Development of a method for estimating asari clam distribution by combining three-dimensional acoustic coring system and deep neural network

Tokimu Kadoi[1], Katsunori Mizuno[2*], Shoichi Ishida[1], Shogo Onozato[2], Hirofumi Washiyama[3], Yohei Uehara[3], Yoshimoto Saito[4], Kazutoshi Okamoto[4], Shingo Sakamoto[5], Yusuke Sugimoto[5] and Kei Terayama[1,6,7*]

[1]Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045, Kanagawa, Japan.

[2]Department of Environment Systems, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwanoha, Kashiwa, Chiba, 277-8561, Japan.

[3]Shizuoka Prefectural Research Institute of Fishery and Ocean, 5005-3, Bentenjima, Maisaka-cho, Chūō-ku, Hamamatsu-shi, Shizuoka, 431-0214, Japan.

[4]Marine Open Innovation Institute, 2nd Floor, Shimizu Marine Building, 9-25, Hinode-cho, Shimizu-ku, Shizuoka-shi, Shizuoka, 424-0922, Japan.

[5]Windy Network Corporation, 1-19-4, Higashi-Hongo, Shimoda-shi, Shizuoka, 415-0035, Japan.

[6]RIKEN Center for Advanced Intelligence Project, 1-4-1, Nihonbashi, Chuo-ku, 103-0027, Tokyo, Japan.

[7]MDX Research Center for Element Strategy, Tokyo Institute of Technology, 4259, Nagatsuta-cho, Midori-ku, Yokohama, 226-8501, Kanagawa, Japan.

*Corresponding author(s).

 E-mail(s): kmizuno@edu.k.u-tokyo.ac.jp and terayama@yokohama-cu.ac.jp

Abstract

Developing non-contact, non-destructive monitoring methods for marine life is crucial for sustainable resource management. Recent monitoring technologies and machine learning analysis advancements have enhanced underwater image and acoustic data acquisition. Systems to obtain 3D acoustic data

from beneath the seafloor are being developed; however, manual analysis of large 3D datasets is challenging. Therefore, an automatic method for analyzing benthic resource distribution is needed. This study developed a system to estimate benthic resource distribution non-destructively by combining high-precision habitat data acquisition using high-frequency ultrasonic waves and prediction models based on a 3D convolutional neural network (3D-CNN). The system was applied to asari clams (*Ruditapes philippinarum*), whose population has been declining in recent years in Japan. Clam presence and count were successfully estimated in a voxel with an ROC-AUC of 0.9 and a macro-average ROC-AUC of 0.8, respectively. This system visualized clam distribution and estimated numbers, demonstrating its effectiveness for quantifying marine resources beneath the seafloor.

Introduction

The seafloor harbors numerous benthic organisms, which are indispensable as marine resources targeted by fisheries and for marine ecosystems and material cycles[1,2]. However, as most benthic organisms are always concealed in sediments, determining their population size and observing their behavior is difficult. Surveys are inevitably destructive, time-consuming, and costly. Conventional sampling methods cannot cover large areas, making it difficult to detect changes over time. This limitation is an impediment to sustainable management of sub-benthic resources and environment.

Benthic organisms such as asari clam (Japanese littleneck clam, *Ruditapes philippinarum*) have become a concern regarding stock management[3]. The asari clam is a bivalve that lives in the shallows of inner bays and is an important target fishery species. The harvest of asari clams in Japan has been on a downward trend, with a significant decrease from 160,000 tons in the 1980s to less than 10,000 tons since 2016. Various factors have been revealed as causes for this trend, including overfishing, feeding damage, disease and insect damage, and the prey environment[4]. Monitoring of clams has recently been conducted with the aim of sustainable fisheries management and resource conservation; however, this relies on manual digging and counting[5]. The current methods are very costly, and comprehensively monitoring the dynamics trends in the asari clam ecosystem over time is difficult.

Various monitoring methods have recently been developed by linking technological improvements in image and acoustic data acquisition with machine learning analysis, including deep learning[6–10]. For example, two-dimensional (2D) imaging of tracer particles[11] and three-dimensional (3D) computed tomography (CT) imaging[12] have been developed as data acquisition methods under the seafloor. Acoustic systems with various operating frequencies are commonly used to detect objects buried in seafloor sediments. For instance, wooden wrecks buried on the seafloor can be visualized using chirp signals with sweep pulses ranging from 1.5 to 13 kHz[13]. Recently, a new monitoring tool, a 3D acoustic coring system, has been developed and used to precisely survey buried roots of plants with an outer diameter of 3 to 5 cm using ultrasonic waves with a center frequency of 100 kHz[14]. In addition, high-frequency signals with a center frequency of 1 MHz have recently been used to survey small creatures

