains important, along with a variety of emerging themes related to evolution and biodiversity. These studies though need to use a larger fraction of the assemblage, distinguish closely related forms, and/or include rarer species. (Indeed it seems to be an instance of Murphy's Law that important biostratigraphic markers are so hard to find in many slides...). The usability of this system will only become apparent when it has grown to include more taxa, including many closely related and rare forms. The ms should make these limitations clear.

There is no general answer to what level of accuracy is 'adequate', but I would suggest that for many biostratigraphic and biodiversity studies error rates should be closer to 1% than 10%, which may prove quite challenging for AI systems to achieve. Alternatively analytic methods in these fields will have to be revised to handle much noisier data.

Response: We acknowledge that our study does not encompass all the diversity of modern radiolarians, but is a proof of concept study based on a real test case. We are continuously trying to improve the system, so it can be used in a variety of studies, including its ability to distinguish very similar species for biostratigraphy or evolutionary studies. For example, more Pliocene and Miocene images were added to the database and which now exhibits new species of *Dydimocyrtis* and closely related *Diartus* species. In the original neural network from the first submission, *Dydimocyrtis tetrathalamus tetrathalamus* was recognized with an accuracy of 91%. In the new version of the neural network that was recently trained with more images, and corrected names, *Dydimocyrtis tetrathalamus tetrathalamus* is still recognized with an accuracy of 91%, while added species groups *Dydimocyrtis antepenultima penultima* is recognized with an accuracy of 96%, and *Diartus hughesi petterssoni* with 83%). More species are thus not likely to decrease the accuracy of the network, is enough images are present in each class. In the same way, we are confident that with more and more images, classes composed of 2 or 3 species, and included all the species of a genus will be progressively divided in several distinct classes.

Comment: There is an additional issue in the adequacy of counting only a few hundred individuals to represent an assemblage of radiolarians. This is adequate only for the rather small fraction of species (usually <10% of the species richness) that are present at several percent abundance in the assemblage, as a closer reading of the short paper the authors cite (Fatela and Taborda) would reveal. While a cut-off of several percent is indeed frequently used in paleoenvironmental proxy studies, this is a seriously inadequate degree of coverage for either biostratigraphy or evolution/biodiversity research. Indeed, this problem is indirectly illustrated in the inadequate numbers of specimens for many species in their training-test image sets. There is in fact a very large body of sophisticated literature in ecology on determining the adequacy of sample sizes for different degrees of assemblage diversity and desired completeness of the resulting sample. See as starting points Chao et al. (2020) *Ecol. Res.*; Dornelas et al. (2012) *Proc. R.Soc. B*, and the brief discussion related to radiolarian assemblages in Lazarus et al. (2018) *PeerJ*. For radiolarians, in many cases the appropriate sample size is several thousand specimens.

Response: Depending on the goal and accuracy of the study, this issue can be easily addressed in the sample preparation by pouring a solution of the same sample in the 8 tanks of the decanter (or more or less tanks according to the abundance of radiolarian in the sediment). This way, no changes are required for the image acquisition part of the workflow. Several versions of the decanter for bigger cover slides (32x24 mm and 40x22 mm as seen in some papers) were added to the download platform, and then required a change in the acquisition part of the workflow (the acquisition software was developed so you can directly enter the number and size of cover slide used, so no change in the code is required). New versions of the decanter for other cover slide sizes can also be generated on demand. All versions are available at: <https://github.com/microfossil/Decanter>

Sample coverage

Comment: Detailed information is given for this in the SOM, but essentially all material is from a single location in the western equatorial Pacific. I miss a discussion, or at least a disclaimer, of how geographic variation in morphology, or variation over time in lineages might affect the system's performance by blurring between species distinctions. It would also be nice to know in more detail the ages of the samples and the sources of the age information.

Response: More information was added about cores location. More material was added as more images were acquired since the initial submission. This material originates from other cores that cover a larger area in the WPWP. A discussion about variation in morphology and morphological and assemblages variation over time in lineages will be proposed in the following study as these parameters will be measured and are already observed in our data with consistent and good results (e.g. *Dydimocyrtis tetrathalamus tetrathalamus* with an accuracy of 91%, *Dydimocyrtis antepenultima penultima* with 96%, and *Diartus hughesi petterssoni* with 83%), again with the aim of increasing the accuracy and distinguishing radiolarians species group in groups of species, or genera, as more images are added to the database.

New sample preparation method

Comment: In this section a new variant of a coverslip holder is described. Although the goal of making slides with very few individuals seems to me to somewhat quixotic given the sample adequacy issues mentioned above, the idea of a custom designed holder that can be manufactured by 3-D printing is novel and is, adapted perhaps to full size cover slips, a useful

addition to the literature. The chemical and other preparation steps are fairly standard and, though important to include, could be moved to the SOM.

Response: As mentioned above, new versions of the decanter of different sizes are available at: <https://github.com/microfossil/Decanter>

Figures and Tables

Comment: The confusion matrix (fig. 4) is useful but a much more readable table listing each species, numbers/percents correctly and incorrectly classified and the top 3 error categories they were assigned to would be very helpful. I spent too much time scrolling figure 4 around my screen.

Response: As the confusion matrix is always used in deep learning studies, we would like to keep it if possible. We emended it to make it more easy to read, and add the family accuracy scores. To address the issue, the original excel file with a fixed class names column that enable a more efficient scrolling is provided as a supplementary material (Appendix B). As suggested, the % accuracy for each class, number of images in the test set, and top 3 error categories was also added to this supplementary material in a second excel file (Appendix C).

Comment: I think the citation to Lazarus et al. 2015 line 23 should be Lazarus 2005 while line 235 should be to Lazarus et al. 2015 - they have inverted these.

Response: Citations were corrected, thanks.

The original Figure 5 and associated paragraph were removed as they were originally used to show the accuracy of the network and the recall scores for each class, but we believe that the new Figure 5 (update of the original figure 6) and the updated confusion matrix (Figure 4) are more usual and efficient ways to show the accuracy of the network and individual class scores.

Thore Friesenhagen (Referee)

thore.friesenhagen@unibas.ch

Received and published: 14 August 2020 General comments

Comment: The development of an automated system for the collection of radiolarian census-data using a neuronal network is a consequent step to give over time-consuming workflows to machines. Posting the codes as well as the organisation of a discretionary image-based radiolarian (training) dataset are a good practice, but it also means that maintaining the dataset will be one of the most important tasks for the future. The manuscript is well done and requires only minor revisions. The following annotations and questions should be considered and/or answered in the final publication.

Response: We thank the reviewer for its comments about the MS, a detailed answer to each of its raised point is given below.

Comment: I miss a short introduction about neuronal networks and the “k-nearest neighbors” algorithm for readers who are not familiar with these terms.

Response: A few sentences about CNNs were added to the introduction. However, as everything including the CNN principles, architecture, algorithms, software, etc, is detailed in Marchant et al. (in press), we did not repeat these information here.

Comment: As mentioned in the script's introduction, one and the same specimen may be referred to different species/classes depending on the experience, subjective interpretation and/or taxonomic “education” of the researcher (e.g. Fenton et al. (2018) for planktonic foraminifera). Thus, to reduce the number of possibly mistakenly identified specimens in the training dataset, having at least one more taxonomic expert checking the correctness of the species determination of specimens within the dataset could increase the reliability of the dataset.

Response: Overall, three expert reviewed the database: Martin Tetard, based on numerous and consistent publications, Giuseppe Cortese, and the first reviewer, David Lazarus, that point out numerous correction on the database.

Comment: Does the transparency of the radiolarian shells produce any problems for the image stacking, especially in case of smaller and more delicate specimens? Figure 3j) shows a specimen of the species *Collosphaera tuberosa*. Its contours are diffuse. Is this a common “problem” for this species? Does this affect the identification accuracy for this species and may be one reason for the relatively high value of confusion with *Solenosphaera zanguebarica*?

Response: As the shells are outlined by the different refractive indices between the shell and the mounting medium, we don't experience any issue with the stacking step, even with small and / or delicate shells. Lots of time was spent for finding the best parameters with regard to the stacking method on Helicon Focus. The diffuse contours of the specimen of *Collosphaera tuberosa* is due to its position on the slide. This specimen was slightly out of focus from the 1500 μm range imaged using the automated stacking technique. However, this specimens is the best preserved we have for now, explaining why it was chosen. This problem is not common but on some slides, it may happen that some specimens are out of focus from the stacking range. This does not usually affect the identification. Most of the confusion between *C. tuberosa* and *S. zanguebarica* was due to the few number of specimens that was imaged in both classes. With more images, these two classes are no longer confused with each other in the new version of the CNN.

Comment: The collection of census data for planktonic foraminifera avoid juvenile specimens (e.g. Davis et al. (2019) only investigated the $>125\mu\text{m}$ fraction), because their identification is often very difficult (Fenton et al., 2018). Is there a lower size limitation for radiolarian specimens to be detected and identified by the new system? Is the system able to distinguish between early ontogenetic stages and broken specimens? Does the size of specimens affect the accuracy of the automated species determination?

Response: We only work on the fraction $> 50 \mu\text{m}$ to avoid lots of juvenile specimens that are often difficult to identify as early features are shared between numerous species, and lots of broken shells. However, the ontogenetic stages are

visible in numerous classes and, when they are sufficiently imaged, can be distinguished. The size of specimens do not seems to affect the accuracy of the system as most of the images are above the images size (256 px) used to train the CNN.

Comment: What is the procedure for (intact) specimens which extend over the borders of the 324 FOV and are parted/bisected? Is the program able to identify these specimens as being intact? In this case, are these specimens prevented from being “double-counted” by the system?

Response: For specimens was are “cut” between two FOVs, if a part of the shell contains the first chambers (usually for nassellaria) it should be identified as the correct class, and the second part should be identified as “broken” , to prevent a double identification in the correct class.

Comment: Closely related species tend to show a similar morphology and are often only distinguishable by details. Since the sample preparation bases on random settling, the orientation of a single specimens may not be ideal to enable the program to recognise these morphological details. What is the procedure for specimens which do not show an ideal orientation for determination?

Response: The intra-specific morphological variability generated by the orientation of specimens in images is mostly covered by the number of images present in each class. However, most of the specimens settled the same way for each specimen of a class. Also, if a specimen is oriented in a way that its identification is not possible (e.g. the under view of a nassellaria, which is very rare), it is identified as “broken” .

Comment: I give the authors credit for implementing morphometric measurements. In combination with census data they may provide additional and valuable information for palaeoenvironmental reconstructions and evolutionary studies. Although this paper clearly focuses on the collection of census data, the accuracy of the morphometric measurements should be given as well. To what extend do differences in specimen orientation affect the accuracy and intraspecific comparability?

Response: The morphometric measurements of every specimens, averaged for each class, will be presented for two cores in a next study that is in preparation. As the morphometric measurements are performed on the outline of each specimens, we are confident regarding their accuracy. Regarding the bias generated by the specimens’ orientation, as specimens usually fall on the same way for every specimen of each class, we are confident that most of the observed variability is due to actual change in shape, and not to change in their orientation between samples.

Comment:

-l. 62-63: The sentence contains two times the phrase “promising results”.

-l. 86: A comma is missing after “6.3ka”. “[. . .] 3-4cm, (6.3ka[,] de Garidel-Thoron et al., 2005) [. . .]”

-l. 273: there is a closing bracket at the end of the sentence, but I could not figure out the corresponding, opening counterpart. “[. . .]and that 150 images represent the C3 original dataset for this class; Fig. 5 green square). [. . .]”

