Verification of the accuracy of rock material determination by hammer strike sound using deep learning

*Daisuke Sugeta[1], Hirokazu Furuki[1], and Shigeru Miyamura[2]
1 . Nippon Koei Co., Ltd., Reseach & Development Center, Japan 2. Nippon Koei Co., Ltd., Infrastructure Engineering Operations, Japan

Sharing is encouraged

Keywords: Deep Learning, audio classification, rocks material, japan

Abstract: We developed a deep learning model trained on hammer impact sounds to assess rock quality. As the result, the proposed model estimated rock quality with accuracy comparable to geological engineers. In addition, the 2D-CNN trained on the Log Mel Spectrogram was confirmed as the most effective.

1. Introduction (Motivation and Purpose of the Study)

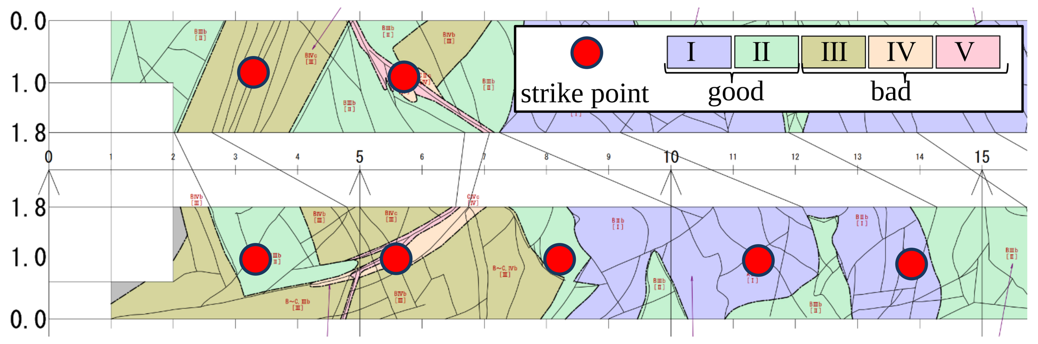

Fig.1 Acoustic inspection (top),
Geological map (bottom)

- In the inspection of rocks, the quality of the material is judged using the sound of hammer impacts.
- The heterogeneity of rocks affects the sound of hammer impacts. Therefore, even experienced engineers may have varying judgments.
- Additionally, the shortage of engineers with such judgment skills is a major social problem in Japan.

- In the field of civil engineering, deep learning is applied to assess the condition of concrete based on the sound of hammer impacts.
- However, there are no cases where deep learning has been applied to assess the quality of rocks in the field of geology.
- In this study, we developed a deep learning model trained on the sound of hammer impacts to assess the quality of rocks.

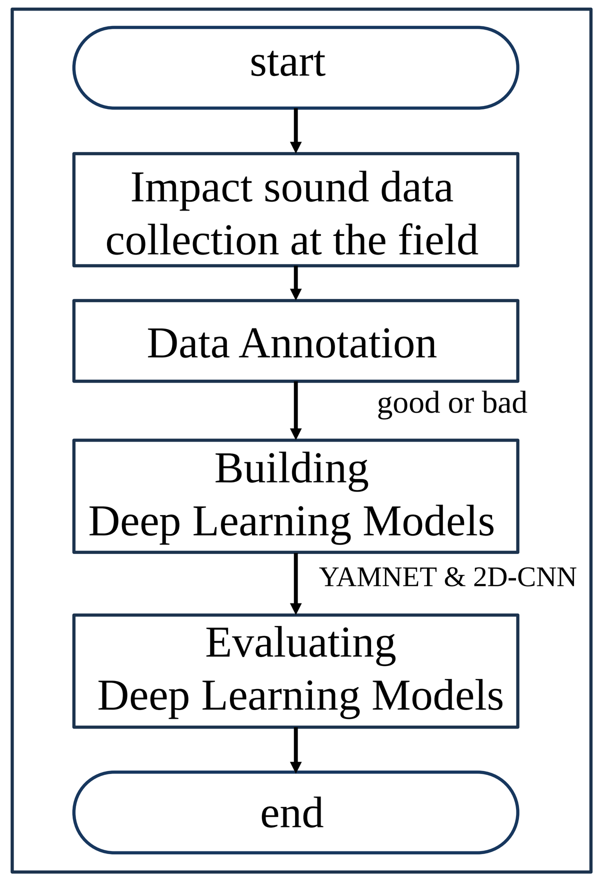

Fig.2 Flow-chart in this study

2. Data and AI Model

2.1 Data Collection And Creation

- Impact sound data was collected in the dam tunnels by striking rocks with a hammer. During the data collection, there was no noise such as car traffic sounds.
- The rocks were Mesozoic Cretaceous sandstones.
- We used an Estwing rock pick hammer with a mass of 900g and a total length of 330mm.
- The hammer impact sounds were recorded one by one using an iPhone 11 placed near the impact point. The iPhone11 was selected because it can easily record on-site.
- The recorded data is in WAV format, and the iPhone's sampling frequency is 48,000Hz.
- After recording, each sound data was labeled individually. In this study, the labels were either "good" or "bad". In addition, we did not use missed hits sound in this study.