of 3–5 cm, such as the clams[15]. Other deep learning analysis methods have been proposed for monitoring seafloor data, such as corals and seaweeds[6–8]. Furthermore, attempts have been made to classify organisms under the seafloor in the laboratory from data acquired with the 3D acoustic coring system[9]. However, such analysis methods need to be verified outside of the laboratory for the practical use of the 3D acoustics to estimate the distribution of targets in the obtained 3D data and the validation of such methods for benthic resource managements.
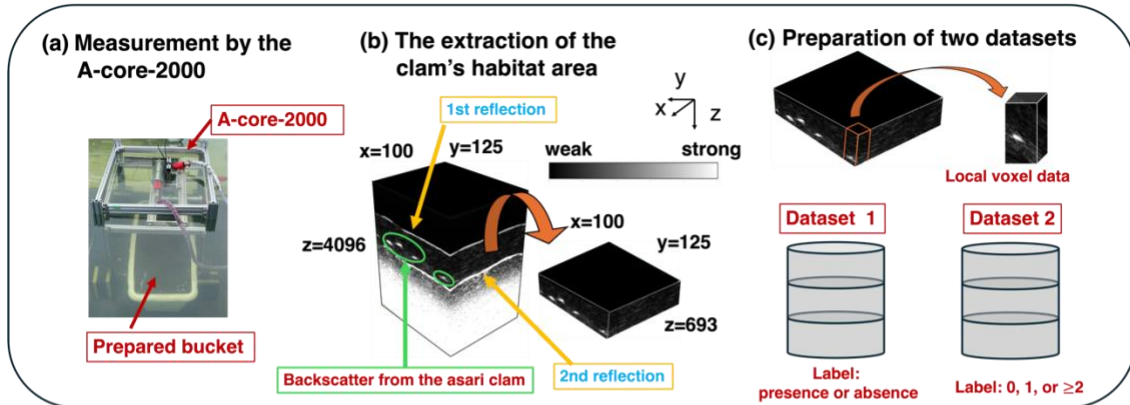
In this study, we developed a system to estimate the distribution of asari clams in a non-contact and non-destructive manner. Our system first acquires 3D benthic data containing clams using the 3D acoustic coring system. Then, the system predicts whether and how many clams are present in a local voxel region using a 3D convolutional neural network (3D-CNN)[16], a method that has been used successfully in a wide range of tasks involving 3D data[10,17–22]. The proposed system can also estimate the distribution and number of targets by integrating the prediction results at multiple voxels within a certain region. To validate the proposed system, benthic organisms, such as clams and mussels, were obtained to verify the estimation. We also reported examples of the distribution visualization and estimation results of the count distribution within a certain region. Furthermore, we showed that Gradient-weighted Class Activation Mapping (Grad-CAM)[23], a visual explanation method for interpretation in the prediction of neural network models, can be used to analyze and interpret regions that are important for the prediction of 3D data.
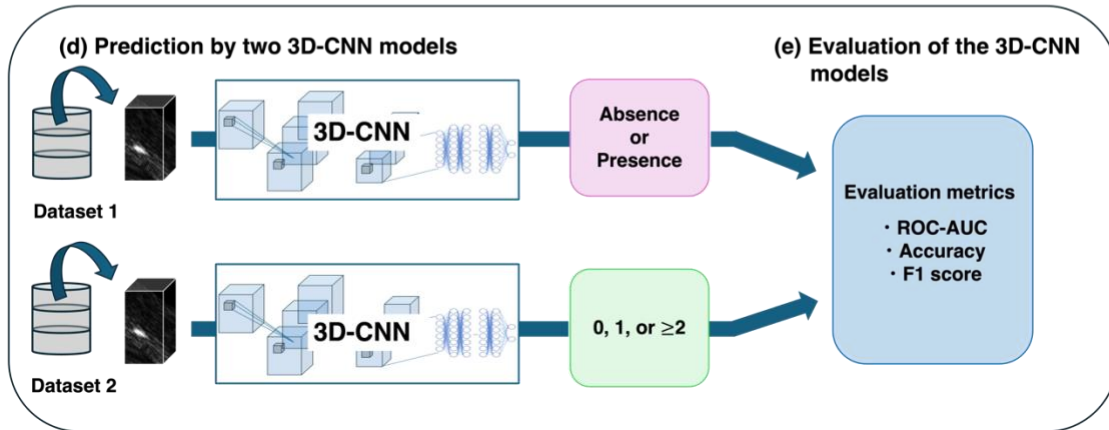
Results and Discussion

Overview

The workflow of this study is shown in Figure 1. The proposed system was divided into two parts: 1) preparation of input data for the deep learning model and 2) prediction by two 3D-CNN models. The 3D-CNN models predicted the presence or absence of clams and the count of clams in the local voxel data.

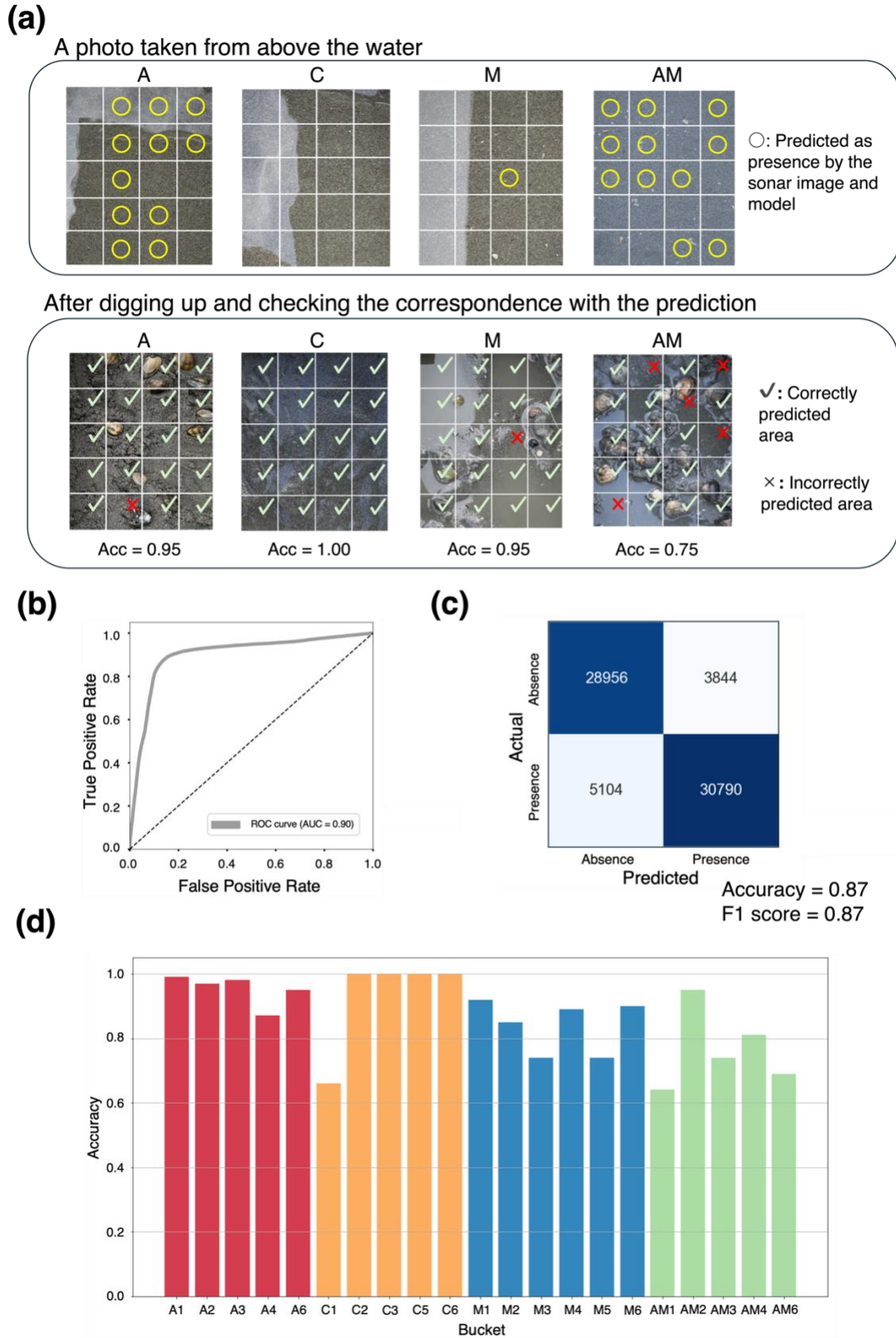## 1) Preparation of input data for the deep learning model