-l. 359: The semicolon may be replaced by a closing bracket. “[. . .] palaeoenvironmental proxies such as SSTs (e.g., radiolarian-based palaeotemperatures for [. . .], Kamikuri, 2017;[]) and paleoproductivity [. . .]”

-Fig. 2: A space is missing in the text for step 7. “7.[]Identification of every single particle using a trained CNN.”

-Fig. 4: The printed version is difficult to read, because the font size of the species names is relatively small. The digital figure requires a lot of scrolling.

-Fig. 5: Several names of species overlap and make it impossible to read them.

-Fig. 6 e,f: The percentage numbers are difficult to read, because they overlap with black bars within the figure.

Response:

-l. 62-63: The was corrected and the double word was removed.

-l. 86: The comma was added.

-l. 273: The closing bracket was removed.

-l. 359: The semicolon was replaced by a closing bracket.

-Fig. 2: A space was added.

-Fig. 4: This figure was revised and the species is now slightly bigger. To help with the reading of this figure, an excel spreadsheet was added as a supplementary material, with a fixed first column (classes name) that help with the scrolling.

-Fig. 5: This figure was removed as discussed in the reviewer 1’ s response to comment.

-Fig. 6 e,f: This was corrected.

Thank you for all your comments.

Anonymous referee.

Received and published: 18 August 2020

Comment: Structure/composition of the Paper: While the paper looks overwhelming due to the number of pages, the content is actually concise. The information written there is not too long nor short. The structure of the paper follows the usual format (introduction, methodology, results and discussion).

Response: Thank you very much for your review. Indeed, lots of things are discussed in the manuscript, and it is difficult to reduce it too much, but we try to keep things concise.

Comment: Methodology/approach to the Problem: Their workflow is a whole and complete system, which starts at image acquisition and ends at classification. The workflow hopes to ease the tedious task of identifying specimen from samples, which requires extensive and consistent taxonomic knowledge of the observer to correctly identify radiolarians. Research was done well, as it can be seen that they have explained the steps in great detail (including the measurements).

Response: We aimed at developing a complete system for radiolarian research by using the expertise developed at CEREGE for several decades now. Everything was detailed as much as possible to enable scientists from other laboratories to use this method for their research.

Comment: As for the AI specific topic, they have used a usual Deep Learning approach. They have used ResNet and finetune it on their dataset. They have also included non-Radiolarian classes, which I believe provided an edge especially since they will be classifying things straight from the image acquisition (w/o humans to remove the non-Radiolarian particles). They have acquired a lot of samples, so they did not struggle that much on this part. Overall what I can see here is that the acquisition and segmentation of images are more tedious than the actual training and classification of Radiolarians.

Response: Indeed, non radiolarian classes were used to enable the system to identify the sediment material straight from the vial. Acquisition and segmentation parts were indeed the more tedious parts due to the complex morphology and composition of radiolarians shells.

Comment: For the purpose of training the Convolutional Neural Network (CNN) for identification of Radiolarians, they developed and released AutoRadio (Automated Radiolarian). To encourage participation and contributions on adding more images to AutoRadio, they provided a very detailed protocol to standardize the way of obtaining images. Even the file for 3D printing Decanter, used to prepare the slides, is provided for everyone to use. The repository for Decanter also includes a video for the modified random settling protocol.

Response: We tried to make this whole new protocol as accessible as possible for future radiolarian studies.

Comment: It is suggested that a section briefly discussing the convolutional neural network model should have been included. The approach fundamentally relies on the model and hence it is necessary to detail how it is applied so as to properly justify the solution for automated identification. As such, the section shall essentially include the following: CNN overview, model architecture, and training approach (transfer learning, loss, etc.).

Response: A small discussion was added in the introduction, as suggested by reviewer 1 and anonymous reviewer. More details about the used CNN are provided in Marchant et al. (in press), so no detail was given in this manuscript.

Comment: A minor concern is that I noticed that the Random Settling Protocol, as discussed starting in line 95 and the video (<https://www.youtube.com/watch?v=veRmKI4rGT0>) differ in the series of steps taken for the preparation of the radiolarian slides. I recognize that some steps are possibly not filmed for brevity, and the difference in steps might suggest that what is

written on paper may not be strictly followed. But the motivated reader who wishes to contribute and follow the protocol may feel confused at first. I also noticed that in the video, the sample taken only amounted to 0.1 mg, but in the protocol the recommended amount is 0.6 mg (line 130, step #15), as it corresponds to the best compromise to ensure that a sufficient number of radiolarian specimens are covered and at the same time the specimens are not crowded and not touching one another, as discussed in subsequent sections that overlaps might affect the ability of the workflow to identify radiolarians. What I thought is that in cases where the amount of samples is limited, taking at most 0.6 mg would be enough. All things taken, the inclusion of the video is very helpful.

Response: The video shows how to use the decanter. The part where samples are chemically prepared is thus not included in the video. The preparation protocol was very slightly emended since the original publication and now matches the protocol that was actually used to prepare 400 samples for the next study and that is visible in the video. Regarding the amount of sample taken in the video (0.1 mg), it was an annotation mistake that was corrected. It is indeed recommended to use between 0.6 and 1.0 (not 0.1) mg of material. This was corrected in the text and in the video.

Comment: Another concern is about imbalance in classes, which is actually common among Radiolarian studies. Reading on the documentation of ParticleTrieur, the recommended number of images per class is 50 at minimum and preferably at least 200 images per class, which can be very difficult to achieve especially on rare radiolarian species. Commonly, data augmentation is performed to address the issue of class imbalance. But augmenting the data has to ensure that variations applied to the image still preserve the class/label after applying transformations. Hence, careful application of augmenting data must be ensured. ParticleTrieur also makes use of weighted loss functions, which is another good way of handling class imbalance.

Response: Indeed, some radiolarian species can be very rare to tricky to found. 200 specimens are recommended per class, 50 is a minimum for accurate results, and here we decided to use classes with at least 10 specimens to train as many classes as possible where more images will be progressively added. This even if these classes are not very accurate, the system can start to recognize them and already help with the identification. The presence of classes with few images do not decrease the accuracy of the overall network and do not affect the other classes.

Comment: I agree that ideally, adding more data on rare species would improve the trained model so paper emphasizing the possibility of collaboration through adding more images to AutoRadio and detailing on how one can contribute is really a good step.

Response: More images will be progressively added to the database as we process samples, to cover the rare and under-represented species. We also encourage people to send us their own pictures to cover the morphological variability that is not due to actual change in the shape of specimens, but to different acquisition settings, material, and so on. With active collaboration, the database should be expanding quickly.

Comment: Discussion of the Results: They have used the usual metrics (Accuracy, Precision, Recall, Confusion Matrix). The results were good since image acquisition and segmentation methodology is already profound, their data is quite large (17k samples total), and they have reported an overall accuracy of 90%.

Response: Again, we thank the reviewer for its comments about the MS.

List of all relevant changes made to the manuscript:

- Incorrect species names were emended according to D. Lazarus' comments.
- The database was increased more images (about 4.000), and a new neural network was trained. All classes accuracies, overall accuracy, precision and recall were emended in the text.
- To answer D. Lazarus' comments on the visibility of FIG.4 and families accuracies, Fig. 4 was emended and now includes family groupings. Appendix B was added to scroll more easily in the confusion matrix.
- To answer D. Lazarus' comments, Appendix C was added to show the misclassified classes for each class.
- Some missing references were added.
- Other versions of the decanter were added to the website.
- Typos were corrected.
- The original Figure 5 was deleted as other was to estimate accuracy were presented such as species and family level in the actual Figure 4, the actual Figure 5, the misclassification in Appendix C.

A NEW AUTOMATED RADIOLARIAN IMAGE ACQUISITION, STACKING, PROCESSING, SEGMENTATION, AND IDENTIFICATION WORKFLOW

Martin Tetard¹, Ross Marchant^{1, 2}, Giuseppe Cortese³, Yves Gally¹, Thibault de Garidel-Thoron¹, and Luc Beaufort¹

¹Aix Marseille Univ, CNRS, IRD, Coll France, INRAE, CEREGE, Aix-en-Provence, France.

²Present address: School of Electrical Engineering and Robotics, Queensland University of Technology, Brisbane, Australia.

³GNS Science, Lower Hutt, New Zealand.

Correspondence: M. Tetard (tetard@cerege.fr)

Abstract. Identification of microfossils is usually done by expert taxonomists and requires time and a significant amount of systematic knowledge developed over many years. These studies require manual identification of numerous specimens in many samples under a microscope, which is very tedious and time consuming. Furthermore, identification may differ between operators, biasing reproducibility. Recent technological advances in image acquisition, processing, and recognition now enable automated procedures for this process, from microscope image acquisition to taxonomic identification.

A new workflow was developed for automated radiolarian image acquisition, stacking, processing, segmentation, and identification. The protocol includes a newly proposed methodology for preparing radiolarian microscopic slides. We mount 8 samples per slide, using a recently developed 3D-printed decanter that enable the random and uniform settling of particles, and minimise the loss of material. Once ready, slides are automatically imaged using a transmitted light microscope. About 4000 specimens per slide (500 per sample) are captured in digital images which include stacking techniques to improve their focus and sharpness. Automated image processing and segmentation is then performed using a custom plugin developed for the *ImageJ* software. Each individual radiolarian image is automatically classified by a convolutional neural network (CNN) trained on a [Neogene to Quaternary](#) radiolarian database (currently ~~17,065~~ 21,746 images, corresponding to ~~112~~ 132 classes) using the software, *ParticleTrieur*.

The trained CNN has an overall accuracy of about 90 %. The whole procedure, including the image acquisition, stacking, processing, segmentation and recognition, is entirely automated via a *LabVIEW* interface, and takes approximately 1 hour per sample. Census data count and classified radiolarian images are then automatically exported and saved. This new workflow paves the way for the analysis of long-term, radiolarian-based palaeoclimatic records from siliceous remains-bearing samples.

Copyright statement. TEXT

The term radiolarians currently refers to the polycystine radiolarian orders Spumellaria and Nassellaria, whose shell is made of opaline silica, relatively well preserved in the fossil record by comparison with the Acantharia and Phaeodaria groups. They are marine micro-organisms whose siliceous shells are found in the sedimentary record since their appearance during the Cambrian period (Boltovskoy, 1999; Lazarus et al., 2015; Suzuki and Not., 2015) (Boltovskoy, 1999; Lazarus et al., 2005; Suzuki and Not., 2015).

25 While they have been originally neglected for a long time for biostratigraphical studies due to several documented cases of recurrent evolution in the overall morphology of some taxa (e.g. Schrock and Twenhofel, 1953; Campbell, 1954; Bjørklund and Goll, 1979), radiolarian taxonomy and stratigraphy have significantly progressed due to Deep Sea Drilling Project (DSDP) studies since 1968 (Sanfilippo et al., 1985) and are currently of major interest. Radiolarians are commonly used in biostratigraphy by documenting the presence / absence of key marker species, as well as in palaeoceanographic reconstructions of past
30 productivity, temperature, and variability of water masses, wherein these approaches rely instead on relative species abundances. For both these approaches, radiolarians are particularly useful in high latitude settings (e.g. the Southern Ocean) where both the preservation and species diversity of calcareous microfossils are very low.

