Automated microfossil classification by image recognition and machine learning to develop an AI system for age-dating

*Tatsuhiko Hoshino[1,2], Yousuke Taira[3], Hitoshi Saitou[3], Kyoko Hagino[4], Jonaotaro Onodera[5], Takuya Itaki[6], Tatsuhiko Yamaguchi[4], Fumio Inagaki[1,2,7]

1. Kochi Institute for Core Sample Research, JAMSTEC, 2. Research and Development Center for Submarine Resources, JAMSTEC, 3. NEC Corporation, 4. Center for Advanced Marine Core Research, Kochi University, 5. Research and Development Center for Global Change, JAMSTEC, 6. Marine Geology Research Group, AIST, 7. Research and Development Center for Ocean Drilling Science, JAMSTEC

The occurrence and/or disappearance of planktonic microfossils in marine subsurface sediments (e.g., diatoms, foraminifera, radiolarians, and calcareous nannoplankton) is diagnostic for identifying age of sediment cores in scientific ocean drilling. For age-dating, significant experience and expertise on taxonomic identification of microfossils are required, which processes including extraction, preparation of slides, microscopic observation, and identification that need laborious onboard efforts. Because the age-dating is one of the crucial measurements for scientific ocean drilling, micropaleontologists have a significant role with huge efforts on the microscopic work onboard since DSDP in 1960’s.
To date, computers can replace human beings as the drivers of cars, because of rapid improvement of process capability. A defeat of the world’s best Go player by the Google DeepMind’s artificial-intelligence program ”AlphaGo” is still fresh in our mind. This news shocked the world because Go create much more possible scenarios than chess and Shogi which had already surrendered to computer software and thus was considered to be the last bastion of human superiority. Deep learning, which is a kind of machine learning, enabled the victory of computer. In machine learning, by giving data consisting of input objects and desired output values, an algorithm analyzes and produces an inferred function automatically, providing rapid and accurate autonomous decision-making. The machine learning technology is used for pattern recognition, allowing recognition of many types of images. For example, face authentication system has been applied to immigration control at international airports, security gate system at amusement parks and concert venues.
The aim of this study is to apply machine learning and image recognition technologies to taxonomic identification of microfossils and develop an AI system which can automatically determine the age of cored sediments. As a feasibility study, we tested the NEC’s AI software “RAPID Machine Learning” to identify two kinds of calcareous nanofossils with relatively simple structures. At first, we prepared smear slides and took images of nannofossils of *Pseudoemiliania* (P) and *Reticulofenestra* (R) by polarization microscopy. For training data, 32 images of each P and R were rotated by every 90 degrees to 270 degrees and 256 images were obtained in total. For the test, 10 images of each P and R were prepared. A statistical model was built automatically by inputting all the training images and data (i.e., correct answer of P or R) into the RAPID Machine Learning, and then the model was tested for identification of the nannofossils.
As a result, 6 out of 10 images of P, and 4 out of 10 images of R were correctly identified with a confidence of 60% or more. The correctly identified images with high degree of confidence had distinctive shape to P or R, confirming the feasibility of applying machine learning to identification of microfossils. On the other hand, ambiguous images without the distinctive shapes were not correctly identified with low degree of statistical confidence. In the future study, the accuracy of identification will be improved by preprocessing the images as sharpening by using graphics software and increasing the number of training data (about 1,000 images per a object are required generally).

Keywords: age-dating, microfossil, machine learning, image recognition, AI

