Japanese mayfly family classification with a vision transformer model

Yuichi Iwasaki[1,5,*], Hiroko Arai[2,6,9*], Akihiro Tamada[3,7], Hirokatsu Kataoka[4,8]

[1] National Institute of Advanced Industrial Science and Technology (AIST), Research Institute of Science for Safety and Sustainability, 16-1 Onogawa, Tsukuba, Ibaraki 305-8569, Japan

[2] National Institute of Advanced Industrial Science and Technology (AIST), Research Center for Emerging Computing Technologies, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

[3] Graduate School of Life Sciences, Tohoku University, Miyagi 980-0812, Japan

[4] National Institute of Advanced Industrial Science and Technology (AIST), Artificial Intelligence Research Center (AIRC), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8560, Japan

Corresponding authors: Hiroko Arai (arai-h@aist.go.jp) and Yuichi Iwasaki (yuichiwsk@gmail.com, yuichi-iwasaki@aist.go.jp)

*These authors equally contributed to this work

Total number of words (text body): 2443

Total number of figures: 5

Total number of tables: 2

Abstract

Benthic macroinvertebrates are a frequently used indicator group for biomonitoring and biological assessment of river ecosystems. However, their taxonomic identification is laborious and requires special expertise. In this study, we aimed to assess the capability of a vision transformer (ViT) model for family-level identification of mayflies (order Ephemeroptera). Specifically, we focused on evaluating the model's capacity to classify three commonly found mayfly families (Baetidae, Ephemerellidae, and Heptageniidae) as well as other families that were grouped together. For the modeling, we originally constructed two different image datasets containing a total of 1,110 images of mayflies, which were split into training and validation datasets, and a test dataset was prepared from two different online photo galleries. The developed ViT model achieved reasonable accuracy, reaching 94.2% and 82.9% for the validation and test datasets, respectively. Given the use of a relatively small number of images in the training process, as well as some variations in the visual styles of the test dataset compared to the training dataset, we consider the level of accuracy to be high. Our results are encouraging toward the use of computer vision for taxonomic identification of macroinvertebrates, although there is still a need to develop specific designs and plans for this purpose, which can vary depending on regional differences in biodiversity as well as sampling and survey methods.

Keywords

Macroinvertebrate; Aquatic insect; Machine learning; Pattern recognition; Computer vision

Introduction

Freshwater covers less than 1% of the Earth's surface (Garcia-Moreno et al. 2014) and accounts for only 2.5% of the Earth's water resources (Garcia-Moreno et al. 2014; Oki &Kanae 2006). Despite their relatively small size, freshwater ecosystems support approximately 10% of all known species (Román-Palacios et al. 2022) and provide vital material, non-material, and regulating services for humans (Lynch et al. 2023). However, the marked degradation of freshwater ecosystems and loss of freshwater biodiversity highlight the importance of their conservation (Loh et al. 2005; Poff et al. 2007; Reid et al. 2019; Tickner et al. 2020). To assess the biological/ecological status of freshwaters such as streams and rivers, biomonitoring using algae, macroinvertebrates, fish, and other species is essential. Among these groups, river benthic macroinvertebrates are most frequently used as a bioindicator group worldwide (Birk et al. 2012; Buss et al. 2015; Eriksen et al. 2021; Namba et al. 2020).

Benthic macroinvertebrates have many characteristics that make them useful for biomonitoring, such as their relatively sedentary nature, ease of sampling, and diverse sensitivities to stressors (Buss et al. 2015; Eriksen et al. 2021; Rosenberg et al. 2008), but sorting and taxonomic identification of benthic macroinvertebrate samples are laborious tasks that require the expertise of specialists (Ärje et al. 2020b). Particularly when a single family can be represented by multiple genera and species, species or genus-level identifications usually require the examination of morphological characteristics under a microscope because they are unlikely to be visible in whole-body photographs of individuals. In contrast, family-level identification can often be accomplished with the naked eye. The level of identification required for environmental assessments has been a topic of discussion (Buss et al. 2015; Jones 2008), but many macroinvertebrate indices based on family-level identifications, such as the Biological Monitoring Working Party (BMWP) system (Armitage et al. 1983), are used globally (Buss et al. 2015). Therefore, particularly in cases where a single family comprises multiple genera and species, endeavoring family-level identification through image recognition represents a feasible and pragmatic goal.

Image recognition is one of the most successful advancements in machine learning technology. Thanks to the development of convolution neural networks (CNNs) and their derivative techniques, the recognition performance of computer vision is now as good as human recognition in some cases (He et al. 2016; Russakovsky et al. 2015). Although transformers were originally developed for natural language processing, the vision transformer (ViT) has emerged as a useful technique for image recognition in the field of computer vision. The recognition performance of ViT-based models outperforms that of CNN-based models in some aspects (Dosovitskiy et al. 2021). However, to our knowledge, the use of ViT-based models for the species identification of river macroinvertebrates has not been explored.