(a) Measurement by the A-core-2000

A-core-2000

Prepared bucket

(b) The extraction of the clam's habitat area

1st reflection

x=100 y=125

z=4096

weak strong

x=100

y=125

z=693

Backscatter from the asari clam

2nd reflection

(c) Preparation of two datasets

Local voxel data

Dataset 1

Dataset 2

Label: presence or absence

Label: 0, 1, or ≥2

2) Prediction by deep learning models

(d) Prediction by two 3D-CNN models

Dataset 1

3D-CNN

Absence or Presence

Dataset 2

3D-CNN

0, 1, or ≥2

(e) Evaluation of the 3D-CNN models

Evaluation metrics
· ROC-AUC
· Accuracy
· F1 score

Figure 1. **The workflow of the proposed system.** This system is divided into two parts: 1) preparation of input data for the deep learning model and 2) prediction using the deep learning model. In 1), the A-core-2000 was first used to measure the reflected waves from the subseafloor (a), extract the areas where clams are present (b), and create datasets on the presence/absence and the count of clams for training and evaluation of the deep learning model (c). In 2), the datasets created in (c) were used to train two independent 3D-CNN models (d), and finally, the 3D-CNN models were evaluated using three evaluation metrics (e).

Model for classifying the presence or absence of clams

First, we constructed the 3D-CNN model to discriminate the presence or absence of clams from the 3D data measured with the A-core-2000. The model training and evaluation were conducted using Stratified Group 5-fold cross-validation with the aforementioned data types A and C, including data types M containing mussels and sand and AM containing clams, mussels, and others (Figure 2(a)). Figure 2(b) and (c) shows the model's performance using ROC-AUC curves and confusion matrices. The prediction achieved an ROC-AUC of 0.90, an accuracy of 0.87, and an F1 score of 0.87. The detailed predictions for each data are shown in Figure 2(d). Data type A containing only clams consistently exhibited high predictive accuracy, with accuracy exceeding 0.8 across all buckets. In contrast, AM and M showed several buckets with accuracy below 0.80, indicating lower overall accuracy than A. Additionally, most of the data in C showed high accuracy, except for C1, which had a value below 0.8. Predictions using only the intensity information of each voxel without 3D-CNN were also performed for comparison. Supplementary Figure 1(a) and (b) shows the classification results using the mean and maximum intensity values for each voxel, respectively, with very low ROC-AUC values of 0.51 and 0.53. These results indicate that the presence or absence can be estimated with a certain level of accuracy using 3D-CNN.

(a)

A photo taken from above the water

○: Predicted as presence by the sonar image and model

After digging up and checking the correspondence with the prediction

A   Acc = 0.95   C   Acc = 1.00   M   Acc = 0.95   AM   Acc = 0.75

✔: Correctly predicted area

✕: Incorrectly predicted area

(b) ROC curve (AUC = 0.90)

(c)

|	Absence	Presence
Absence	28956	3844
Presence	5104	30790

Accuracy = 0.87
F1 score = 0.87

(d)

Figure 2. Prediction performances of the model for predicting the presence or absence of clams in a voxel. The model was trained and evaluated using Dataset 1 (presence or absence of clams). (a)

Predicted distribution examples for each data type. Note that the presence or absence of clams is predicted from the corresponding voxel data, not from these photo images. The prediction is performed for the voxel corresponding to each square on these images. Regions marked with $\times$ indicate failed predictions, and regions marked with ✓ indicate successful predictions. (b) and (c) The prediction accuracy and the ROC curve for each voxel. (d) The prediction accuracy for each bucket.