Indeed, radiolarian's delicate siliceous remains have been proved important for decades in micropalaeontological studies focussing on palaeoenvironmental reconstructions from various oceanic areas to investigate primary and export productivity
35 (e.g. Welling et al., 1992; Lazarus, 2002; Abelman and Nimmergut, 2005; Lazarus et al., 2006; Hernández-Almeida et al., 2013; Matsuzaki et al., 2019), sea surface temperature (e.g. Abelman et al., 1999; Lazarus, 2002; Cortese and Abelman, 2002; Lüer et al., 2008; Panitz et al., 2015; Kamikuri, 2017; Hernández-Almeida et al., 2017; Matsuzaki et al., 2019), water masses (e.g. Welling et al., 1992; Kamikuri et al., 2009; Kamikuri, 2017; Hernández-Almeida et al., 2017; Matsuzaki et al., 2019) and oxygenation (e.g. Matsuzaki et al., 2019) across the Cenozoic. At present, radiolarians assemblages are considered
40 to be consistent and valuable micropalaeontological bio-indicators as they are largely distributed in all oceans since their appearance and can be very abundant in sediments (e.g. Sanfilippo et al., 1985; Boltovskoy, 1998; Hernández-Almeida et al., 2017).

However, despite their usefulness for such investigations, radiolarians are not as used as other microfossil groups such as benthic and planktic foraminifera, or nannofossils such as coccolithophorids. Experts on living and fossil radiolarians are
45 relatively few, and some radiolarian species still lack a satisfactory taxonomy, especially for taxa within the order Spumellaria (Riedel, 1967; Sanfilippo et al., 1985). Identification of a substantial and sufficient number of specimens per sample (usually about 300 for reliable assemblage composition estimations Fatela and Taborda, 2002) is very time-consuming and requires a consistent and detailed taxonomic knowledge. Moreover, as it is common for all microfossil groups, and especially true for radiolarians, determination and taxonomy of recovered specimens can be different between studies as it can be biased by the
50 subjective appreciation of the operator, influencing reproducibility of the census counts.

Recent technological advances in image acquisition, processing, and recognition now enable automated procedures, from microscopic slide field-of-view acquisition to taxonomic identification, that can ease radiolarian studies. In the early 1980's, some authors had already proposed to automatically analyse the size and shape of a large number of digitised images of

assemblages of microfossils (Budai et al., 1980), in order to investigate the variability of their morphology and use it as
55 a palaeoenvironmental descriptor. For more than 20 years now, the CEREGE laboratory has been a pioneer in automated
image acquisition and recognition for several microfossil groups. Dollfus and Beaufort (1999) developed a structured multi-
layer fat Neural network for coccolith recognition, that was first applied in 2001 to Late Pleistocene primary productivity
reconstructions (Beaufort et al., 2001). This formed the base for the following Système de Reconnaissance Automatique de
Coccolithes (SYRACO) workflow, that used dynamic neural networks (Beaufort and Dollfus, 2004) and is still operating today.
60 ~~The~~ For the past few years ~~have~~, the field of computer vision has seen the emergence and development of convolutional neural
networks (CNNs), a deep-learning approach that enables the automated classification of large sets of images. Convolutional
neural networks are a class of deep neural networks that consist of an input and an output layer, and multiple convolutional
layers. This architecture is similar to the organisation of the visual cortex in the humain brain.

Several workflows inspired by SYRACO and now using CNNs were successively developed at CEREGE and applied
65 to microfossil taxa (e.g. Marchant et al., accepted; Bourel et al., 2020). ~~For radiolarians,~~ Regarding radiolarians, previous
attempts mainly focussed on the identification step. Apostol et al. (2016) used morphometrical measurements and support
vector machine on four radiolarians species recovered from Triassic sediments. In 2017, Keçeli et al. investigated SEM images
of 27 selected triassic species. Renaudie et al. (2018) recently achieved promising results focussing on the automated iden-
tification of species from the same genus with ~~promising results~~ transmitted light microscope images. They obtain an overall
70 identification accuracy of 73 %, achieved over 16 species from 2 genera, where the morphological difference between species
can be very tricky.

In this paper, we also propose a workflow for the automated identification of radiolarians. Our approach differs in that
we wanted to generate a neural network that could recognise most of the common radiolarian species, rather than those of a
specific genus, in order to investigate their abundance (relative and absolute) and diversity, and thus use them as bio-indicators
75 to reconstruct palaeoenvironmental parameters. It is necessary to obtain a large database of images covering the common
species in order to train the network. Out of the modern living 400 to 500 polycystine species, about 100 are relatively common
(Boltovskoy, 1998), however, they have yet to be imaged to create a database for automated recognition purposes. Some online
Cenozoic radiolarian databases have already existed for a few years (e.g. WoRaD, Boltovskoy et al., 2010; radiolaria.org,
Dolven and Skjerpen, 2006; Radworld, Caulet et al., 2006; see Lazarus et al., 2015 for an extensive review of the existing
80 databases), however these are more directed to creating a catalogue for taxonomic purposes. As such, we created an exhaustive
and participative database specifically for CNN training and automated recognition purposes, called AutoRadio (Automated
Radiolarian database, visible at: <https://autoradio.cerege.fr>). To achieve this goal, a new protocol to obtain standard images for
inclusion in the database was required (square images of individual white specimens on a black background, using a stacking
technique if possible) which was also developed in this study.

2.1 Material

Radiolarian microfossils to be used in this study were extracted from several sediment cores. Core MD97-2140 was retrieved from the centre of the West Pacific Warm Pool (WPWP; latitude: 2°02'N; longitude: 141°46'E) at a water depth of 2547 m during the Marion Dufresne IMAGES III-IPHis cruise in 1997 (Beaufort et al., 1997). This core is currently stored at the CEREGE laboratory, France. The sediments consist of a greyish and compact calcareous nannofossil ooze, also containing abundant radiolarian and foraminiferal faunas (de Garidel-Thoron et al., 2005).

Several samples were chosen to extract siliceous microfossils and thus construct a radiolarian images database. Their depths within the recovered core are: 3-4 cm, (6.3 ka de Garidel-Thoron et al., 2005) (6.3 ka, de Garidel-Thoron et al., 2005), 48-49 cm (11.8 ka), 82-83 cm (16.4 ka), 98-99 cm (18.8 ka), 245-246 cm (38.0 ka), 363-364 cm (53.3 ka), 405-406 cm (63.0 ka), 417-418 cm (67.8 ka), 487-488 cm (77.7 ka), 648-650 cm (120.4 ka), 727-728 cm (141.4 ka). For detail on sample processing and slide preparation, the reader is referred to section 2.2.

A few other Miocene to recent Middle Miocene to Quaternary samples retrieved from the ~~Warm Pool were also used later~~ WPWP were subsequently used to increase the number of rare and absent species in the database. These cores were also taken during the Marion Dufresne IMAGES III-IPHis cruise: Core MD97-2138 (1°25'S, 146°24'E; 1960 mbsl; samples 1760-1761 cm, 2670-2671 cm, 3151-3152 cm); and from IODP Expedition 363 (Rosenthal et al., 2018): Holes U1483A (13°05.24'S, 121°48.25'E; samples from sections 9H-4W, 14H-5W), U1483B (samples 6H-6W, 18H-2W), U1486B (2°22.34'S, 144°36.08'E; samples 3H-3W, 6H-4W, 13H-4W), and U1486C (21H-4W) as well as about 150 samples from Hole U1488A (2°02.59'N, 141°45.29'S; samples E6H-3W to 35F-2W).

2.2 Random Settling Protocol

A new protocol was developed as a proposed standard methodology for preparing radiolarian microscopic slides. It places 8 samples per standard 76x28 mm slide using 12x12 mm cover slides on which radiolarians are randomly and uniformly decanted using a new 3D-printed decanter (Fig. 1a-b). The 3D file for this new decanter was designed online, using the Autodesk, Inc. 3D design platform *Tinkercad* (<https://www.tinkercad.com/>), and is available for free at <https://github.com/microfossil/Decanter>. ~~It was~~ Two other versions of the decanter were designed for standard 32x24 mm and 40x22 mm cover slides, also commonly used in micropalaeontology, and are also available online. Custom-sized decanters can also be designed on demand. Our decanters were printed on a Raise3D fused deposition layer type printer using 1.75 mm R3D Premium PLA filament for a material cost of about 1 euro. Approximately 30 g of filament was used, and 4.5 hours were needed to print the model using a standard resolution layer height of at least 0.20 mm.

A random settling technique was preferred to a standard smear slide preparation as the objective of this study is a detailed quantitative faunal analysis with investigation of the relative abundances of each taxon (Sanfilippo et al., 1985). Indeed, a random settling technique provides a more uniform distribution of the residue resulting in less clumped particles, which are also easier to capture digital images of each specimen. The new decanter minimises the loss of material, and a slide guide

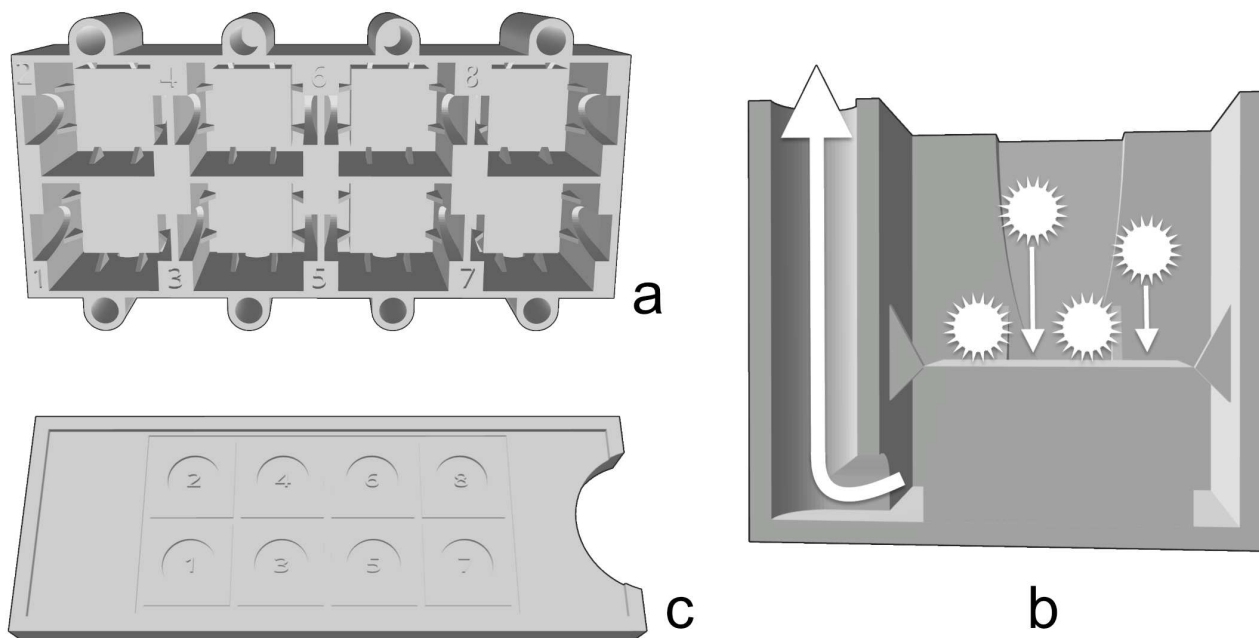


Figure 1. a. Upper view of the new 3D-printed decanter, showing 8 tanks. b. Cross-section of a single tank of the new 3D-printed decanter. c. Upper view of the slide guide.