While several previous studies have implemented automated species identification for river benthic macroinvertebrates using machine learning techniques through image recognition (Ärje et al. 2020b; Joutsijoki et al. 2014; Larios et al. 2011; Lytle et al. 2010; Milosavljević et al. 2021; Raitoharju et al. 2018), many of these studies used datasets with a limited number of species within a single family (but see Larios et al. 2011; Milosavljević et al. 2021). In this study, our objective was to assess the capability of a ViT model for family-level identification of mayflies (order Ephemeroptera), considering the presence of multiple taxa (i.e., genus/species) within a single family. Specifically, our focus was to evaluate the capacity for classifying the three mayfly families (Baetidae, Ephemerellidae, and Heptageniidae) found commonly in Japanese rivers, as well as another group that contained several other families within the order Ephemeroptera. These three families have varied sensitivities to organic and metal pollution. For example, ephemerellid and heptageniid mayflies are highly responsive to metal and organic pollution in the environment, whereas baetid mayflies are often found in metal- and organic-contaminated rivers (Armitage et al. 1983; Iwasaki et al. 2018a; Iwasaki et al. 2018b). Recently, an attempt was made to assess the levels of ecological impacts in metal-contaminated rivers mainly on the basis of changes in the abundances or presence/absence of these three families (Iwasaki et al. 2023), indicating their importance as indicators for environmental impact assessments.

Materials and Methods

Dataset construction

The image datasets used in this study were originally created by two individuals—A. Tamada (T-dataset) and M. Monma (M-dataset)—who independently collected and identified mayflies, along with other insects, from rivers in Japan for their own personal interests. Although the mayflies were identified generally to genus or species level, we used the family-level results in this study. The images of mayflies were captured either directly with a digital camera or using a digital camera attached to stereo microscope, manually labeled, and saved in JPG format. Both datasets consisted of close-up photos of insects captured mostly from an overhead perspective, either against a black background (T-dataset, Fig. 1a) or with a ruler in the background (M-dataset, Fig. 1b). Both datasets contained images of 8 aquatic insect families in the order Ephemeroptera: Baetidae, Ephemerellidae, Heptageniidae, Ameletidae, Ephemeridae, Leptophlebiidae, Oligoneuriidae, and Siphlonuridae. Because both datasets had more images of the former three families than of the latter four families, the latter families were combined into one class (D, other mayflies; Table 1). The T- and M-datasets included 5 and 12 taxa (unique species or genus) in Baetidae, 17 and 13 taxa in Ephemerellidae, and 13 and 14 taxa in Heptageniidae, respectively.

The images were randomly divided: 80% into a training dataset and 20% into a validation dataset for each class in each dataset. The respective training and validation datasets were then combined (Table 1). The total number of images in the training dataset (N_{train}) was 885, and the total in the validation dataset (N_{val}) was 225. To examine the potential impact of dataset composition, five train/val datasets with different configurations were prepared by randomly selecting images for training and validation splits.

We also prepared a test dataset consisting of images collected from two different online photo galleries (Table 2). The styles of the images in these galleries differed from those of the images used for training; for instance, the background was white/gray or blue. As above, the "other mayflies" group included five families (Ameletidae, Ephemeridae, Leptophlebiidae, Oligoneuriidae, and Siphlonuridae).

Mayfly classification model

To develop a recognition model for classifying mayfly families, we adopted the fine-tuning technique, which additionally trains the "classifier" of the pre-trained model and is an effective way to build a recognition model of a specific dataset that is difficult to scale up (Brigato et al. 2022). A schematic illustration of the three steps used to construct the mayfly classification model is shown in Fig. 2. Here, we used the pre-trained ViT model provided by Kataoka et al. (2022), which was constructed in two steps (Fig. 2). In the first step, a model was trained from scratch with a fractal database created by using mathematical information of fractal images. In the second step, the model was fine-tuned with ImageNet, which is a large-scale real image dataset. In these steps, the model gains the ability to recognize real images. In the third step, the pre-trained ViT model was further fine-tuned with our mayfly dataset. The latter fine-tuning step was performed using the training codes provided by Kataoka et al. (2022)

(available at https://github.com/masora1030/CVPR2022-Pretrained-ViT-PyTorch), with the default settings, except for the number of classes and the method used to calculate accuracy. The training code was modified to calculate the "top-1" accuracy of 4-class classification task. The top-1 accuracy was calculated as the ratio of the number of images predicted correctly by the class with the highest confidence score to the total number of validation images (N_{val}). Similar to CNN-based image recognition, ViT-based models output a confidence score for each class. This score indicates the probability that the image belongs to a particular class. The class with the highest confidence score was assigned as the predicted class. In the default settings of the training code, the images input into the pre-trained model are resized to 224 × 224 × 3 RGB images. During training, additional data argumentation methods were applied to the resized images. They were randomly cropped after they had been randomly converted with different aspect ratios, and then the brightness, contrast, and saturation of color (i.e., color jitter) were also randomly changed. During validation, the resized images were directly input to the model without data argumentation. The recognition task with the test dataset was performed by using the timm library (Wightman 2019).

Results and Discussion

Mayfly classification model

Results obtained from one of the five randomly prepared train/val datasets are shown in Fig. 3; similar results were obtained from the other datasets (accuracy: 91.7–94.3%). The learning curves of loss function for training and validation converged to a low value (Fig. 3a), indicating that the fine tuning was successful. The confusion matrix (Fig. 3b) shows the number of images predicted for individual classes. For example, in class A, 54 images were placed correctly, and 1 was placed incorrectly in each class C and class D. The top-1 accuracy was 94.2%. The confidence score and the predicted class of all images in the validation dataset are summarized in Fig. 3c. In classes with true labels A (Baetidae) and C (Heptageniidae), 83% and 87% of the images, respectively, had confidence scores above 80%, and they all were classified correctly. In the case of true label B (Ephemerellidae), 77% of the images had confidence scores above 80%, but a few images with high confidence scores (>80%) were misclassified. In the case of true label D (other mayflies), only 54% of the images had high confidence scores. Of all the misclassified images, 69% had confidence scores below 70%.