We used Grad-CAM[23] to analyze where the prediction model focuses on in a 3D acoustic image to determine the presence or absence of clams. Grad-CAM enables the visualization of regions that constitute the basis for the predictions made by the model. Figure 3 shows the original images and the Grad-CAM visualizations averaged every 10 pixels in the z-direction. The upper rows of Figure 3 show examples of successful predictions visualized by Grad-CAM. Figure 3(a) and (b) are cases in which clams are present in the voxel, and (c) and (d) are cases in which clams are absent. Figure 3(a) and (b) shows that when the model predicted the voxel as the presence of clams, it responded strongly to signals that appeared to be clams in relatively deep areas. In contrast, the model focused on shallow areas when predicting the voxel as the absence of clams. The lower row of Figure 3 shows examples of failed predictions, whereas (e) and (f) represent cases where clams were predicted to be present even though there were no clams. Relatively strong signals were observed in the acoustic images; thus, the model responded to these signals and predicted the voxel as the presence of clams. As observed in (i) and (j) in Figure 3, even C, which contained only sand, contained small clams and stones, which may have adversely affected the prediction. Although signals that appeared to be clams were present in the acoustic images and were focused by the model to some extent, the predictions were incorrect. These results suggest that although there remains room for improvement in accuracy, the prediction model focuses on signals that appear to be clams and the prediction of the presence of clams.
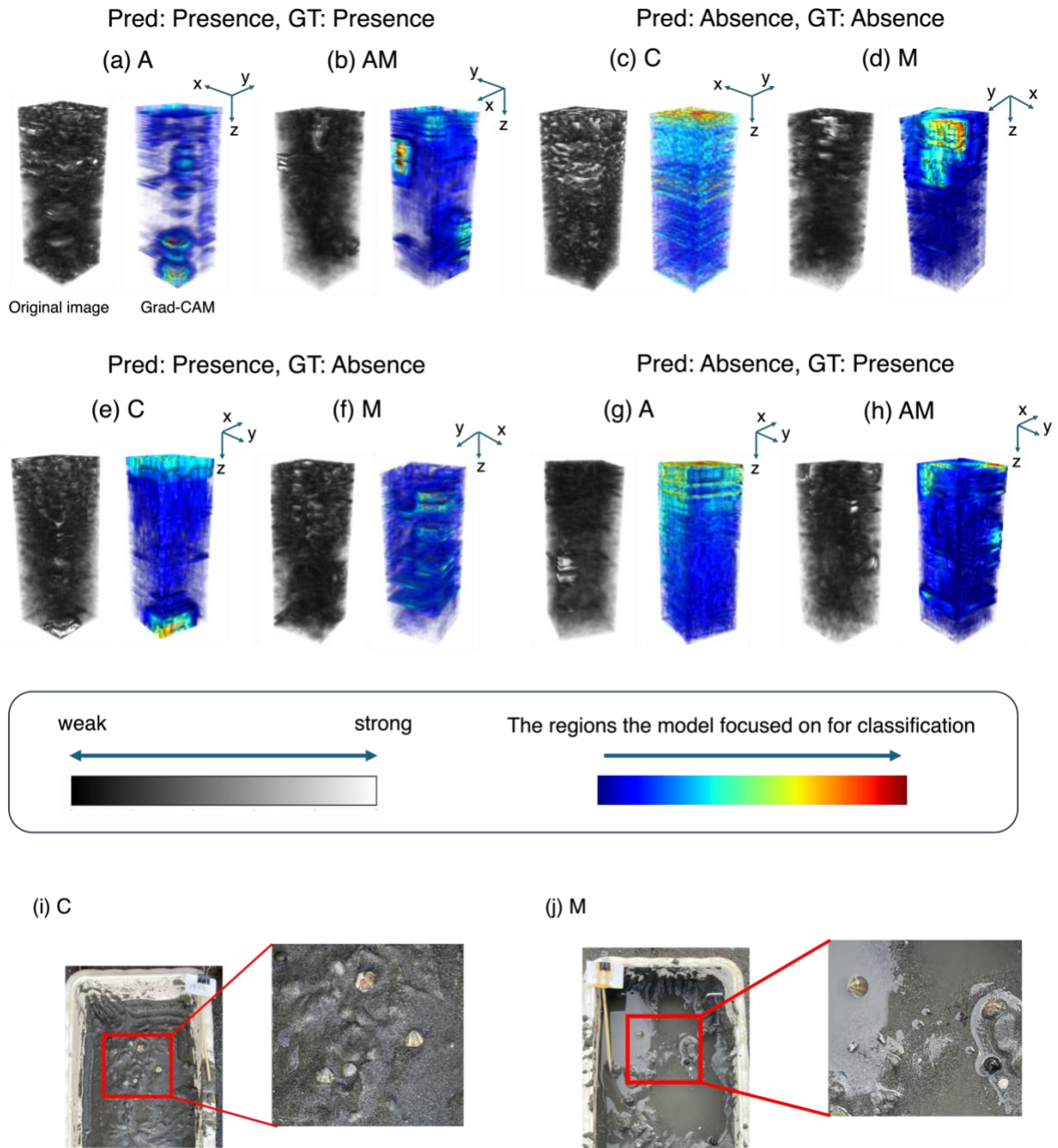
Figure 3. Visualized attention by the prediction model using Grad-CAM. The visualized data were averaged every 10 pixels in the z-direction. (a) to (h) The original input data visualized in grayscale (left) and the corresponding heatmap calculated by Grad-CAM (right). (a) A data where both prediction and actual were "presence." (b) AM data where both prediction and actual were "presence." (c) C data where both prediction and actual were "absence." (d) C data where both prediction and actual were "absence." (e) C data where the actual was "absence" but predicted as "presence." (f) M data where the actual was "absence" but predicted as "presence." (g) A data where the actual was "presence" but
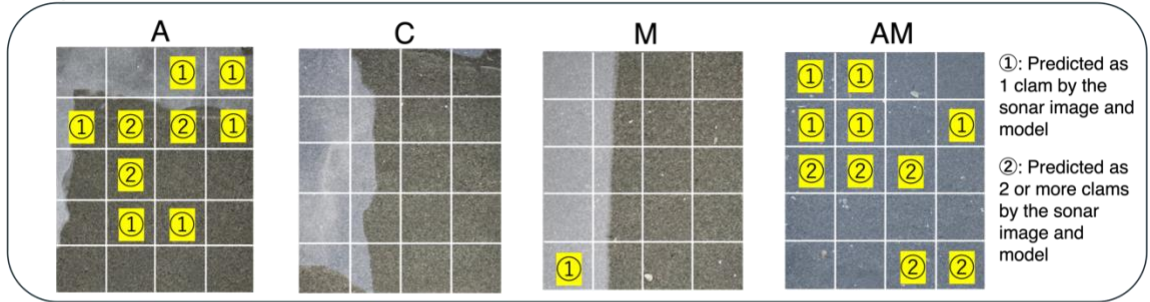
predicted as "absence." (h) AM data where the actual was "presence" but predicted as "absence." (i) Excavation scene for C data where the actual was "absence" but predicted as "presence." (j) Excavation scene for M data where the actual was "absence" but predicted as "presence."