2.2 Deep Learning Model

- In this study, we used two types of Deep Learning models. These models are widely used for sound classification.
- ① YAMNET is reported to have high performance in sound classification tasks because it is pre-trained on a large audio corpus available on YouTube. The input format for YAMNET is WAV format.
- ② 2D-CNN is a deep neural network model that learns features within images and makes judgments based on them.
- In this study, we applied a Fourier transform to the WAV files and created 3 cases of image data. The image data is given to 2D-CNN.

- These images show the frequency and amplitude of the sound. Engineers judge the quality of materials based on the pitch and volume of the sound, so these transformations are considered effective.

Fig.3 Adit and rocks in this study

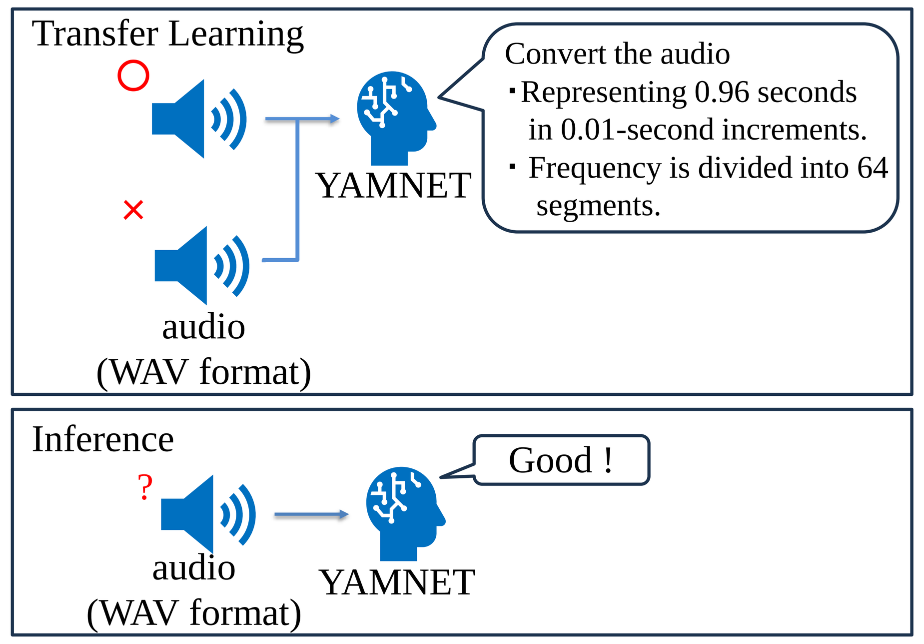

Fig.4 Training and inference image of YAMNET

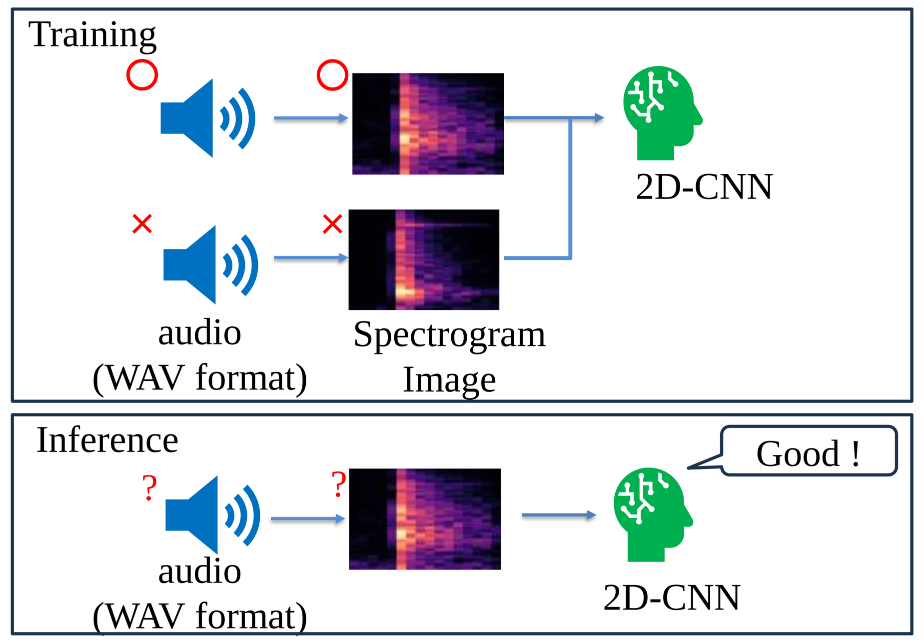

Fig.5 Training and inference image of 2D-CNN

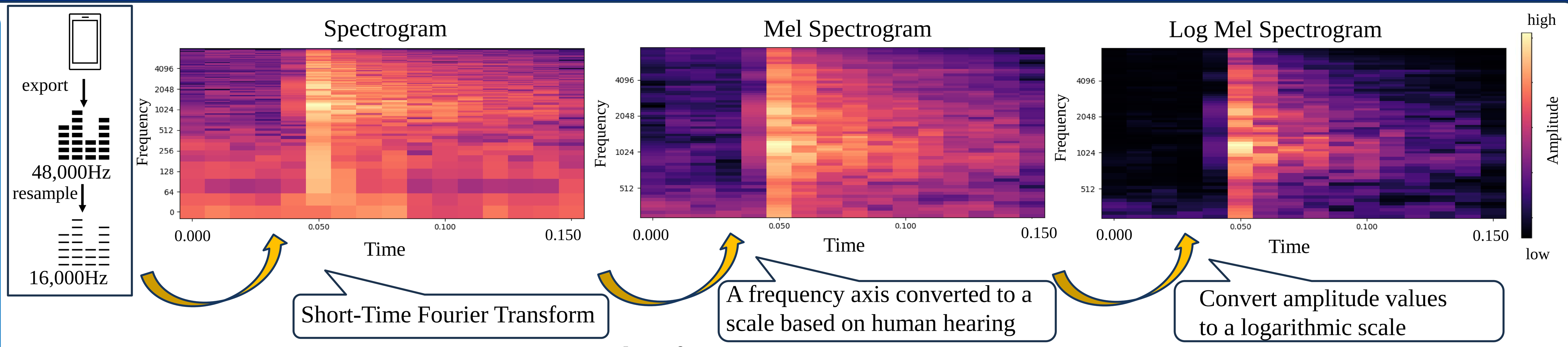

Fig.6 Procedure for creating spectrogram images

※ In this validation, we applied the same resampling conditions and input size of YAMNET.

3. Results

- In this study, Case 4 is the best result.
- As a result, it was confirmed that among the three types of features, the Log Mel spectrogram was the most effective.

Table.1 datasets in this study

data	label	quantity
training	good	267
	bad	359
test	good	67
	bad	90

		case1	case2	case3	case4
Deep Learning Model		YAMNET	2D-CNN	2D-CNN	2D-CNN
input data		wav	STFT	Mel Spectrogram	Log Mel Spectrogram
epoch		40	26	21	22
score	Accuracy	84.08%	86.62%	86.62%	87.26%
	F1 Score	86.63%	86.64%	86.67%	87.26%

Fig.7 Comparison of Model Score

- Using Optuna, we explored the optimal hyperparameters for the number of convolutional layers, neurons, and the dropout rate.
- Specifically, we performed stratified 5-fold cross-validation on 80% of the data, excluding the test data.
- After setting the optimal hyper-parameters, we trained on 80% of the data.

- Although YAMNET is reported to be a high-performance model, it was confirmed to have the lowest accuracy in this study.
- The short duration of the audio was considered a possible cause.
- Specifically, the input length for YAMNET is 0.96 seconds, but the audio data was 0.15 seconds.

- Therefore, it was thought that the model could not efficiently learn the audio features.
- Visualizing it, training data is such as Fig-8.