Example images of classes A, B, and C in the validation dataset are shown in Fig. 4. The distinctions between correctly classified and misclassified images were not readily apparent. Some misclassified images notably lacked the head portion of the mayflies (IDs 73, 91, 134, and 141) due to the resizing of the original images in the default setting of the training code. In all cases, the original image in the dataset included the entire body of the mayflies, and these resized images were also used for fine-tuning. Despite using a relatively small number of images in the training process and employing the default setting (i.e., both the fine-tuning and validation were performed with the resized images, including ones missing body parts), a high recognition rate was attained. This is likely an advantage of using a well pre-trained ViT model on a large natural image dataset such as ImageNet.

To evaluate the recognition performance on entirely new images that the model had not been exposed to during training, we used the test dataset, and the results of the classification are shown in Fig. 5. Despite some variations in the visual styles between the test and training datasets, the classification accuracy remained high (82.9%). The overall decrease in accuracy was attributed to that in classifying class D because the accuracies for classes A–C were almost identical to those achieved on the validation dataset. Thus, we conclude that the model we built can provide a reasonable level of accuracy for classifying three common mayfly families.

Future Directions

This is the first study to develop a ViT model for the identification of Japanese mayfly families. Despite the reasonable level of model accuracy attained, there are issues that need to be addressed for practical implementation. First, to assess biological conditions on the basis of family-level macroinvertebrate identification (Paisley et al. 2014; Torii et al. 2023; Wright 2000) using the ViT model, it is necessary to include other families within mayflies (Ephemeroptera), as well as other orders (e.g., Plecoptera, Trichoptera, and Diptera), which are commonly found in river macroinvertebrate surveys. Second, although biological assessments based on the family-level identification of macroinvertebrates can provide the appropriate amount of information for a given purpose (Jones 2008; Wright & Ryan 2016), more detailed assessments may require identifications to the genus and species levels. As previously noted (Joutsijoki et al. 2014; Raitoharju et al. 2018), a critical barrier to address these two issues is the lack of appropriate images for model development. To this end, the original images used in this study have been made available (see the GitHub website at https://github.com/yuichiwsk/images_mayfly_families). Species- and genus-level identifications also require checking smaller and more detailed morphological characteristics (see, e.g., Merrit et al. 2019), many of which are not visible in the overhead-perspective images used in this study. Thus, such identifications using computer vision and acquiring relevant images become even more challenging. Finally, since the mayfly images used in this study were manually captured from fixed directions, the implementation of semi- or fully-automated imaging and identification poses a further challenge for integrating computer vision into biological assessments based on macroinvertebrates (but see Ärje et al. 2020a; Jaballah et al. 2023; Raitoharju et al. 2018). Developing specific designs and plans about how to use computer vision in macroinvertebrate identification, which may vary depending on diverse regional differences in inherent biodiversity as well as sampling and survey methods, is a fundamental challenge that needs to be addressed in future studies.

Data availability

All the original image data used for developing the vision transformer model (i.e., training and validation datasets) are available on the GitHub website (https://github.com/yuichiwsk/images_mayfly_families).

Acknowledgements

We are deeply grateful to K. Monma for providing the mayfly images. During the preparation of this study, the authors used ChatGPT to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

Yuichi Iwasaki: Conceptualization, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. Hiroko Arai: Conceptualization, Methodology, Software, Formal analysis, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Akihiro Tamada: Investigation, Data Curation, Writing - Review & Editing. Hirokatsu Kataoka: Software, Writing - Review & Editing, Supervision.

References

Ärje J, Melvad C, Jeppesen MR, Madsen SA, Raitoharju J, Rasmussen MS, Iosifidis A, Tirronen V, Gabbouj M, Meissner K, Høye TT (2020a) Automatic image-based identification and biomass estimation of invertebrates. Methods in Ecology and Evolution 11:922–931. doi:10.1111/2041-210X.13428

Ärje J, Raitoharju J, Iosifidis A, Tirronen V, Meissner K, Gabbouj M, Kiranyaz S, Kärkkäinen S (2020b) Human experts vs. machines in taxa recognition. Signal Process Image Commun 87:115917. doi: 10.1016/j.image.2020.115917

Armitage PD, Moss D, Wright JF, Furse MT (1983) The performance of a new biological water-quality score system based on macroinvertebrates over a wide-range of unpolluted running-water sites. Water Res 17:333–347. doi:10.1016/0043-1354(83)90188-4

Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund WV, Zampoukas N, Hering D (2012) Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18:31–41. doi:10.1016/j.ecolind.2011.10.009

Brigato L, Barz B, L. I, J. D (2022) Image classification with small datasets: overview and benchmark. IEEE Access 10:49233–49250. doi:10.1109/ACCESS.2022.3172939