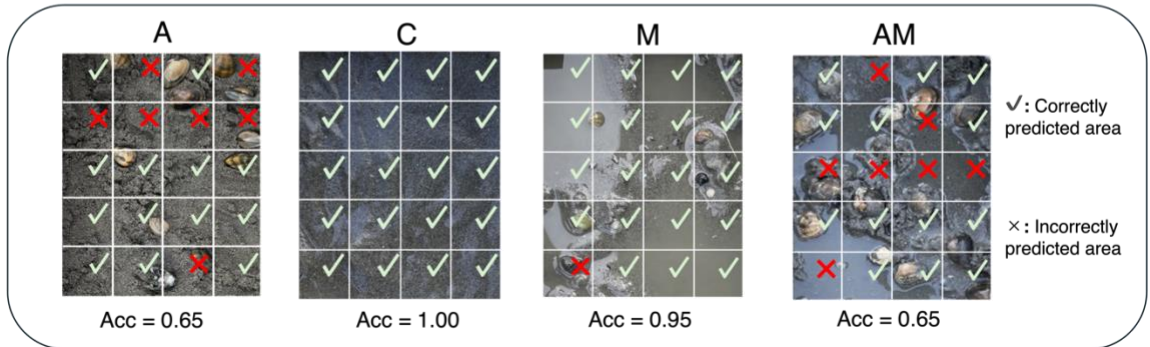**Model for classifying the count of clams**

Next, the model for estimating the count of clams was trained and evaluated. The dataset regarding the count of clams (Table 3) was used, and model training and evaluation were conducted through Stratified Group 5-fold cross-validation. The output was performed as a 3-class classification: "0", "1", and "2 or more." Figure 4(a) shows the prediction examples for each data type. Figure 4(b) and (c) shows the model's performance using ROC-AUC curves and confusion matrix. The prediction achieved a macro-average ROC-AUC of 0.90, an accuracy of 0.64, and a macro-average F1 score of 0.64. The confusion matrix indicated that the discrimination accuracy of "1" and "2 or more" was lower than the prediction accuracy of "0". This is probably because predicting the count of clams becomes difficult when the reflections overlap due to the high density of clams. The prediction results for each data type regarding accuracy are also shown in Figure 4(d). The accuracy of data type C, which contained only sand, was high, whereas that of data types A and M was slightly low. The accuracy of data type AM, containing various objects other than clams, was further reduced. This result suggests that small stones or clams may make classification more difficult, as discussed in the model predicting presence.
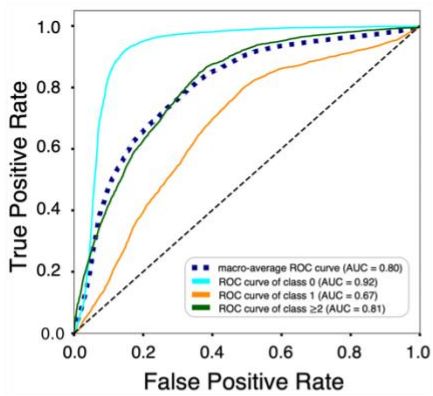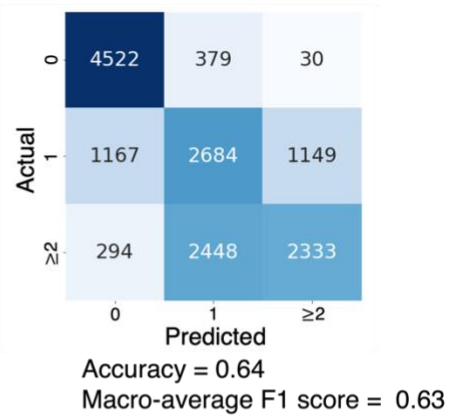
(a)

A photo taken from above the water

A   C   M   AM

①: Predicted as 1 clam by the sonar image and model

②: Predicted as 2 or more clams by the sonar image and model

After digging up and checking the correspondence with the prediction

A   C   M   AM

Acc = 0.65   Acc = 1.00   Acc = 0.95   Acc = 0.65

✓: Correctly predicted area

✗: Incorrectly predicted area

(b)

macro-average ROC curve (AUC = 0.80)
ROC curve of class 0 (AUC = 0.92)
ROC curve of class 1 (AUC = 0.67)
ROC curve of class ≥2 (AUC = 0.81)
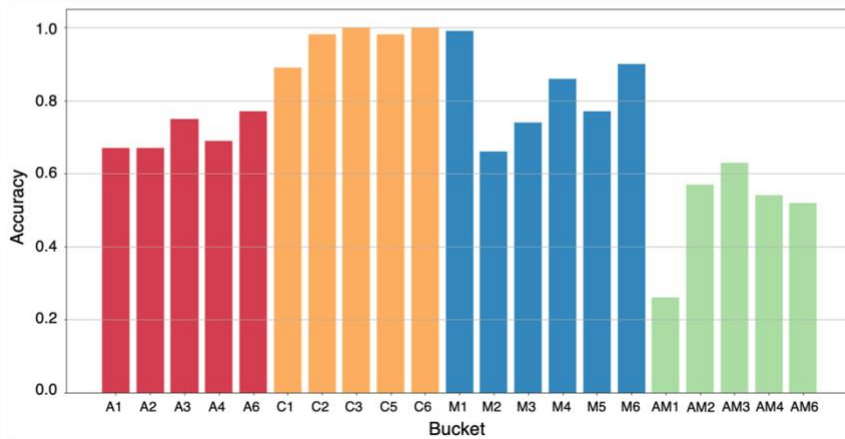
**(c)**

Accuracy = 0.64
Macro-average F1 score = 0.63

(d)

Figure 4. Prediction results of the model for estimating the clams count. The model was trained using Dataset 2 (labels: 0 clams, 1 clam, and 2 or more clams). (a) Examples of predictions for each data type. Regions marked with $\times$ indicate failed predictions, and regions marked with ✓ indicate successful predictions. (b) and (c) The prediction model's performance for each voxel regarding ROC curves and confusion matrix. (d) The prediction accuracy for each bucket.