(Fig. 1c) can also be used to align cover slides. During development, various shapes and sizes of tank were tested, and the one presented herein was the best compromise between the quantity of sample material required, loss of residue that would not settle on the slide, and quantity of microfossil residue recovered. This method is an improved version of the original random settling method developed by (Moore, 1973), adapted to radiolarian studies (Boltovskoy, 1998) that provides an even and random distribution of the shells on a slide, and modified by (Beaufort et al., 2014) to mount up to 8 samples on a single micropalaeontological glass slide. The use of this new device is simple: a 12x12 mm cover slide is placed in the middle of each tank and maintained centred by the fins. A solution containing radiolarians in suspension for each sample is then poured onto each tank and after a few minutes of settling on the cover slides, water is vacuumed out from each hole.

The new radiolarian slide preparation protocol follows the following steps (#2 to 7 have been adapted from a similar procedure used to process limestone and calcareous sediments):

1. Weigh the sediment.
2. Put about 1 g of sediment (depending on the abundance of radiolarians) in a 200 mL beaker and add a few drops of distilled water to disaggregate it.
3. Add a few mL of 37 % Hydrochloric acid (HCl) until the end of the effervescence.

4. Further add a few ~~mL of 10–15 drops of~~ % HCl to ensure the end of the effervescence.
5. Pour the solution and rinse the beaker over a 50 μ m sieve.
6. Clean the residues in the sieve using a pressure sprayer until they appear whitish.
- 135 7. Rinse the residues using distilled water.
8. Weigh a clean glass storage vial.
9. Pour the residues from the sieve to the vial using ~~ethanol~~distilled water.
10. Once the residues have decanted, remove the excess water using a pipette.
11. Place the vial into the oven (about 50 °C) until the residues are dry.
- 140 12. Weigh again the vial to calculate the weight of the recovered residues.
13. Gently tap the vial to unstick the residues from the bottom of the vial.
14. Put a 12x12 mm licked or flame burned cover slide into one tank of the decanter.
15. Take about 0.6 to 1 mg of siliceous residue and drop it onto 3.5 mL of distilled water.
16. Shake this solution to suspend the residue and quickly pour it into the corresponding tank.
- 145 17. Wait until the residues have decanted (few seconds to minutes) and slowly vacuum out the water from the hole (Fig. 1b).
18. Place the decanter in the oven (about 50 °C) to dry the cover slide.
19. When dry, remove the cover slide from the tank using plastic tweezers and glue it to a standard glass slide (76x28 mm) using optical glue (e.g., NOA81, refractive index of 1.56).

Regarding step #2, the reader should take into consideration the fact that the absolute abundance of radiolarians varies
150 massively in sediment samples from various parts of the ocean. The amount of sediment dissolved into HCl should thus be customised according to the expected abundance.

Regarding step #15, 0.6 mg of residue corresponds to the best compromise between having a sufficient number of radiolarian specimens and them touching or overlapping too much, for a 12x12 mm cover slide (see Appendix A). This is desirable as often touching specimens cannot be individually segmented from images, leading to "double" images containing two or more
155 specimens, which cannot be easily classified or assigned to a species count. Distilled water was preferred to ethanol as it leads to less clustering of specimens.

The volume above the cover slide in the tank corresponds to about 45 % of the total volume of the tank. According to the average weight of a radiolarian specimen (about 0.5 μ g Takahashi and Honjo, 1983), 0.6 mg of siliceous residues after

chemical treatment should then contain about 1,200 radiolarians, if not "contaminated" by other siliceous particles, of which
160 about 600 should fall on the cover slide, thus resulting in at least 300 specimens that should be available for identification
(~~minimum required to characterize an assemblage by most of the statistical studies, e.g. Fatela and Taborda, 2002~~)(minimum required to c)
This was confirmed by our tests showing an average of 473 complete identifiable radiolarian specimens per sample, or at least
exhibiting more than 50 % of their shell, including at least the medullary shells for spumellarians, and cephalis and thorax
for nassellarians (excluding specimens touching each other and broken specimens). Other testing found that Norland Optical
165 Adhesives NOA81 glue was preferred to other mounting media such as NOA74 or Naphrax due to its refractive index, con-
sistency and long-term preservation. Although time consuming, metal coating (using C or Au/Pd for example) is also a very
efficient way of increasing contrast prior to mounting specimens on the slides. The darkfield illumination technique was too
inconsistent in the produced images that further tests were not carried out.

2.3 Automated Image acquisition

170 Particular emphasis was placed on acquiring high-quality slide images, as being able to recognise different radiolarian species
depends on having clearly visible features. However, it has to be noted that, no matter the image quality, very small features
that can be taxonomically important (for example bladed vs cylindrical spines) are likely to be difficultly learned by the
network as the morphological variability between every pictures of each class is likely to play a most important role, before
the network can focus on such small features. For each radiolarian microscopic slide, the 8 cover slides (corresponding to
175 8 samples) are automatically and consecutively imaged using a Leica DMR 6000 B automated transmitted light microscope
(200x magnification using a HCX PL FLUOTAR 20 x magnification Leica lens) and a Hamamatsu ORCA Flash4.0 LT camera,
controlled via a *LabVIEW* (National Instruments) interface. The microscope parameters were set as: Intensity: 10; Depth of
field: 38; Aperture: 33; and the condenser was lowered by 9 mm from the glass slide. The *LabVIEW* acquisition software
parameters were set as: Exposure: 9 ms; Gain: 1. These settings provided the maximum contrast between the glass shells and
180 their mounting medium.

For each sample, 324 fields of view (18 x 18 FOVs of 660 x 660 μm each within each 12x12 mm cover glass) were imaged
using a multi-focal technique (Fig. 2). For each FOV, 15 images were acquired by incrementally stepping the Z focus position
through the microscopic slide (step size: 10 μm) to cover a total focal distance of 150 μm , which corresponds to the thickness
of most radiolarian species. This acquisition step takes exactly 1 hour per sample, and thus 8 hours per slide.

185 2.4 Automated Image Processing and Segmentation

Image processing and segmentation is performed via a second *LabVIEW* interface. For each FOV, the batch of 15 images is
automatically stacked using *Helicon Focus 7* (Helicon Soft) and saved following a CoreName-SampleName-FOVNumber.jpg
pattern (Fig. 2). Every stacked FOV image is then processed and segmented into individual specimen images using a custom
plugin (AutoRadio_Segmenter.ijm) developed for the *ImageJ / Fiji* software (V1.52n Schneider et al., 2012). The processing
190 steps are:

1. Open a stacked FOV image.
2. Subtract its background.
3. Adjust the minimum and maximum greyscale value to increase its contrast.
4. Invert the image and create a mask.
- 195 5. Threshold it in order to binarise it.
6. Blur it and threshold it again to obtain the overall shape of each particle.
7. Separate particles that are in contact with each other (require the configurable Biovoxxel "Water Irregular Features" plugin, available at: https://github.com/biovoxxel/BioVoxxel_Toolbox).
8. Define regions of interest (ROIs) for each particle.
- 200 9. Restore ROIs corresponding to every particle on the original FOV image.
10. Create a square vignette for each particle.
11. Save it into the corresponding "Core" folder and "Sample" subfolder.

Each sample results in approximately 1,000 to 3,000 individual segmented vignettes after the automated image processing and segmentation step.

205 2.5 Database building and CNN training

ParticleTrieur is a dedicated software program developed at CEREGE (Marchant et al., accepted), that enables the operator to visualise and assign vignettes to manually defined classes, and uses the k-NN (k-nearest neighbours) algorithm to aid in identification by self-learning and progressively suggesting identification once enough radiolarian pictures are identified ([the reader is referred to Marchant et al. \(accepted\) for more information](#)). Using this software, a large dataset of radiolarian taxa images (called the AutoRadio Database) was progressively built (the ~~original~~^{current} version of the database used in this study can be downloaded at: <http://microautomate.cerege.fr/dat>). It is currently composed of ~~17,065~~^{21,746} images, corresponding to ~~142~~¹³² classes/taxa. Each class contains between 1 and about 1,000 images.

Once labelled, this database was used to train a CNN (convolutional neural network) for the automated taxonomical identification of radiolarian vignettes resulting from the automated microscope image acquisition, processing, and segmentation steps([Fig. 7](#)). The best results were obtained using a resnet50 topology with added cyclic and gain layers (resnet50_cyclic_gain_tl), greyscale images resized to 256x256 px, a batch size (number of images presented per training iteration) of 64, 30 epochs and four drops for the adaptive learning rate system (ALR), and augmentation (Marchant et al., accepted). This training process lasts about 30 min and generates two files that can then be used for automated recognition (network_info.xml and frozen_model.pb files).

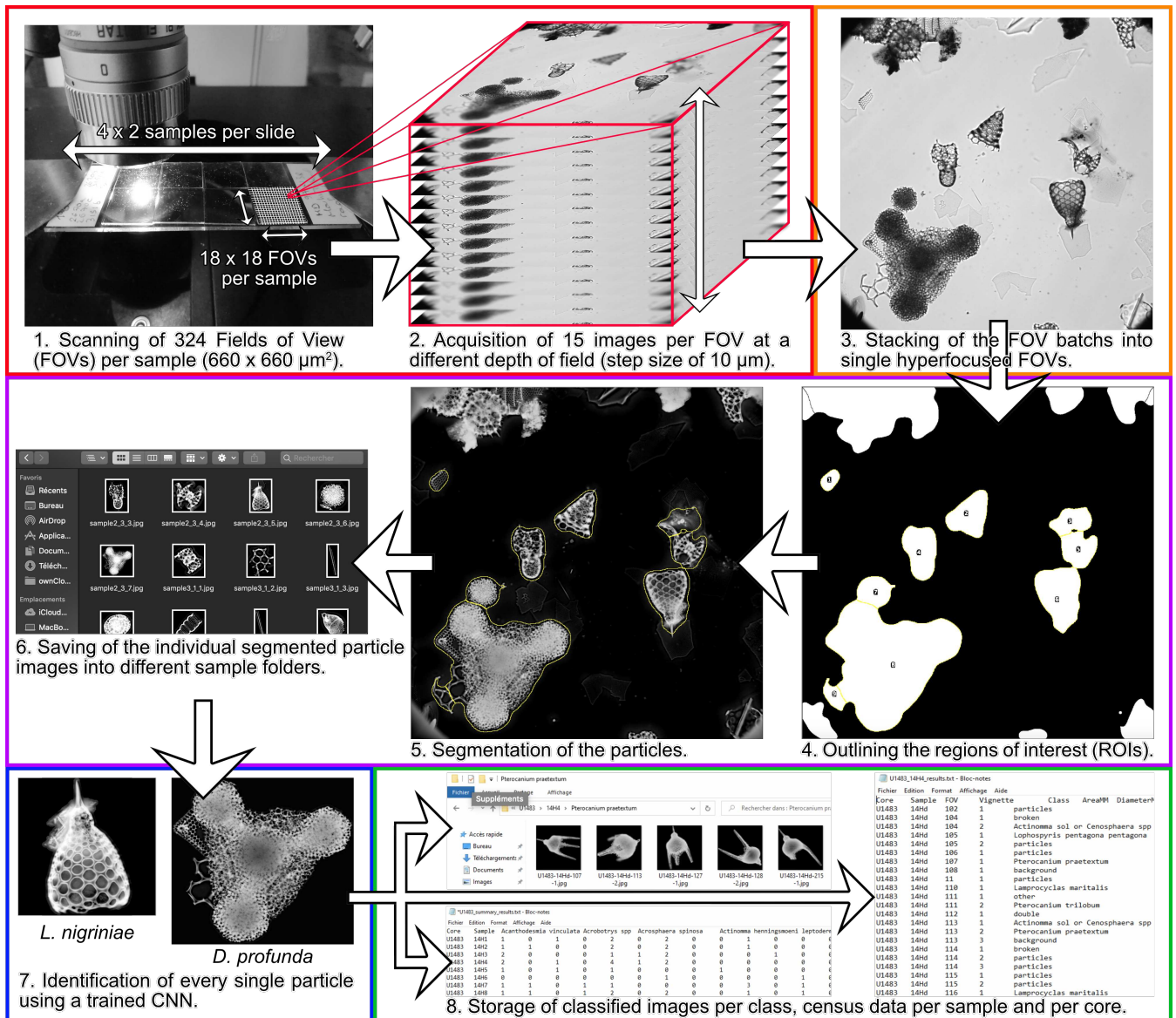


Figure 2. Automated radiolarian image acquisition, processing and identification workflow. 1, 2 (red rectangle): Automated acquisition steps. 3 (Orange rectangle): Automated FOV images stacking step. 4, 5, 6 (Purple rectangle): Automated FOV images processing and segmentation steps. 7 (Blue rectangle): Automated recognition step. 8 (Green rectangle): Automated export of classified images, census counts, and morphometric measurements.