Buss DF, Carlisle DM, Chon T-S, Culp J, Harding JS, Keizer-Vlek HE, Robinson WA, Strachan S, Thirion C, Hughes RM (2015) Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environmental Monitoring and Assessment 187:4132. doi:10.1007/s10661-014-4132-8

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021): An image is worth 16x16 words: Transformers for image recognition at scale, International Conference on Learning Representations. https://openreview.net/forum?id=YicbFdNTTy

Eriksen TE, Brittain JE, Søli G, Jacobsen D, Goethals P, Friberg N (2021) A global perspective on the application of riverine macroinvertebrates as biological indicators in Africa, South-Central America, Mexico and Southern Asia. Ecological Indicators 126:107609. doi: 10.1016/j.ecolind.2021.107609

Garcia-Moreno J, Harrison IJ, Dudgeon D, Clausnitzer V, Darwall W, Farrell T, Savy C, Tockner K, Tubbs N (2014): Sustaining Freshwater Biodiversity in the Anthropocene. In: Bhaduri A, Bogardi J, Leentvaar J , Marx S (Editors), The Global Water System in the Anthropocene: Challenges for Science and Governance. Springer International Publishing, Cham, Switzerland, pp. 247–270

He K, Zhang X, Ren S, Sun J (2016): Deep residual learning for image recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778

Iwasaki Y, Kagaya T, Matsuda H (2018a) Comparing macroinvertebrate assemblages at organic-contaminated river sites with different zinc concentrations: Metal-sensitive taxa may already be absent. Environ Pollut 241:272–278. doi:10.1016/j.envpol.2018.05.041

Iwasaki Y, Schmidt TS, Clements WH (2018b) Quantifying differences in responses of aquatic insects to trace metal exposure in field studies and short-term stream mesocosm experiments. Environ Sci Technol 52:4378–4384. doi:10.1021/acs.est.7b06628

Iwasaki Y, Mano H, Shinohara N (2023) Linking levels of trace-metal concentrations and ambient toxicity to cladocerans to levels of effects on macroinvertebrate communities. Environ Adv 11:100348. doi:10.1016/j.envadv.2023.100348

Jaballah S, Garcia GF, Martignac F, Parisey N, Jumel S, Roussel J-M, Dézerald O (2023) A deep learning approach to detect and identify live freshwater macroinvertebrates. Aquatic Ecology 57:933–949. doi:10.1007/s10452-023-10053-7

Jones FC (2008) Taxonomic sufficiency: The influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. Environ Rev 16:45–69. doi:10.1139/A07-010

Joutsijoki H, Meissner K, Gabbouj M, Kiranyaz S, Raitoharju J, Ärje J, Kärkkäinen S, Tirronen V, Turpeinen T, Juhola M (2014) Evaluating the performance of artificial neural networks for the classification of freshwater benthic macroinvertebrates. Ecol Inform 20:1–12. doi:10.1016/j.ecoinf.2014.01.004

Kataoka H, Hayamizu R, Yamada R, Nakashima K, Takashima S, Zhang X, Martinez-Noriega EJ, Inoue

N, Yokota R (2022): Replacing labeled real-image datasets with auto-generated contours, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 21232–21241

Larios N, Lin J, Zhang M, Lytle D, Moldenke A, Shapiro L, Dietterich T (2011): Stacked spatial-pyramid kernel: An object-class recognition method to combine scores from random trees, 2011 IEEE Workshop on Applications of Computer Vision (WACV), pp. 329–335

Loh J, Green RE, Ricketts T, Lamoreux J, Jenkins M, Kapos V, Randers J (2005) The Living Planet Index: using species population time series to track trends in biodiversity. Philos Trans R Soc B 360:289–295. doi:10.1098/rstb.2004.1584

Lynch AJ et al. (2023) People need freshwater biodiversity. WIREs Water:e1633. doi:10.1002/wat2.1633

Lytle DA, Martínez-Muñoz G, Zhang W, Larios N, Shapiro L, Paasch R, Moldenke A, Mortensen EN, Todorovic S, Dietterich TG (2010) Automated processing and identification of benthic invertebrate samples. J N Am Benthol Soc 29:867–874. doi:10.1899/09-080.1

Merrit RW, Cummins KW, Berg MB (2019): An introduction to the aquatic insects of North America. Kendall Hunt Publishing, Dubque, Iowa, 1480 pp

Milosavljević A, Milošević Đ, Predić B (2021) Species identification for aquatic biomonitoring using deep residual cnn and transfer learning. Facta Universitatis, Series: Automatic Control and Robotics. doi:2010.22190/FUACR201118001M

Namba H, Iwasaki Y, Heino J, Matsuda H (2020) What to survey? A systematic review of the choice of biological groups in assessing ecological impacts of metals in running waters. Environ Toxicol Chem 39:1964–1972. doi:10.1002/etc.4810

Oki T, Kanae S (2006) Global hydrological cycles and world water resources. Science 313:1068–1072. doi:10.1126/science.1128845

Paisley MF, Trigg DJ, Walley WJ (2014) Revision of the biological monitoring working party (BMWP) score system: Derivation of present-only and abundance-related scores from field data. River Res Appl 30:887–904. doi:10.1002/rra.2686

Poff NL, Olden JD, Merritt DM, Pepin DM (2007) Homogenization of regional river dynamics by dams and global biodiversity implications. Proc Nat Acad Sci USA 104:5732–5737. doi:10.1073/pnas.0609812104