Estimating the distribution of clams

The proposed method enabled us to estimate the distribution of clams (Figures 2(a) and 4(a)) because the trained models predicted the presence and count of clams in each voxel. Furthermore, the count of clams within a certain region containing multiple voxels was estimated by integrating the prediction results. Here, we evaluated the accuracy of estimating the clams contained within each bucket's region, consisting of 20 voxels (Figure 5). The estimated and measured counts for each bucket are listed in Supplementary Table 1. The overall correlation coefficient was 0.92, and the correlation coefficient for only the A and AM data was 0.68, confirming the correlations of the clam counts. The mean absolute error (MAE) and mean relative error (MRE) for each data type are presented in Table 1. The MRE values were 0.20 for A and 0.12 for AM, indicating that the estimation is possible with an error of approximately 10–20%. These results suggest that the count of clams in each region can also be estimated, although with some error.
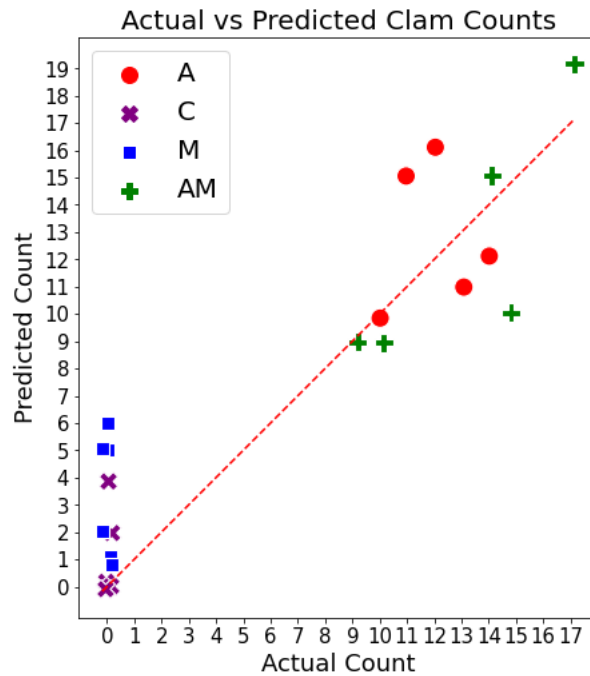
Figure 5. Estimated and actual count of clams per local region (bucket). The count of clams per bucket was calculated by integrating the results predicted by the model to estimate the count of clams.

Table 1. MAE and MRE for predicting the count of clams for each data type. The count of clams was estimated for each bucket using the prediction model of the count of clams. The MAE in the count of clams was calculated for each data type. MRE was calculated only for A and AM because relative absolute error cannot be calculated when the count of clams is zero.

	A	C	M	AM
MAE	2.4	1.2	3.3	1.8
MRE	0.20	-	-	0.12

Conclusion

In this study, we prepared and measured benthic data, including clams, to validate the resource survey method using the three-dimensional acoustic coring system. Using the measured data, two prediction models based on 3D-CNN were trained and evaluated to estimate the presence and count of clams. We successfully estimated the presence or absence of clams in each voxel with an ROC-AUC of approximately 0.9 and the count of clams in local areas with a macro-average ROC-AUC of 0.8. The count of clams within a certain area (bucket) was confirmed to be estimated by MAE values of 0.12 or 0.20. These results indicate the potential of the proposed method for estimating benthic resources.

However, this study has the potential for improvement. First, the prediction accuracy tended to decrease when factors other than clams were introduced, such as AM. To solve this problem, it is beneficial to prepare a larger amount of various data and to use more advanced neural networks, such as vision transformers, to process voxel data for learning. Because the measurement speed of data by A-core-2000 is slow and the measurement area is relatively narrow, considerably faster measuring devices (array sonar system) are needed. As this study was conducted in a relatively controlled environment, verification in the outside field is desired in the future.

Methods

Data preparation and measurement using the acoustic coring system

Preparation of benthic data containing clams. Benthic data, including clams, were obtained from a portable pool at the Hamanako Branch, Shizuoka Prefectural Research Institute of Fishery, Shizuoka Prefecture, Japan. Four types of benthic data were prepared to verify whether estimation could be performed for various types of data: control data containing sand only (C), data containing asari clams and sand (A), data containing a mixture of mussels and sand (M), and data containing a mixture of clams, mussels, and other materials (AM). Each benthic data was prepared by placing sand and mussels from Lake Hamana in a rectangular bucket with internal dimensions of $580 \times 155 \times 200$ mm. Six buckets were prepared for each data type. For the A and AM data, 40 live clams from the nearby Mikawa

Bay were placed in each bucket three days before the measurement date and stored until the measurement date.

Measurement using the acoustic coring system. The A-core-2000[24] was used to observe benthic data in the prepared buckets (Figure 1(a)). The bucket's 250×200 mm area was scanned at 2 mm intervals while continuously irradiated with ultrasound at a center frequency of 500 kHz, and the reflected sound wave was measured and recorded. Envelope processing was performed on the recorded waveform data. The acoustic image constructed by the Viewer[24] showed that the reflection from the clams was between reflections from the soil surface and reflection from the soil surface again (the 2nd reflection in Figure 1(b)). In this study, the range of analysis was defined as the area between the first and second reflections (multiple reflections) from the sediment surface. The total reflection intensity in the XY plane was calculated for each z-coordinate in the 3D data of the acquired reflection intensity, the region of clam presence was extracted, and the size was standardized for each bucket (Sl). Therefore, the 3D data of reflectance intensity measured at $125 \times 100 \times 693$ points in each bucket were used in this study.

Manual identification of positions of asari clams. The positions of asari clams in the measured data were identified to evaluate the prediction models. First, x and y coordinates, the candidate positions of asari clams in the horizontal direction, were manually obtained from the reflection intensity data. Next, the positions of the asari clams were identified by comparing them with the image data of the dug clams taken at the time of measurement. The backscatter of the clams was measured, and the center position of the backscatter was defined as the clam position. The data on the positions of the clams were recorded in Supplementary Tables 2 and 3. The data in A5, C4, and AM5 were excluded from the data set because mechanical issues with the probe resulted in excessive noise, rendering the measurements inaccurate.

Preparation of datasets for model training and evaluation. Datasets 1 and 2 were created to construct models to predict the presence or absence of clams in a local voxel and the count of clams in a local voxel. First, the data obtained from each bucket was divided into 25×25 pixels, shifting every 1 pixel in the horizontal direction, resulting in 124,689 local voxel data, each with size $25 \times 25 \times 693$ pixels

(Figure 1(b)). Here, backscatter data of clams could be included in the boundary of the divided data. Therefore, if the position of a clam recorded above was included in a local voxel or the shortest distance of the position from a local voxel was less than 11 pixels, the local voxel data was regarded as including the clam. Based on the positions of clams identified in the above Section, for Dataset 1, a label of "Absence" was assigned if no clams were included in a local voxel data, and "Presence" was assigned if at least one clam was included in a local voxel data. The number of labeled data is presented in Table 2; for C and M, all data were labeled as "absence" because no clams were included in the buckets. Based on the positions of clams identified in the above Section 3.1.3, for Dataset 2, the count of clams was assigned as a label (Table 3). The number of data containing three or more clams in one local voxel was very small (1,137); therefore, data containing three or more clams were treated as data containing two or more clams. Datasets 1 and 2 were created by randomly extracting data from the 124,689 local voxel data, each with a size of $25 \times 25 \times 693$ pixels; thus, the number of data for each label was approximately the same.