Once individual vignettes of radiolarian specimens are generated and saved during the *ImageJ* processing and segmentation step, they are automatically opened in *ParticleTrieur* using its server mode, controlled by the second *LabVIEW* interface. These vignettes are then automatically assigned to a class using the trained CNN. Individual vignettes are then automatically moved into folders corresponding to their core and sample, and subfolders corresponding to their assigned class. Using one
 225 microscope, about 8,000 individuals from two slides (16 cover slides corresponding to 16 samples) can be imaged per day (about 500 specimens per sample). This fully automated stacking, processing, segmentation and identification step takes about 50 min per sample and operates in parallel to the image acquisition step.

Two types of data are then automatically exported (Fig. 2):

(1) For each sample, a "sample results" file is generated which assembles metadata and morphometric measurements. Each
 230 taxonomic ID is then returned to the *LabVIEW* interface and indexed with its corresponding vignette name (also containing the core, sample, FOV and vignette numbers in column) into a .txt file for each vignette (in row). For each specimen, morphometric measurements, such as "Area", "Diameter", "Major Axis", "Minor Axis", "Circularity", "Roundness", "Solidity", and "Eccentricity" are also automatically appended to the .txt file.

(2) For each core, census data counts of each sample are automatically compiled. A "core results" file is generated during
 235 this process where the abundance of each taxon (in column) for each sample (in row) is automatically incremented.

3 Results and discussion

3.1 Description of the database

Of the ~~17,065~~21,746 images used to construct the database, ~~112~~132 morphoclasses were created. Of all these classes, ~~104~~124 belong to Neogene to Quaternary radiolarian taxa (116 classes corresponding to species or groups of ~~species~~(94 ~~two to three~~ species and containing 11,126 images, 7 classes corresponding to ~~species,~~9 to genera, genera and containing 1,932 images, and 1 ~~to family, see Table 1~~corresponding to family and containing 677 images) and are part of the Spumellaria families Actinommidae, Coccodiscidae, Heliodiscidae, Litheliidae, Pyloniidae, Spongodiscidae, and Tholoniidae; and of the Nassellaria families Artostrobiidae, Cannobotryidae, ~~Collozoidae, Carpoecaniidae, Carpocaniidae, Collozoidae,~~ Plagiacanthidae, Pterocorythidae, Theoperidae, and Trissocyclidae (see Fig. 3, which includes some example images). Eight non-radiolarian classes
 245 (corresponding to "background", "broken" specimens, air "bubble", "diatom", "double", "porous fragments", siliceous "particles", and "spicule" and containing 8,011 images) were also defined to train the network to recognise these non-radiolarian images that usually represent 1/2 to 4/5 of the total vignettes.

An extensive overview of the existing Neogene to Quaternary literature was used for the taxonomy and identification of each class, and to define as accurately as possible our assemblages and the observed taxa (~~including Ling and Anikouchine, 1967; Nigrini and M~~
 250 Synonymies were also taken into account, especially regarding the work of Boltovskoy (1998, 1999). This means that a few species were regrouped into a single class when a significant morphological gradation was observed and when the limit be-

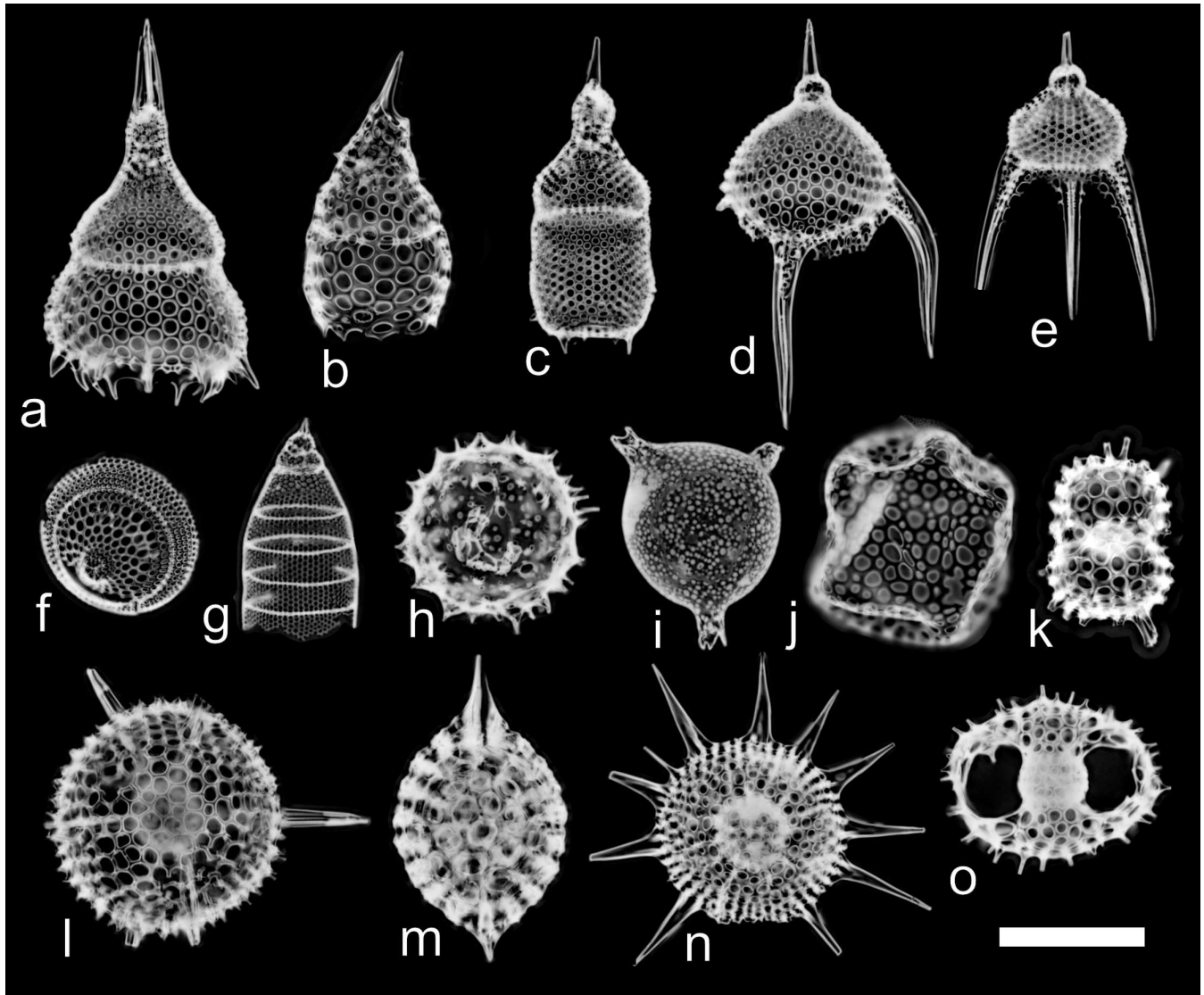


Figure 3. Examples of radiolarian vignettes generated by the automated acquisition, processing and recognition workflow. (a) *Lamprocyclas maritalis*. (b) *Lamprocyrtis hannai*. (c) *Theocorythium trachelium*. (d) *Pterocanium trilobum*. (e) *Pterocanium praetextum*. (f) *Eucecryphalus sestrodiscus*. (g) *Eucyrtidium acuminatum* / *hexagonatum*. (h) *Acrosphaera spinosa*. (i) *Solenosphaera chierchiae*. (j) *Collosphaera tuberosa*. (k) *Didymocyrtis tetralthalamus tetralthalamus*. (l) *Hexacontium* spp. (m) *Stylatractus neptunus*. (n) *Heliodiscus asteriscus*. (o) *Tetrapyle octacantha* group. Scale bar 100 μ m.

tween the considered species was blurry (e.g., *Eucyrtidium acuminatum* and *E. hexagonatum*; *Sithocampe arachnea* and *S. lineata*; *Actinomma henningsmoeni* and *A. leptodermum*). All manual taxonomic IDs during the building of the database were reviewed by a radiolarian taxonomy expert (G. Cortese) to ensure consistent and accurate identifications.

255 3.2 Results of the CNN training

One of the best ways to assess the efficiency of a trained CNN is to look at its confusion matrix (Fig. 4; [the original excel spreadsheet is available as supplementary material, see Appendix B](#)). Right before the training step, the dataset is automatically split into two subsets: one being the training set, and the second one the test set. The data split chosen for this study is 1/5. This means that 4/5 of the original images are used for training (training set) while the remaining 1/5 (test set) of the original images are used for testing the CNN efficiency by calculating several indices. The efficiency results are then represented by the overall accuracy Eq. (1), precision Eq. (2) recall Eq. (3), and individual recall for each class, with these terms defined as:

$$Accuracy = \frac{\text{Number of images correctly classified}}{\text{total number of images}} \quad (1)$$

It is the overall performance of the system regardless of class. If you select a random image from the dataset and classify it, the overall accuracy is the probability (in %) that the returned classification is correct.

$$265 \text{ Precision} = \frac{\text{Number of images that were classified as class } N \text{ and actually belong to class } N}{\text{total number of images classified as class } N} \quad (2)$$

Precision is a metric for a specific class: it is the probability (in %) that an image classified as class N is actually from class N, divided by the total number of images classified as class N.

$$Recall = \frac{\text{Number of images in class } N \text{ that were correctly classified}}{\text{total number of images in class } N} \quad (3)$$

Recall is, for a specific class, the probability (in %) that a random image from class N is correctly classified, divided by the number of images belonging to class N. Recall is basically the accuracy of a single class. Individual recall scores for each class are visible in the confusion matrix (Fig. 4) as the % of class N (in row) that was identified as various classes (in column). For example, for the first row "Acanthodesmia vinculata", ~~73~~95 % of the images belonging to this class were correctly identified, while ~~18~~5 % were classified as "~~Tholospyris~~" and ~~9~~ % as "~~broken~~" Lophospyris pentagona pentagona". If the CNN training was perfect, the diagonal should only exhibit "100" values. The single overall recall and precision scores are the respective values averaged across all the classes.