Raitoharju J, Riabchenko E, Ahmad I, Iosifidis A, Gabbouj M, Kiranyaz S, Tirronen V, Ärje J, Kärkkäinen S, Meissner K (2018) Benchmark database for fine-grained image classification of benthic macroinvertebrates. Image Vis Comput 78:73–83. doi:10.1016/j.imavis.2018.06.005

Reid AJ, Carlson AK, Creed IF, Eliason EJ, Gell PA, Johnson PTJ, Kidd KA, MacCormack TJ, Olden JD, Ormerod SJ, Smol JP, Taylor WW, Tockner K, Vermaire JC, Dudgeon D, Cooke SJ (2019) Emerging threats and persistent conservation challenges for freshwater biodiversity. Biological Reviews 94:849–873. doi:10.1111/brv.12480

Román-Palacios C, Moraga-López D, Wiens JJ (2022) The origins of global biodiversity on land, sea and freshwater. Ecol Lett 25:1376–1386. doi:10.1111/ele.13999

Rosenberg DM, Resh VH, King RS (2008): Use of aquatic insects in biomonitoring. In: Merritt RW, Cummins KW , Berg MB (Editors), An Introduction to the Aquatic Insects of North America. Kendall Hunt, Dubuque, IA, pp. 123–138

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115:211–252. doi:10.1007/s11263-015-0816-y

Tickner D et al. (2020) Bending the Curve of Global Freshwater Biodiversity Loss: An Emergency Recovery Plan. Bioscience 70:330–342. doi:10.1093/biosci/biaa002

Torii T, Abe E, Tare H, Tsuzuki T, Myosho T, Kobayashi T (2023) Prediction of average score per taxon in Japan using mega data from the national census on river environments. Limnology10.1007/s10201-023-00729-2

Wightman R (2019): PyTorch Image Models. GitHub, GitHub repository

Wright IA, Ryan MM (2016) Impact of mining and industrial pollution on stream macroinvertebrates: importance of taxonomic resolution, water geochemistry and EPT indices for impact detection. Hydrobiologia 772:103–115. doi:10.1007/s10750-016-2644-7

Wright JF (2000): An introduction to RIVPACS. In: Wright JF, Sutcliffe DW, Furse MT (Editors), Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques. Freshwater Biological Association, Ableside, UK, pp. 1–24

Table 1. Number of images in each class in the T- and M-datasets.

Class	T-dataset	M-dataset	Total (train/val)
A: Baetidae	56	219	275 (219/56)
B: Ephemerellidae	130	182	312 (249/63)
C: Heptageniidae	137	212	349 (278/71)
D: Other mayflies	100	74	174 (139/35)

Classes were assigned in this study. The images were randomly divided in an 80 (for training, train) to 20 (for validation, val) split for each class in each dataset.

Table 2. Number of images for the test dataset.

Class	Gallery 1	Gallery 2	Total
A: Baetidae	12	6	18
B: Ephemerellidae	12	19	31
C: Heptageniidae	19	20	39
D: Other mayflies	15	17	29

The images were collected from two online photo galleries (Gallery 1: http://museinfo.hitohaku.jp/kawamushi/zukan/kagerou.html; Gallery 2: https://www.eonet.ne.jp/~suiseikontyu/).

Figure captions

Fig. 1. Example images from the (a) T-dataset and (b) M-dataset.

Fig. 2. Schematic illustration of the steps to construct a mayfly classification model.

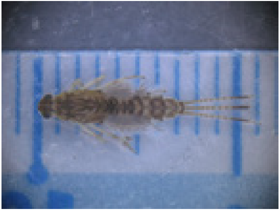
Fig. 3. (a) Loss function learning curves for training (red) and validation (blue) datasets. (b) Confusion matrix of the 4-class classification in the validation dataset. The numbers denote the number of images belonging to the predicted class, and blue indicates a correct classification. (c) Summary of inferences for the validation dataset. Colored regions indicate the true labels for each image: class A (ID 1–56), class B (ID 57–119), class C (ID 120–190), and class D (ID 191–225). The point data are the predicted labels, as defined in the legend.