Table 2. Dataset regarding the presence or absence of clams (Dataset 1)

	Absence	Presence
A	2,125	16,717
C	13,000	0
M	15,600	0
AM	2,075	19,177
Total	32,800	35,894

Table 3. Dataset on the count of clams (Dataset 2)

	0	1	≥2
A	1,147	2,500	1,939
C	1,250	0	0
M	1,500	0	0
AM	1,034	2,500	3,136
Total	4,931	5,000	5,075

Classification using deep neural networks

In this study, we employed the 3D-CNN[16] to develop the prediction models. 3D-CNN has demonstrated success in various tasks involving 3D data, such as classification, detection, and segmentation of medical images[10,17–20], as well as action recognition[21,22]. It mainly consists of convolutional, pooling, and fully connected layers. We modified the 3D-CNN developed in our previous study[9] to build two independent models: one for classifying the presence or absence of clams and another for classifying the count of clams. In the model for classifying the presence or absence of clams, we used Dataset 1, containing information on the presence or absence of clams as input and obtained outputs of either 'Absence' or 'Presence' (Figure 6). The model employed the Adam optimizer, with a batch size of three and a learning rate of 10^{-6}, and was trained for a maximum of 50 epochs, stopping if there was no improvement for 10 epochs. For the model classifying the count of clams, we used Dataset 2, containing information on the count of clams as input, and obtained outputs of '0', '1', or '2 or more clams' (Figure 6). The model employed the Adam optimizer, with a batch size of three and a learning rate of 10^{-6}, and was trained for a maximum of 30 epochs, stopping if there was no improvement for 10 epochs.

In this study, TensorFlow (version 2.4.1)[25] was used to construct two 3D-CNN models, and the performance of the models was evaluated using Stratified Group 5-fold cross-validation[26]. The groups

represented each bucket. Additionally, the dataset was divided into training (60%), validation (20%),

and test (20%) sets.
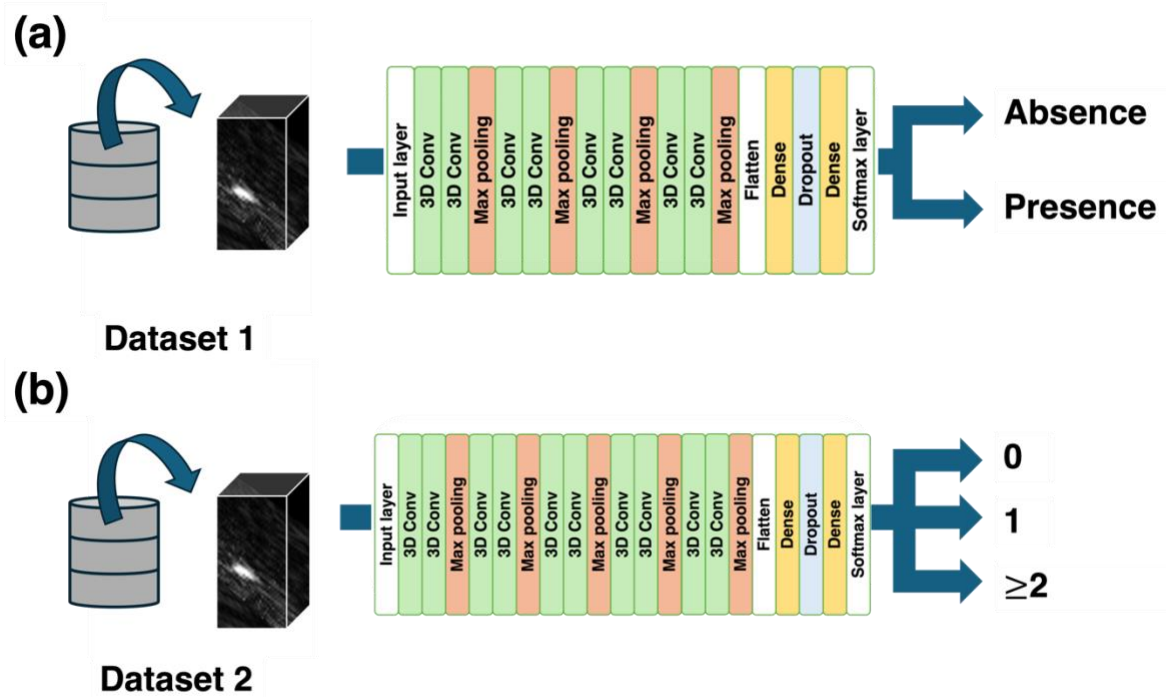
Figure 6. Overview of predictions using 3D-CNN. Training and evaluation of 3D-CNN models were

conducted using Dataset 1 on the presence/absence of clams (a) and Dataset 2 on the count of clams (b).

Evaluation metrics of the prediction model

Three evaluation metrics were employed to evaluate the trained models: accuracy, F1-score, and the

area under the receiver operating characteristic curve (ROC-AUC). Accuracy represents the proportion

of correctly predicted data, while F1-score is defined as the harmonic mean of precision and recall:

$$F1-score = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

ROC-AUC refers to the area under the ROC curve, which plots the true-positive rate against the false-positive rate as the classification threshold varies. Each metric ranges from 0 to 1, with higher values indicating better model performance. Furthermore, to evaluate the model for classifying the count of clams, we used macro-average ROC-AUC and F1-score, which are the averages of the ROC-AUC and F1-score calculated for each class. Macro-average refers to the method of computing the ROC-AUC and F1-score for each class individually and then averaging these scores. This metric considers the performance of each class equally and allows for a balanced evaluation of the overall performance, even in cases with class imbalance.

Data availability

The code for the machine learning is available on GitHub at https://github.com/ycu-iil/Asari3DCNN.