During the CNN training, all classes containing less than 10 images (corresponding to rare species, currently lacking images) were automatically fused into a single "other" class. Of the original ~~112 classes, 84~~132 classes, 109 classes (including ~~76 radiolarian classes~~101 radiolarian classes from Middle Miocene, to Quaternary, and 8 non-radiolarian classes) were then trained to be recognised with a current overall precision of ~~about~~just above 90 % (~~89.6~~90.1 %) over every class. The average precision is above ~~83 %~~(83.185 % (85.6 %) and the average recall is ~~above 80 %~~(80.2about 81 % (80.7 %)). A closer look at the matrix shows that classes with a low recall score usually correspond to classes containing an insufficient number of images (rare species, difficult to get on slide, that would require a significant amount of samples to be processed before

dataset for this class; Fig. 5 green square). When the test set is below 30 images per class, a higher number of classes (28 out of 84 trained classes; Fig. 5 orange square) show a recall score below 80 % (59 % in average). More images, about at least 150 (ideally 300) in total for each class, as defined above, are then likely to increase the recall and accuracy of these under-represented classes up to at least 80 %. To this aim, the database will be updated and populated gradually through the automated processing of new samples. As the aim of this database is to be free-access and participative online, people are encouraged to send and / or add pictures of these under-represented classes.

Surprisingly, a significant number (36 out of 84 trained classes; Fig. 5 blue square) of classes does not seem to require this number of images in their dataset, as less than 30 images in the test set was sufficient to reach a recall of more than 80 % (e.g., *Pterocanium trilobum* and *P. praetextum* showing a recall of While the ability of the network to distinguishing between morphologically / taxonomically very dissimilar taxa is very strong (Fig. 4: almost every values out of the squares corresponding to family groupings in the confusion matrix are 0; most of the specimens are usually assigned to their correct family: actinommiidae: 93 %; coccodiscidae: 93 %; heliodiscidae: 100 % with only 33 and 61 images in their original dataset, respectively, and 7 and 12 images in their test sets; Fig. 5 blue square). Finally, only a single class (*Porodiscus* sp; 1 class out of 84 trained) contained more than 30 images in its test set, and scored just less than; lithelidae: 88 %; pyloniidae: 95 %; spongodiscidae: 96 %; artostrobiidae: 97 %; cannobotryidae: 95 %; carpocaniidae: 93 %; collozoidae: 99 %; plagiacanthidae: 87 %; pterocorythidae: 98 %; theoperidae: 96 %; trissocyclidae: 99 %), we also tested its ability to distinguish between morphologically very similar forms, usually corresponding to closely related taxa (species or genera) by computing the accuracy of each radiolarian family present in the database (using the number of specimens of the test set and recall score for each class). Overall, the intra-family accuracy for each family is very high (actinommiidae: 80 % (79 %; Fig. 5, red square %; coccodiscidae: 89 %; heliodiscidae: 100 %; lithelidae: 84 %; pyloniidae: 90 %; spongodiscidae: 89 %; artostrobiidae: 93 %; cannobotryidae: 95 %; carpocaniidae: 93 %; collozoidae: 91 %; plagiacanthidae: 85 %; pterocorythidae: 92 %; theoperidae: 93 %; trissocyclidae: 94 %). For each class, the three classes where most of the misclassification occurs are summarised in Appendix C. Most of the misclassification usually happens with classes of the same family, or with the "broken" class, where a specific part of the investigated class might be recognised (e.g. part of a cephalis or part of a thorax, but always incomplete).

Plot of the number of images in the test set for each trained class (logarithmic scale) vs recall score per class.

3.3 Accuracy of the trained CNN on a random set of samples

In order to test the reliability and reproducibility of our trained CNN on actual samples, a slide on which 8 cover slides containing siliceous particles from 8 random samples with variable radiolarian abundances from cores MD97-2138 and MD97-2140 was was selected. Four Quaternary samples (400 to 6,400 years BP) from Core U1488A and 4 Miocene sample (10.116 to 10.694 ma) from Core U1483 (both from IODP Expedition 363) were then prepared, and their identification scores computed. This slide was automatically imaged, FOV pictures were automatically segmented and individual vignettes were automatically identified using the trained CNN. After a manual verification of every automated identification, 6 indices were computed: (1) the % of radiolarian images recognised as radiolarians (Fig. 6a); (2) the % of radiolarian images recognised as the correct radiolarian taxa (Fig. 6c); (3) the % of non-radiolarian images recognised as non-radiolarian particles (Fig. 6b); (4) the % of

non-radiolarian images recognised as the correct particle class (Fig. 6d); (5) the % of non-radiolarian images recognised as radiolarian (non-radiolarian false positive; Fig. 6e); and (6) the % of radiolarian recognised as non-radiolarian (radiolarian false positive; Fig. 6f).

Overall, ~~10,288-7,800~~ vignettes were identified and manually checked among ~~the~~ 8 samples containing between ~~623 and 2,372-444 and 1,502~~ images each. The abundance of radiolarians ranges from ~~41 to 340-176 to 697~~ specimens per sample. The results show that the 6 indices exhibit very close values between the 8 samples. In average, the proportion of radiolarians actually recognised as radiolarian is very high, about ~~98-100~~ % (Fig. 6a) and the proportion of radiolarians identified as the correct radiolarian taxa is about ~~90-93~~ % (Fig. 6b). Almost all radiolarian images are thus recognised as radiolarian with a ~~10~~ % error regarding their species identification. Regarding the non-radiolarian images, more than ~~99-95~~ % are recognised as non-radiolarian (Fig. 6c) and ~~about 98 again about 95~~ % are assigned to the correct class (Fig. 6d).

False positive identifications were also investigated and are relatively low. Among all the images identified as non-radiolarians, only ~~0.34-0.08~~ % should be assigned to radiolarians, and among all the images automatically recognised as radiolarians, about ~~4-6~~ % are non-radiolarian images. Within these ~~4-6~~ %, most of the non-radiolarian images confused with radiolarians exhibit radiolarian features and correspond to the non-radiolarian classes "broken" and "double" that either contain incomplete radiolarians, or radiolarians touching each other and cannot be assigned to a single species. These false positives are then usually assigned, in the "broken" class case, to the species partially present in the image, or in the "double" class case, to one of the species that can be distinguished.

3.4 Biostratigraphy

To explore the applicability of the automated radiolarian identification workflow for biostratigraphic studies, radiolarian faunal events, such as first occurrences (FOs) and last occurrences (LOs) of radiolarian taxa (about 30 zones were defined for the Cenozoic (Sanfilippo et al., 1985)) were compiled into an Excel spreadsheet. Here, we decided to focus on the biostratigraphy of the Neogene to Quaternary interval using the existing zonation (Nigrini, 1971; Lazarus et al., 1985; Johnson et al., 1989; Moore, 1995; Sanfilippo and Nigrini, 1998; Nigrini and Sanfilippo, 2001; Vigour and Lazarus, 2002; Nigrini et al., 2005; Sanfilippo et al., 1985; Kamikuri, 2017) and especially the recent work of Kamikuri et al. (2009) who compiled and documented the stratigraphic occurrences of 115 Neogene and Quaternary radiolarian species recovered from Ocean Drilling Program (ODP) Sites 845 and 1241 in the tropical Pacific Ocean.

The known stratigraphic ranges of species included in our database were then compiled into an Excel spreadsheet that automatically suggests the age of any sample, according to the composition of its radiolarian assemblage. As this spreadsheet follows the architecture of the automatically generated census data file for each core (core results file), it can be easily filled by copy-pasting this content. This operative workflow, that is automated from the image acquisition to the census counts and can suggest an age for the processed sample could thus contribute to the field of biostratigraphy.

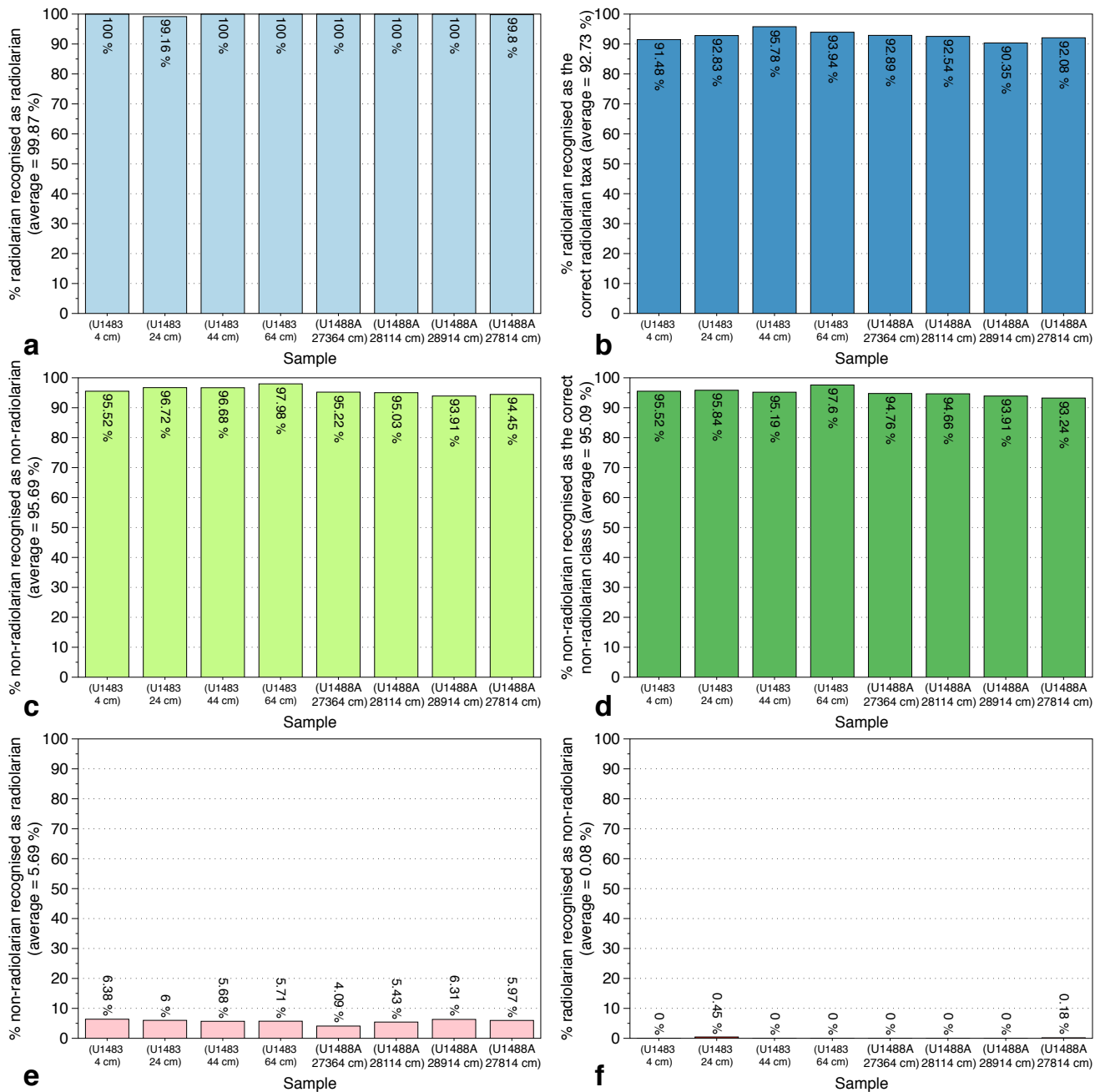


Figure 5. Identification indices evaluated on 8 Quaternary and Miocene random samples recovered from cores MD97-2138-U1483 and MD97-2140-U1488A (IODP Exp. 363).