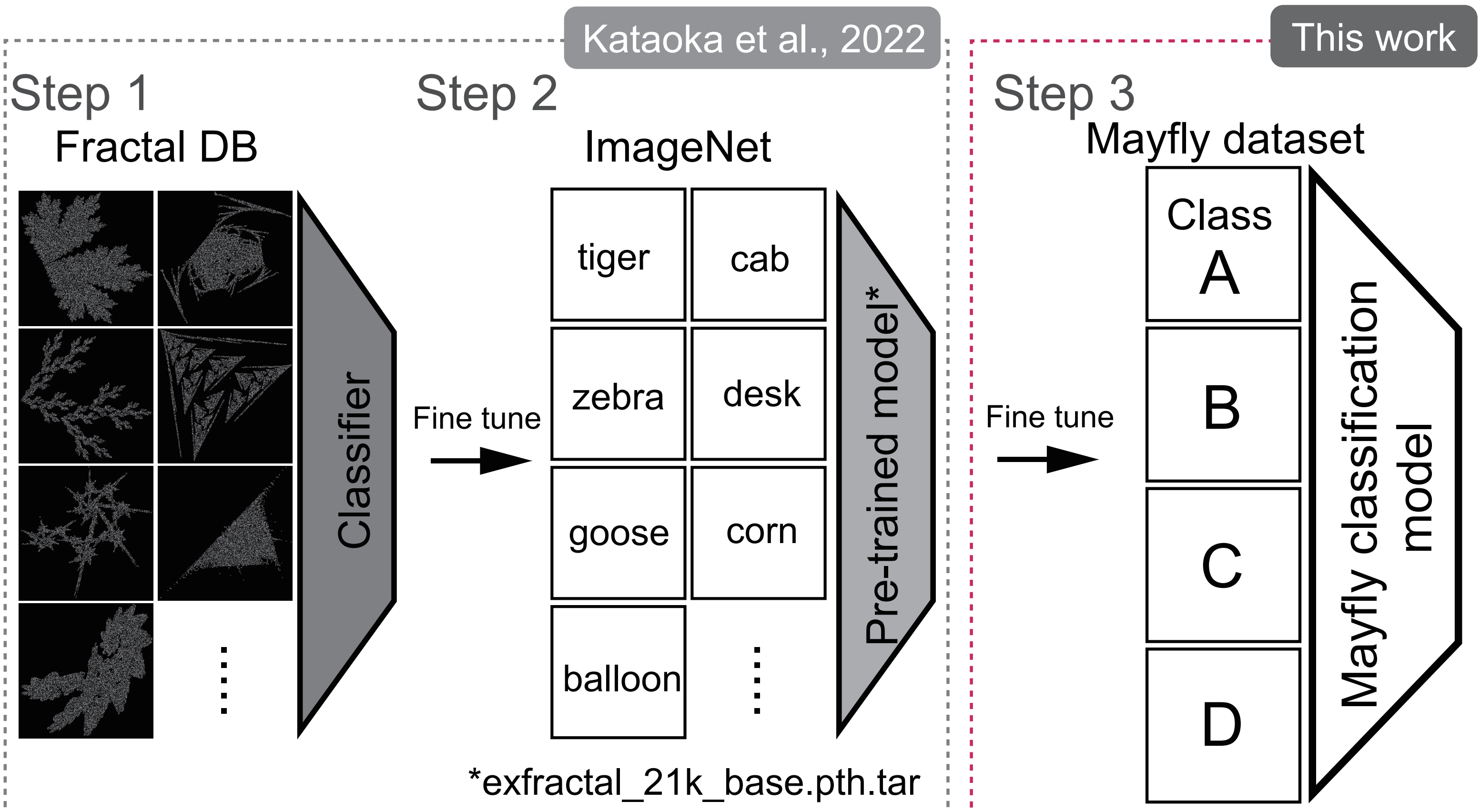
Fig. 4. Example images from classes A–C in the validation dataset. ID numbers are in white; underline indicates misclassification. The incorrectly predicted class and confidence score are shown below each misclassified image.

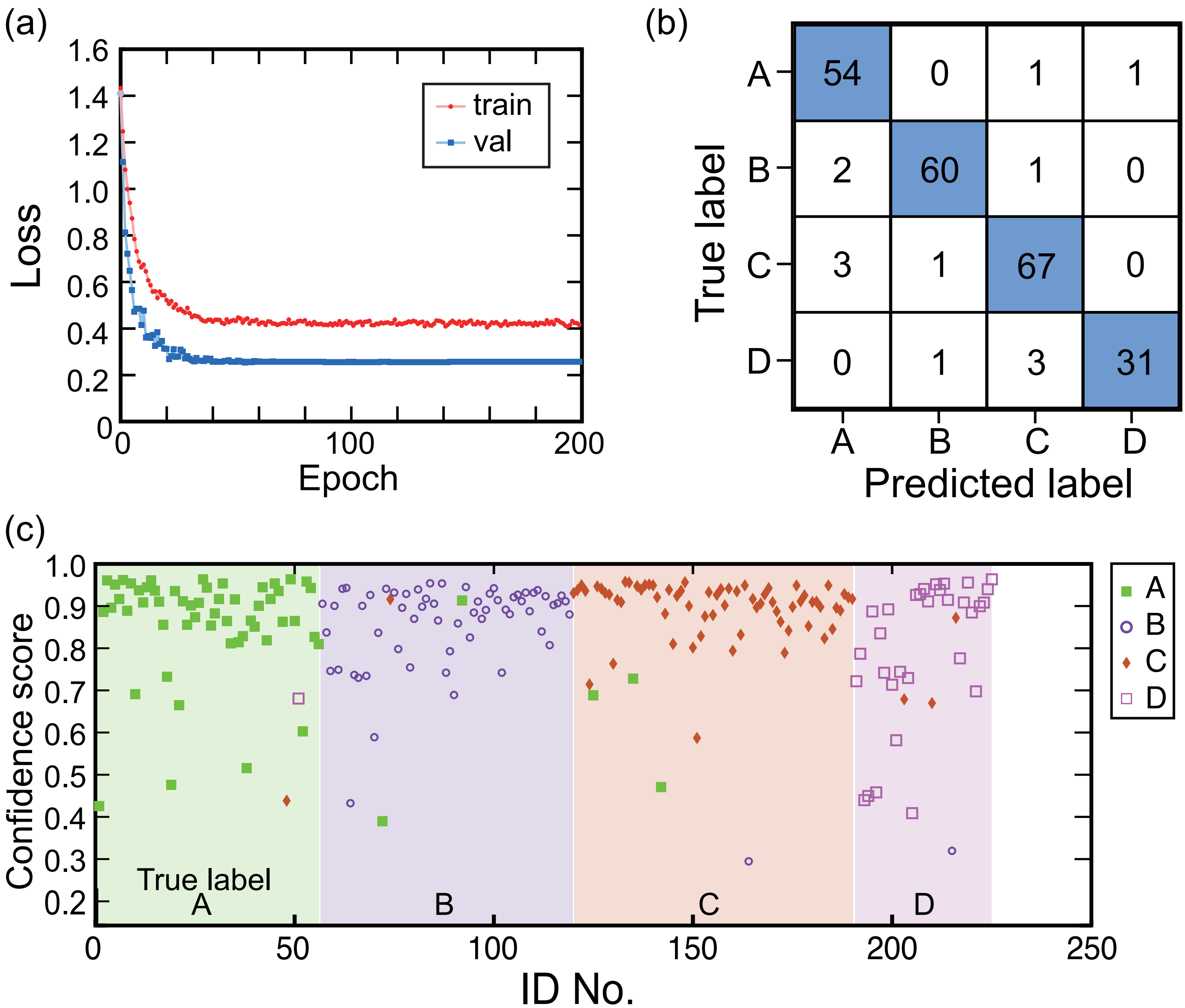
Fig. 5. Confusion matrix for the classification task in the test dataset. Top-1 accuracy was 82.9%.

