

EGU2020-21465

<https://doi.org/10.5194/egusphere-egu2020-21465>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



## Multi-labelled taxonomic prediction using a small benthic foraminifera dataset trained on a FastAI library

**Abduljamiu O. Amao** and Michael Kaminski

King Fahd University of Petroleum & Minerals, College of Petroleum Engineering and Geosciences, Center for Integrative Petroleum Research, Dhahran, Saudi Arabia (amao@kfupm.edu.sa)

The study explored an end-to-end application of a ResNet convolutional neural network (transfer learning) to classify benthic foraminifera images using the FastAI library. 201 SEM images of 13 benthic foraminifera including *Ammonia convexa*, *Ammonia tepida*, *Asterorotalia gaimardi*, *Asterorotalia indica*, *Bulimina biserialis*, *Bulimina marginata*, *Elphidium advenum*, *Elphidium hispidum*, *Elphidium jenseni*, *Elphidium neosimplex*, *Perenoplis arianus*, *Perenoplis pertusus* and *Quinqueloculina* sp. The images were separated into two groups in a 80 -20 split for training and validation dataset respectively. We successfully trained a state-of-the-art image classifier for a very small dataset, achieving 96.5% accuracy in just a handful of lines of code on a very small dataset i.e. accurately predicting the binomial nomenclature of species. The fastai AI/Machine learning library we used offers interesting prospects in taxonomy where it can be used for multilabel image classification. Fastai's recent research breakthroughs are embedded in the library, resulting in significantly improved accuracy and speed over other deep learning libraries, whilst requiring dramatically less code. The library sits on top of PyTorch and provides a single consistent API to the most important deep learning applications and data types. It also offers an opportunity to a novice user, new to data science to apply state of the art deep learning to practical problems quickly and reliably. It has several advantages over other known libraries is its flexibility to import data of various kind and from various sources. It is fast, has a large and friendly community backing and its immune to several limitation of other libraries.