3.5 Application to other datasets and other studies

355 To test the potential application and limits of our trained CNN on existing sets of images, we compiled various individual images of radiolarians from the literature including unstacked optical microscope images and scanning electron microscope (SEM) images. We then performed a simple colour inversion of the optical microscope image to obtain white specimens on a dark background. Of the hundred images tested, about half were correctly recognised while the other ones were mostly assigned to the "background" class, likely due to the blurry shell edges of the unstacked images, and to the "broken" class, as
360 only part of the shell were probably recognised. While this 50 % accuracy on a random set of unstacked optical microscope and SEM images may seem relatively low and arbitrary, it is very encouraging and promising for the development of future and extensive neural networks for automated radiolarian recognition regardless of the imaging method.

4 Conclusions

A new automated radiolarian workflow was developed and consists of a sequence of six steps:

365 1: A new microscopic slide preparation protocol to enable an efficient automated image acquisition on transmitted light microscopes and decrease the loss of material, as this can limit the investigation of samples where radiolarians are scarce.

2: Automated microscope image acquisition that can automatically image microscopic slides bearing up to 8 samples (324 FOV images per sample) at different focal depths (15 images per FOV, every 10 μm in depth).

3: Automated stacking of each batch of FOV images (using depth maps) to generate a single clear FOV with clearly distinct
370 radiolarian specimens.

4: Automated FOV image processing (contrast enhancement, B&W inversion) and segmentation to generate individual images for every radiolarian specimen.

5: Automated radiolarian recognition using a CNN, as well as calculating morphometric measurements.

6: Automated export of census data per sample (usually about 500 radiolarian images per sample), and storage of radiolarian
375 images in folders corresponding to their taxonomic identification for every sample.

The whole procedure is then entirely automated from the image acquisition to the census counts, and only required from the operator to prepare the micropalaeontological slides and put them under the microscope. The operative workflow described in this study can thus perform complex, tedious, time-consuming tasks such as taxonomic identification and census counts by producing reliable, reproducible, and accurate results. Moreover, as the system can identify most of the common Miocene to
380 Quaternary species, taxonomic specialists can focus on unknown and poorly documented forms. This workflow is achieved using a polyvalent and extensive radiolarian image database (currently ~~17,065~~ 21,746 images) and a ResNet CNN trained using transfer learning for modern and Neogene radiolarian identification. The CNN is currently able to recognise ~~84~~ 109 classes with an average precision of about 90 %, an overall score that was also obtained on a test performed on 8 random samples containing ~~more than 10,000~~ about 7,800 images. In order to continue to increase its efficiency, more images are required,
385 particularly so for rare species. To this aim, the database was made free-access and participative to increase the number of

images, especially for rare species where the recall score is relatively low, most likely due to low numbers of training images for these taxa.

This new workflow and associated CNN has the potential to make paleoclimate studies more approachable and feasible, along with biostratigraphy for very long sequences. The radiolarian census data can then be used to investigate the radiolarian assemblages variability for biostratigraphical purposes, and to develop, apply and improve existing assemblage-based palaeoenvironmental proxies such as SSTs (e.g., radiolarian-based palaeotemperatures for the late Quaternary, Cortese and Abelman, 2002; Subtropical (ST) Index, L  r et al., 2008; Radiolarian Temperature Index (RTI), applied to Miocene samples, Kamikuri, 2017) and paleoproductivity (e.g., Upwelling Radiolarian Index (URI), Caulet et al., 1992; Water Depth Ecology index (WADE), Lazarus et al., 2006). It also enables the investigation of evolutionary trends, appearance of new species, and rate of evolutionary change, a fascinating topic regarding radiolarians and other microfossil groups.

This dataset and following studies also enable the fast and accurate measurement of numerous morphometric parameters for each vignette that was assigned a class in the automated recognition step. In addition to the previous research applications, the morphometry aspect provides the possibility to investigate the link between the morphological variability of a species or an assemblage through time along a sedimentary record and elaborate/test scenarios to explain such variability. This new workflow will now be used on two Neogene to Recent sedimentary records from IODP Expedition 363 (Hole U1483A, Hole U1488A), recovered in the West Pacific Warm Pool.

Data availability. The original version of the AutoRadio database used in this study can be downloaded at: <http://microautomate.cerege.fr/dat>. It is currently composed of 21,746 images, corresponding to 132 classes/taxa.

Code and data availability. A manual version of the AutoRadio_Segmenter.ijm plugin (automated image processing performed on *ImageJ / Fiji*), developed to process a root folder ("Core"), containing subfolders ("Samples") of images ("FOVs") is available online for free at <https://github.com/microfossil/ImageJ-LabView-Scripts>. To use it, download the .ijm file and save it into the ImageJ/plugins folder, and it will be available to use after restarting *ImageJ / Fiji*.

Appendix A

~~Number of specimens counted and % for different weight of radiolarian material drop in the decanter. The dashed blue line corresponds to the minimum counts required per sample.~~

Author contributions. MT designed the experiment, performed its technical aspects, including the image preprocessing and write the first draft of the manuscript. RM developed ParticuleTrieur. MT and GC established the taxonomy of the training set. YG developed the automation of the microscope. LB and TdGT were at the origin of the project. All authors participated in the writing of the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

415 *Disclaimer.* TEXT

Acknowledgements. We thanks IODP-France for financial support for this project. This work was also supported by the French National Research Agency (ANR) as part of the French platform called Nano-ID (EQUIPEX project ANR-10-EQPX-39- 01) and the ANR project FIRST (ANR-15-CE4-0006-01). We also thanks the program Ocean Acidification from the french Foundation for Research on Biodiversity (FRB), and the Ministry for the Ecological and Inclusive Transition (MTES) in supporting the project COCCACE.

420 References

- Abelmann, A.: Radiolarian taxa from Southern Ocean sediment traps (Atlantic sector), *Polar Biol.*, 12, 373–385, 1992.
- Abelmann, A., and Nimmergut, A.: Radiolarians in the Sea of Okhotsk and their ecological implication for paleoenvironmental reconstructions, *Deep-Sea Res.*, II, 52, 2302–2331, 2005.
- Abelmann, A., Brathauer, U., Gersonde, R., Siegier, R., and Zielinski, U.: Radiolarian-based transfer function for the estimation of sea surface temperatures in the Southern Ocean (Atlantic Sector), *Paleoceanography*, 14, 410–421, 1999.
- 425 [Apostol, L.A., Marquez, E., Gasmen, P., and Solano, G.: Radss: A radiolarian classifier using support vector machines, 7th International Conference on Information, Intelligence, Systems & Applications \(IISA\), 2016.](#)
- Beaufort, L., Dollfus, D.: Automatic recognition of coccoliths by dynamical neural networks, *Mar. Micropaleontol.*, 51, 57–73, 2004.
- Beaufort, L., Chen, M. T., Chivas, A., Manighetti, B.: Campagne IPHIS - IMAGES III / MD 106 du 23-05-97 au 28-06-97. Les Publications de l'Institut français pour la recherche et la technologie polaires, Les Rapports des campagnes a la mer, 151p, 1997, Open Access version: <https://archimer.ifremer.fr/doc/00629/74140/>
- 430 Beaufort, L., de Garidel-Thoron, T., Mix, A. C., Pisias, N. G.: ENSO-like forcing on oceanic primary production during the Late Pleistocene. *Science* 293, 2440–2444.
- Beaufort, L., Barbarin, N., and Gally, Y., 2014. Optical measurements to determine the thickness of calcite crystals and the mass of thin carbonate particles such as coccoliths, *Nat. Protoc.*, 9, 633–642, 2014.
- 435 Bjørklund, K. R., and Goll, R. M.: Internal skeletal structures of *Collosphaera* and *Trisolenia*: A case of repetitive evolution in the Collosphaeridae (Radiolaria), *J. Paleontol.*, 53, 1293–1326, 1979.
- Boltovskoy, D.: Classification and distribution of South Atlantic Recent polycystine Radiolaria, *Palaeontol. Electron.*, 1, 6A, 111p, 1998.
- Boltovskoy, D.: Radiolaria Polycystina, in: *South Atlantic Zooplankton*, edited by Boltovskoy, D., Backhuys Publishers, Leiden, 149–212, 440 1999.
- Boltovskoy, D., and Jankilevich, S. S.: Radiolarian distribution in east equatorial Pacific plankton, *Oceanol. Acta*, 8, 101–123, 1985.
- Boltovskoy, D., and Vrba, A.: Classification and geographic distribution of *Stylodictya*-type radiolarians, *Micropaleontology*, 34, 332–340, 1998.
- Boltovskoy, D., Kling, S. A., Takahashi, K., and Bjorklund, K.: World atlas of distribution of living radiolaria, *Palaeontol. Electron.*, 13, 445 1–230, 2010.
- Boltovskoy, D., Anderson, O. R., and Correa, N. M.: Radiolaria and Phaeodaria, in: *Handbook of the Protists*, edited by Archibald, J. M., Simpson, A. G. B., Slamovits, C., Springer, 1–33, 2017.
- Bourel, B., Marchant, R., de Garidel-Thoron, T., Tetard, M., Barboni, D., Gally, Y., Beaufort, L.: Automated recognition by multiple convolutional neural networks of modern, fossil, intact and damaged pollen grains, *Comput. and Geosci.*, ~~accepted~~ [140, 2020.](#)
- 450 Budai, A., Riedel, W. R., and Westberg, M. J.: A general-purpose paleontologic information decide, *J. Paleontol.*, 54, 259–262, 1980.
- Campbell, A. S.: Radiolaria, Part D: Protista 3, in: *Treatise on Invertebrate Paleontology*, edited by Moore, R.C., Geological Society of America, University of Kansas Press, Lawrence, DI-DI63.
- Calet, J. P., and Nigrini, C.: The genus *Pterocorys* (Radiolaria) from the tropical Late Neogene of the Indian and Pacific Oceans, *Micropaleontology*, 34, 217–235, 1988.

