2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Automatic Training Data Selection for Autoencoder-based Acoustic Defect Detection Robust against Class-Imbalance

Aki Takamura[1], Koki Shoda[1], Jun Younes Louhi Kasahara[2], Takuya Igaue[2], Qi An[1] and Atsushi Yamashita[1]

Abstract— **This study proposes an automatic training data selection method for Autoencoder-based defect detection in hammering inspection, designed to address the severe class imbalance between normal and defective sounds. The proposed method employs a physically grounded indicator, Acoustic Energy per Impact, to automatically select and collect only the sound data considered normal. An Autoencoder is then trained exclusively on the collected normal sounds to identify defects based on reconstruction errors. To evaluate the effectiveness of the proposed method, experiments were conducted using concrete specimens with cracks. The results demonstrate that the proposed method achieves higher defect detection performance than a conventional approach, even under highly imbalanced class conditions.**

I. INTRODUCTION

In recent years, the deterioration of concrete structures, widely used as social infrastructure, has been accelerating. In Japan, approximately 25% of tunnels and 37% of road bridges have already exceeded their service life [1], making maintenance and inspection to preserve their integrity an urgent issue.

The hammering test is widely used, particularly in the primary inspection of concrete structures, as a simple and non-destructive method for detecting internal anomalies. Japan's Ministry of Land, Infrastructure, Transport and Tourism has mandated this method for periodic inspections [2] . Meanwhile, the number of skilled inspectors is decreasing due to a declining birthrate and aging population. To reduce reliance on human resources, research into the automation of the hammering test has become active [3], [4], [5], [6].

Several supervised learning approaches for defect identification have been proposed. These methods require preparing specific features, such as the Fourier spectrum of the hammering sounds, and corresponding labels indicating a healthy or defect state. This data is then used to train a classifier, such as an Artificial Neural Network [7], [8]. However, because the composition and curing conditions of concrete vary, the resulting hammering sounds also differ, necessitating on-site training. Training a defect identifier requires a large number of hammering sounds labeled by skilled inspectors, which significantly increases the inspectors' workload.

[1]Aki Takamura, Koki Shoda, Qi An and Atsushi Yamashita is with Department of Human and Engineered Environmental Studies, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563, Japan {aki-takamura, shoda, an, yamashita}@robot.t.u-tokyo.ac.jp

[2]Jun Younes Louhi Kasahara, Takuya Igaue with Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyoku, Tokyo, 113-8656, Japan {louhi, igaue}@robot.t.u-tokyo.ac.jp

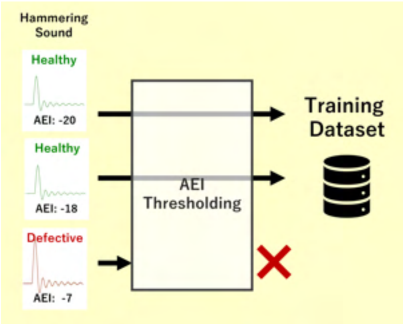
To mitigate this labeling cost, approaches such as transfer learning [9] and semi-supervised learning [10] also exist. Transfer learning is an approach that fine-tunes a model trained at a different site with a small amount of labeled data, while semi-supervised learning similarly utilizes a small amount of labeled data. Although both methods reduce the required volume of labeled data, they cannot eliminate it entirely and thus fall short of fundamentally solving the labeling cost problem.

Therefore, unsupervised learning approaches that do not require prior labeling have also been proposed. Louhi et al. [11] perform clustering of hammering sounds using sound and location. Furthermore, based on the assumption that the majority of hammering sounds gathered during an inspection are from healthy concrete, they identify the cluster with the most data points as the healthy cluster. Shoda et al. [12] proposed a method that clusters hammering sound data based on frequency features and identifies defect clusters from the average value of a physical feature for each cluster. However, this approach faces a new challenge. In primary inspections, the entire structure is sparsely hammered. As a result, the majority of gathered data consists of healthy sounds, while defect sounds are extremely rare. Under such "class-imbalance" conditions, it has been reported that the minority defect sounds cannot form a distinct cluster, leading to a significant drop in identification accuracy.

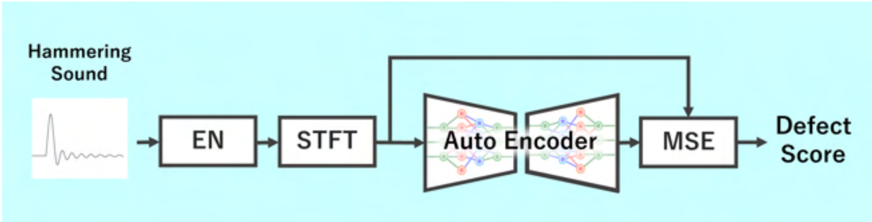
Thus, we focus on an outlier detection approach that determines whether an individual data point is normal or anomalous, because it is difficult to capture defects as a "group" with clustering. In this context, there is Isolation Forest, which does not even require training data. Moghadam et al. [13] applied this method to the acceleration response data of structures and demonstrated its effectiveness. However, because Isolation Forest separates data with axis-parallel splits, it has limitations in accurately capturing the non-linear boundary between healthy and defect states in high-dimensional acoustic data, where various frequency components are intricately correlated.