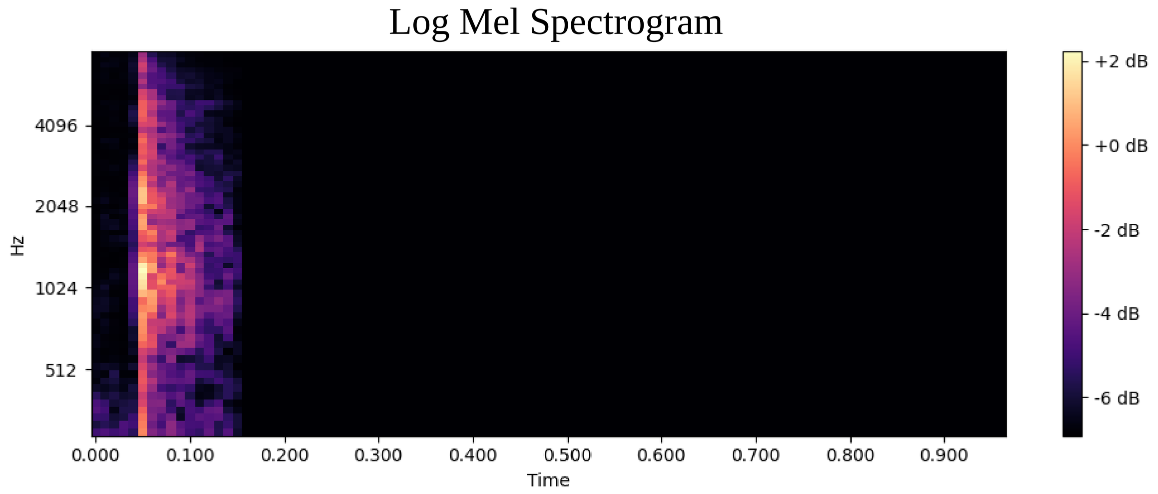

Fig.8 Features input to YAMNET

4. Conclusion

- The AI model developed in this study can estimate the quality of rock materials with accuracy comparable to geological engineers.
- The Log Mel spectrogram was confirmed to be the most effective.

- When using sounds shorter than 0.96 seconds, YAMNET is not necessarily effective.
- In the future, we plan to conduct further verification by subdividing the classification of rock materials. We also introduce the state-of-the-art AI technologies such as transformers.

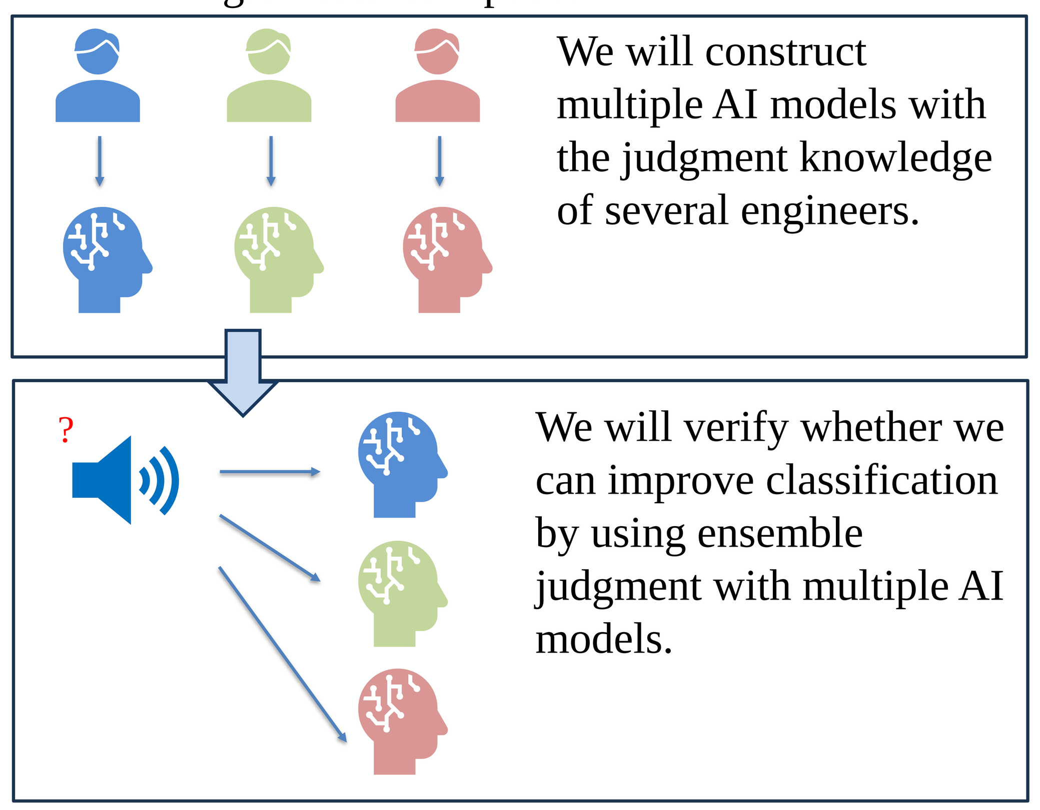

Fig.9 Ensemble judgment image by multiple AIs

An application for recording the impact sound is now available for free!

Contact Sugeta: SUGETA-DS@n-koei.jp , Furuki: FURUKI-HR@n-koei.jp , Miyamura: MIYAMURA-SH@n-koei.jp