- 455 Caulet, J. P., Vénec-Peyré, M. T., Vergnaud-Grazzini, C., and Nigrini, C.: Variation of South Somalian upwelling during the last 160 ka: Radiolarian and foraminifera records in Core MD-85674, in: *Upwelling Systems: Evolution Since the Early Miocene*, edited by Summerhayes, C. P., Prell, W. L., and Emeis, K. C., Geol. Soc. Spec. Publ., 64, Geological Society, London, 379–389, 1992.
- Caulet, J. P., Sanfilippo, A., and Nigrini, C.: "Radworld", a taxonomic relational database for radiolarians, in: *InterRad II and Triassic Stratigraphy Symposium: a joint international conference hosted by the International Association of Radiolarian Paleontologists, IGCP*
- 460 467 and the Subcommission of Triassic Stratigraphy, edited by Lürer, V., Hollis, C., Campbell, H., and Simes, J., GNS Science, Lower Hutt, 47, 2006.
- Cortese, G., and Abelmann, A.: Radiolarian-based paleotemperatures during the last 160 kyr at ODP Site 1089 (Southern Ocean, Atlantic Sector), *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 182, 259–286, 2002.
- Cortese, G., Dolven, J. K., Bjørklund, K. R. and Malmgren, B. A.: Late Pleistocene-Holocene radiolarian paleotemperatures in the Norwegian
- 465 Sea based on artificial neural networks, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 224, 311–332, 2005.
- de Garidel-Thoron, T., Rosenthal, Y., Bassinot, F., and Beaufort, L.: Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years, *Nature*, 433, 294–298, 2005.
- Dollfus, D., and Beaufort, L.: Fat neural network for recognition of position-normalised objects, *Neural networks*, 12, 553–560, 1999.
- Dolven, J. K., and Skjerpen, H. A.: An online micropaleontology database: Radiolaria.org, *Eclogae Geol. Helv., Supplement 1*, 63–66, 2006.
- 470 Fatela, F., and Taborda, R.: Confidence limits of species proportions in microfossil assemblages, *Mar. Micropalaeontol.*, 45, 169–174, 2002.
- Hernández-Almeida, I., Bjørklund, K. R., Sierro, F. J., Filippelli, G. M., Cacho, I., and Flores, J. A.: A high resolution opal and radiolarian record from the subpolar North Atlantic during the Mid-Pleistocene Transition (1069–779 ka): Palaeoceanographic implications, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 391, 49–70, 2013.
- Hernández-Almeida, I., Cortese, G., Yu, P. S., Chen, M. T., and Kucera, M.: Environmental determinants of radiolarian assemblages in the
- 475 western Pacific since the last deglaciation, *Paleoceanography*, 32, 830–847, 2017.
- Itaki, T., Matsuoka, A., Yoshida, K., Machidori, S., Shinzawa, M., and Todo, T.: Late spring radiolarian fauna in the surface water off Tassha, Aikawa Town, Sado Island, central Japan, *Sci. Rep. Niigata Univ. (Geol.)*, 17, 41–51, 2003.
- Johnson, D. A., Schneider, D. A., Nigrini, C., Caulet, J. P., and Kent, D. V.: Pliocene–Pleistocene radiolarian events and magnetostratigraphic calibrations for the tropical Indian Ocean, *Mar. Micropaleontol.*, 14, 33–66, 1989.
- 480 Kamikuri, S.: Late Neogene Radiolarian Biostratigraphy of the Eastern North Pacific ODP Sites 1020/1021, *Paleontol. Res.*, 21, 230–254, 2017.
- Kamikuri, S., and Moore, T. C.: Reconstruction of oceanic circulation patterns in the tropical Pacific across the early/middle Miocene boundary as inferred from radiolarian assemblages, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 487, 136–148, 2017.
- Kamikuri, S., Motoyama, I., Nishi, H., and Iwai, M.: Neogene radiolarian biostratigraphy and faunal evolution of ODP Sites 845 and 1241,
- 485 eastern equatorial Pacific, *Acta Palaeontol. Pol.*, 54, 713–742, 2009.
- [Keceli, A.S., Kaya, A., and Keceli, S.U.: Classification of radiolarian images with hand-crafted and deep features, *Comput. and Geosci.*, 109, 67–74, 2017.](#)
- Lazarus, D. B.: Environmental control of diversity, evolutionary rates and taxa longevities in Antarctic Neogene Radiolaria, *Palaeontol. Electron.*, 32, 1–32, 2002.
- 490 Lazarus, D. B.: A brief review of radiolarian research, *Paläontol. Z.*, 79, 183–200, 2005.
- Lazarus, D., Spencer-Cervato, C., Pika-Biolzi, M., Beckmann, J. P., Von Salis, K., Hilbrecht, H., and Thierstein, H.: Revised chronology of Neogene DSDP Holes from the world ocean, *Ocean Drilling Program Technical Note*, 24, 1–301, 1985.

- Lazarus, D. B., Faust, K., and Popova-Goll, I.: New species of prunoid radiolarians from the Antarctic Neogene, *J. Micropaleontology*, 24, 97–121, 2005.
- 495 Lazarus, D., Bittniok, B., Diester-Haass, L., Meyers, P., and Billups, K.: Comparison of radiolarian and sedimentologic paleoproductivity proxies in the latest Miocene-Recent Benguela Upwelling System, *Mar. Micropaleontol.*, 60, 269–294, 2006.
- Lazarus, D., Suzuki, N., Caulet, J. P., Nigrini, C., Goll, I., Goll, R., Dolven, J. K., Diver, P., and Sanfilippo, A.: An evaluated list of Cenozoic-Recent radiolarian species names (Polycystinea), based on those used in the DSDP, ODP and IODP deep-sea drilling programs, *Zootaxa*, 3999, 301–333, 2015.
- 500 Ling H. Y., and Anikouchine W. A.: Some spumellarian Radiolarian from the Java, Philippine, and Mariana Trenches, *Jour. Paleon.* 41, 1481–1491, 1967.
- Lüer, V., Hollis, C. J., and Willem, H.: Late Quaternary radiolarian assemblages as indicators of paleoceanographic changes north of the subtropical front, offshore eastern New Zealand, southwest Pacific, *Micropaleontology*, 54, 49–69, 2008.
- Marchant, R., Tetard, M., Pratiwi, A., de Garidel-Thoron, T.: Classification of down-core foraminifera image sets using convolutional neural networks, *J. Micropalaeontol.*, submitted, doi: 10.1101/840926
- 505 Matsuoka, A.: Catalogue of living polycystine radiolarians in surface waters in the East China Sea around Sesoko Island, Okinawa Prefecture, Japan, *Sci. Rep. Niigata Univ. (Geol.)*, 32, 57–90, 2017.
- Matsuzaki, K. M., Suzuki, N., Nishi, H., Hayashi, H., Gyawali, B. R., Takashima, R., and Ikehara, M.: Early to middle Pleistocene paleoceanographic history of southern Japan based on radiolarian data from IODP Exp 314/315 Sites C0001 and C0002, *Mar. Micropaleontol.*, 118, 17–33, 2015.
- 510 Matsuzaki, K. M., Itaki, T. and Tada, R.: Paleoceanographic changes in the Northern East China Sea during the last 400 kyr as inferred from radiolarian assemblages (IODP Site U1429), *Prog. Earth Planet. Sci.*, 6, 1–21, 2019.
- Moore, T. C.: Method of randomly distributing grains for microscopic examination, *J. Sediment. Petrol.*, 43, 904–906, 1973.
- Moore, T. J. Jr.: Radiolarian stratigraphy, Leg 138, *Proc. Ocean Drill. Prog. Sci. Results*, 138, 191–232, 1995.
- 515 Motoyama, I., Yamada, Y., Hoshiba, M., and Itaki, T.: Radiolarian Assemblages in Surface Sediments of the Japan Sea, *Paleontol. Res.*, 20, 176–206, 2016.
- Nigrini, C.: Radiolarian zones in the Quaternary of the equatorial Pacific Ocean, in: *The Micropalaeontology of Oceans*, edited by Funnell B. M., and Riedel, W. R., Cambridge University Press, Cambridge, 443–461, 1971.
- Nigrini, C., and Moore, T. C.: A guide to modern Radiolaria - with taxonomic descriptions and illustrations of radiolarian species, Cushman Foundation for Foraminiferal Research, Sp. Pub., 16, 1979.
- 520 Nigrini, C., and Lombardi, G.: A guide to Miocene Radiolaria, Cushman Foundation Foraminiferal Research, Sp. Pub., 22. S1–S102, N1–N206, 1984.
- Nigrini, C., and Sanfilippo, A.: Cenozoic radiolarian stratigraphy for low and middle latitudes with descriptions of biomarkers and stratigraphically useful species, ODP Tech. Note 27, 2001, available from World Wide Web: <http://www.odp.tamu.edu/publications/tnotes/tn27/index.html>
- 525 Nigrini, C., Sanfilippo, A., and Moore, T. J., Jr.: Cenozoic radiolarian biostratigraphy: a magnetobiostratigraphic chronology of Cenozoic sequences from ODP Sites 1218, 1219, and 1220, equatorial Pacific, in: *Proc. ODP, Sci. Results 199*, edited by Wilson, P. A., Lyle, M., and Firth, J. V., 1–76, 2005.
- Panitz, S., Cortese, G., Neil, H. L., and Diekmann, B.: A radiolarian-based palaeoclimate history of Core Y9 (Northeast of Campbell Plateau, New Zealand) for the last 160kyr, *Mar. Micropaleontol.*, 116, 1–14, 2015
- 530

- Renaudie, J., Gray, R., and Lazarus, D. B.: Accuracy of a neural net classification of closely-related species of microfossils from a sparse dataset of unedited images, *PeerJ Preprints*, 6:e27328v1, 2018.
- Riedel, W. R.: Subclass Radiolaria, in: *The fossil record*, edited by Harland W.B. et al., Geol. Soc. London, 291–298, 1967.
- Rosenthal, Y., Holbourn, A. E., Kulhanek, D. K., and the Expedition 363 Scientists: Western Pacific Warm Pool, *Proceedings of the International Ocean Discovery Program*, 363: College Station, TX (International Ocean Discovery Program), 2018.
- 535 Sandoval, M. I.: Miocene to recent radiolarians from southern pacific coast of Costa Rica, *Rev. Geol. Amér. Central*, 58, 115–169, 2018.
- Sanfilippo, A., and Nigrini, C.: Code numbers for Cenozoic low latitude radiolarian biostratigraphic zones and GPTS conversion tables, *Mar. Micropaleontol.*, 33, 109–156, 1998.
- Sanfilippo, A., Westberg-Smith, M. J., and Riedel, W. R.: Cenozoic Radiolaria, in: *Plankton Stratigraphy (Vol. 2): Radiolaria, Diatoms, Silicoflagellates, Dinoflagellates, and Ichthyoliths*, edited by Bolli, H. M., Saunders, J. B., and Perch-Nielsen, K., Cambridge Univ. Press, Cambridge, UK, 631–712, 1985.
- 540 Schneider, C. A., Rasband, W. S., and Eliceiri, K. W.: NIH Image to ImageJ: 25 years of image analysis, *Nat. Methods*, 9, 671–675, 2012.
- Sharma, V., Singh, S., and Rawal, N.: Early Middle Miocene Radiolaria from Nicobar Islands, Northeast Indian Ocean, *Micropaleontology*, 45, 251–277, 1999.
- 545 Schrock, R. R., and Twenhofel, W. H.: *Principles of Invertebrate Palaeontology*, New second edition, New York-London, McGraw Hill, 816p, 1953.
- Suzuki, N., and Not, F.: Biology and Ecology of Radiolaria, in: *Marine Protists*, edited by Ohtsuka, S., Suzaki, T., Horiguchi, T., Suzuki, N., and Not, F, Springer, Tokyo, 2015.
- Takahashi, K.: Radiolaria: flux, ecology, and taxonomy in the Pacific and Atlantic, *Woods Hole Oceanogr. Inst., Ocean Biocoenosis Ser.*, 3, 1–303, 1991.
- 550 Takahashi, K., and Honjo, S.: Radiolarian skeletons: size, weight, sinking speed, and residence time in tropical pelagic oceans, *Deep-Sea Res.*, 30, 543–568, 1983.
- Vigour, R., and Lazarus, D.: Biostratigraphy of late Miocene–early Pliocene radiolarians from ODP Leg 183 Site 1138, in: *Proc. ODP, Sci. Results*, 183, edited by Frey, F. A., Coffin, M. F., Wallace, P. J., and Quilty, P. G., 1–17, 2002.
- 555 Welling, L. A., Pisias, N. G., and Roelofs, A. K.: Radiolarian microfauna in the northern California Current System: indicators of multiple processes controlling productivity, in: *Upwelling Systems. Evolution since the Early Miocene*, edited by Summerhayes, C. P., Prell, W. L. and Emeis, K. C., London Geological Society: Geological Society Special Publication, 64, 177–195, 1992.
- Zhang, L. L., and Suzuki, N.: Taxonomy and species diversity of Holocene pylonoid radiolarians from surface sediments of the northeastern Indian Ocean, *Palaeontol. Electron.*, 20.3.48A, 1–68, 2017.
- 560 Zhang, L. L., Chen, M. H., Xiang, R., Zhang, J. L., Liu, C. J., Huang, L. M., and Lu, J.: Distribution of polycystine radiolarians in the northern South China Sea in September 2005, *Mar. Micropaleontol.*, 70, 20–38, 2009.