(a)

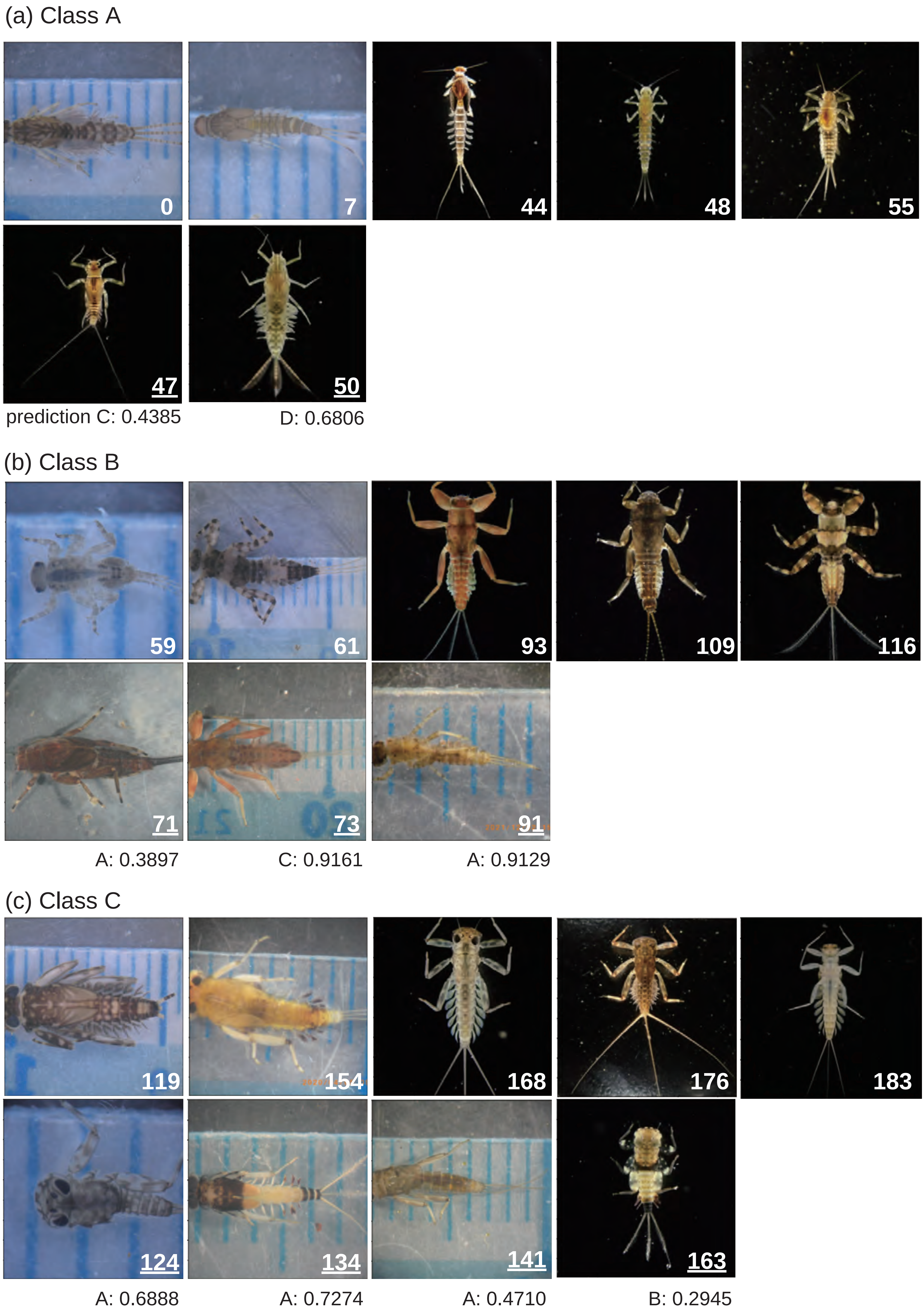
(b)