References

1.   Solan, M. *et al.* Extinction and ecosystem function in the marine benthos. *Science* **306**, 1177–1180 (2004).

2.   Danise, S., Twitchett, R. J., Little, C. T. S. & Clémence, M. E. The impact of global warming and anoxia on marine benthic community dynamics: an example from the Toarcian (Early Jurassic). *PLoS One* **8**, e56255 (2013).

3.   Ito, H. What kind of animal is the clam Ruditapes philippinarum? –Introduction to its ecology and fishery. Asari to wa donna ikimono ka: Asari no seitai, oyobi gyogyou seisan no suii (in Japanese). *Jpn J Benthol.* **57**, 134–138 (2002).

4.   Toba, M. Revisiting recent decades of conflicting discussions on the decrease of Asari clam Ruditapes philippinarum in Japan: A review. Asari shigen no genshou ni kansuru giron e no saihou (in Japanese). *Nippon Suisan Gakkaishi* **83**, 914–941 (2017).

5. Murai, M. Trends and considerations on the variation in the number of clams at 'Umi no Kouen.' 「Umi no Kouen」 ni okeru Asari kotaisuu no hendou ni kansuru keikou to kousatsu (in Japanese). *Enkangiki Gakkaishi (Journal of Coastal Zone Studies)* **32**, 19–30 (2019).

6. Wang, S. *et al.* An efficient segmentation method based on semi-supervised learning for seafloor monitoring in Pujada Bay, Philippines. *Ecol. Inform.* **78**, (2023).

7. Terayama, K. *et al.* Cost-effective seafloor habitat mapping using a portable speedy sea scanner and deep-learning-based segmentation: A sea trial at Pujada Bay, Philippines. *Methods Ecol. Evol.* **13**, 339–345 (2022).

8. Mizuno, K. *et al.* An efficient coral survey method based on a large-scale 3-D structure model obtained by Speedy Sea Scanner and U-Net segmentation. *Sci. Rep.* **10**, (2020).

9. Mizuno, K., Terayama, K., Ishida, S., Godbold, J. A. & Solan, M. Combining three-dimensional acoustic coring and a convolutional neural network to quantify species contributions to benthic ecosystems. *R. Soc. Open Sci.* **11**, (2024).

10. Gu, Y. *et al.* Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. *Comput. Biol. Med.* **103**, 220–231 (2018).

11. Solan, M. *et al.* In situ quantification of bioturbation using time-lapse fluorescent sediment profile imaging (f-SPI), luminophore tracers and model simulation. *Mar. Ecol. Prog. Ser.* **271**, 1–12 (2004).

12. Hale, R. *et al.* High-resolution computed tomography reconstructions of invertebrate burrow systems. *Sci. Data* **2**, (2015).

13. Plets, R. M. K. *et al.* The use of a high-resolution 3D Chirp sub-bottom profiler for the reconstruction of the shallow water archaeological site of the Grace Dieu (1439), River Hamble, UK. *J. Archaeol. Sci.* **36**, 408–418 (2009).

14. Mizuno, K. *et al.* Automatic non-destructive three-dimensional acoustic coring system for in situ detection of aquatic plant root under the water bottom. *Case Studies in Nondestructive Testing and Evaluation* **5**, 1–8 (2016).

15. Suganuma, H., Mizuno, K. & Asada, A. Application of wavelet shrinkage to acoustic imaging of buried asari clams using high-frequency ultrasound. *Jpn. J. Appl. Phys.* **57**, 07LG08 (2018).

16. Ji, S., Xu, W., Yang, M. & Yu, K. 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 221–231 (2013).

17. Kamnitsas, K. *et al.* Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017).

18. Zhou, J. *et al.* Weakly supervised 3D deep learning for breast cancer classification and localization of the lesions in MR images. *J. Magn. Reson. Imaging* **50**, 1144–1151 (2019).

19. Dou, Q., Chen, H., Yu, L., Qin, J. & Heng, P.-A. Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. *IEEE Trans. Biomed. Eng.* **64**, 1558–1567 (2017).

20. Yang, C., Rangarajan, A. & Ranka, S. Visual Explanations From Deep 3D Convolutional Neural Networks for Alzheimer's Disease Classification. *AMIA Annu. Symp. Proc.* **2018**, 1571–1580 (2018).

21. Molchanov, P. *et al.* Online Detection and Classification of Dynamic Hand Gestures with Recurrent 3D Convolutional Neural Networks. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 4207–4215 (IEEE, 2016). doi:10.1109/CVPR.2016.456.

22. Jie Huang, Wengang Zhou, Houqiang Li & Weiping Li. Sign Language Recognition using 3D convolutional neural networks. in *2015 IEEE International Conference on Multimedia and Expo (ICME)* 1–6 (IEEE, 2015). doi:10.1109/ICME.2015.7177428.

23. Selvaraju, R. R. *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. in *2017 IEEE International Conference on Computer Vision (ICCV)* 618–626 (IEEE, 2017). doi:10.1109/ICCV.2017.74.

24. Mizuno, K., Nomaki, H., Chen, C. & Seike, K. Deep-sea infauna with calcified exoskeletons imaged in situ using a new 3D acoustic coring system (A-core-2000). *Sci. Rep.* **12**, (2022).

25.   Abadi, M. *et al.* TensorFlow: a system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) USENIX Association.* 265–283 (2016).

26.   Scikit-learn, User Guide, 3.1. Cross-validation: evaluating estimator performance. Accessed July 24, 2024. https://scikit-learn.org/stable/modules/cross_validation.html

Acknowledgements

Author contributions

T. K.: data acquisition, data curation, formal analysis, investigation, software, methodology, writing—original draft; K.M.: conceptualization, data acquisition, data curation, investigation, methodology, project administration, resources, writing—review and editing; S.I.: methodology, resources, software, validation, writing—review and editing; S. O.: data acquisition, investigation; H. W.: data acquisition, investigation, conceptualization; Y. U.: data acquisition, investigation; Y. S.: data acquisition, investigation; K. O.: data acquisition, investigation; S. S.: data acquisition, investigation; Y. S.: data acquisition, investigation; K.T.: formal analysis, funding acquisition, software, validation, writing—original draft, review and editing.

Competing interests

The authors declare no competing financial interests.