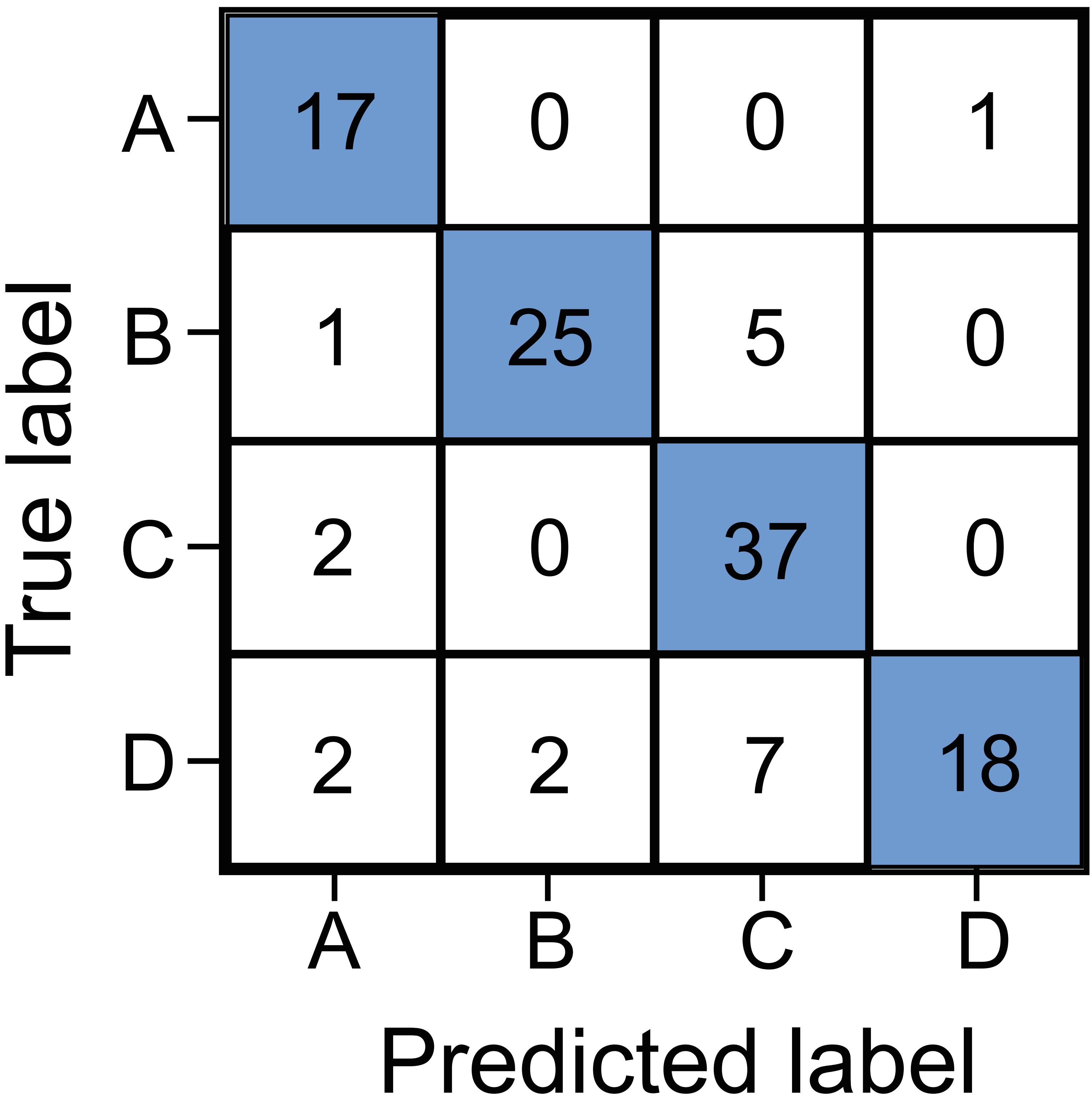



Kataoka et al., 2022
This work
Step 1
Fractal DB
Classifier
Fine tune
Step 2
ImageNet
tiger
cab
zebra
desk
goose
corn
balloon
Pre-trained model*
*exfractal_21k_base.pth.tar
Fine tune
Step 3
Mayfly dataset
Class
A
B
C
D
Mayfly classification model



(a)
Loss
Epoch
train
val
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
0
100
200
(b)
True label
Predicted label
A
B
C
D
54 0 1 1
2 60 1 0
3 1 67 0
0 1 3 31
(c)
Confidence score
ID No.
True label
A
B
C
D
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
50
100
150
200
250


(a) Class A
0
7
44
48
55
47
50
prediction C: 0.4385
D: 0.6806
(b) Class B
59
61
93
109
116
71
73
91
A: 0.3897
C: 0.9161
A: 0.9129
(c) Class C
119
154
168
176
183
124
134
141
163
A: 0.6888
A: 0.7274
A: 0.4710
B: 0.2945


True label
Predicted label
A
B
C
D
A
B
C
D
17
0
0
1
1
25
5
0
2
0
37
0
2
2
7
18