Therefore, this research focuses on an anomaly detection method using an Autoencoder [14], which can automatically learn features from high-dimensional, non-linear data. This method learns only from healthy sounds to model their "ideal state." It then identifies defects based on the large reconstruction error that occurs when an unknown sound that deviates significantly from the training data is input. The use of Autoencoders for defect identification in hammering tests has been shown to be effective [15], [16], and this approach is promising as it is less susceptible to the effects of class imbalance. However, conventional applications of Autoen-

978-1-6654-5784-2/26/$31.00 ©2026 IEEE



(a) Automatic selection of sound data using AEI to construct a training dataset for an Autoencoder.



(b) Identification of healthy and defective sounds using an Autoencoder. EN, STFT, and MSE stands for energy normalization, Short-Time Fourier Transform, and Mean Squared Error, respectively.

Fig. 1: Framework of the proposed method

coders assume that "labeled healthy sounds" are prepared in advance. In the context of our problem, where hammering sound characteristics differ from site to site, this ultimately leads back to the labeling cost issue of how to prepare the healthy sounds for training.

Existing research has faced a barrier of either "labeling cost" or "class imbalance," and no definitive method has been established to overcome both simultaneously. The objective of this study is to solve both of these challenges simultaneously by proposing a method to automatically select healthy sound data for training an Autoencoder, without any label information. Through this, we aim to build a practical, automated defect identification framework that is robust to class imbalance and requires no manual labeling whatsoever.

II. PROPOSED METHOD

A. Concept

The objective of this research is to address the unsolved problem of "automated defect identification from hammering tests under class imbalance without labeling costs." As the vast majority of a concrete structure is in a healthy state, the collected hammering sound data is overwhelmingly dominated by healthy sounds, with defect sounds being extremely rare. This class imbalance problem is a major factor that complicates the application of most machine learning methods.

In addressing this challenge, clustering-based approaches, which detect anomalies by analyzing relationships across the entire dataset, are found to be insufficient for providing a fundamental solution. This limitation arises from the fact that defects exhibit substantial variability depending on the size and depth of internal voids, making it difficult to detect them collectively as a single, consistent group. Therefore, this study is based on the premise that a system employing an individual discriminator—which determines whether each sound instance is healthy or not—is more appropriate than methods that rely on inter-sample relationships.

In designing this discriminator, we focused on the physical mechanism of sound generation in concrete hammering. Physically, it is expected that sounds from healthy locations, which have no internal voids, will be uniform. In contrast, sounds from defects are expected to be highly diverse, as their resonant frequencies depend on the size and depth of the internal voids. Based on this asymmetrical relationship—the "uniformity of healthy sounds" versus the "diversity of defect sounds"—we concluded that a One-Class discriminator, which uses only healthy sounds as training data, is the most rational choice.

Furthermore, to enhance the practicality of this method and automate the discriminator's training process, we once again turn to the physical mechanism. In defective areas with internal voids, vibrations are more easily excited by hammering[17], leading to a tendency for the acoustic energy generated per unit of impact energy (AEI) to be larger[12]. Calculating this AEI requires an instrumented impact hammer capable of measuring the impact force, making its use a prerequisite for the proposed method. We propose a framework that leverages this physical index, AEI, to automatically collect data with low AEI as "healthy sounds." While AEI alone cannot perfectly distinguish between healthy and defective sounds—owing to various factors such as the hammer's impact angle and its diversity, the coefficient of restitution, the microphone-to-impact distance, acoustic reverberation, and ambient noise at the inspection site—it is observed that sounds with a low AEI can be reliably collected as 'healthy'. Through this approach, we aim to realize a practical, automated defect identification system by circumventing the problem of class imbalance while also automating the collection and selection of training data.

From a practical standpoint, the proposed method is envisioned for on-site application, identifying defects using only the sounds collected during inspection. It is designed to allow an operator to automatically determine the presence of defects simply by tapping the structure's surface and recording the resulting sounds. This process eliminates the need for prior data labeling by experts or a dedicated pre-training stage, thereby offering substantial reductions in both implementation cost and preparation time.

The overall configuration of the defect identification framework proposed in this study is shown in Fig. 1. First, to construct a training dataset for an Autoencoder, only healthy sounds are automatically selected from the hammering sound

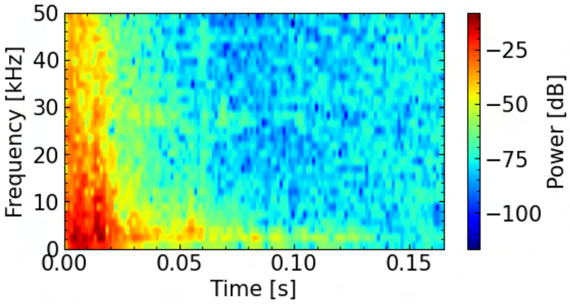


Fig. 2: Spectrogram of a hammering sound.

data using AEI. Next, an Autoencoder is trained using the extracted healthy sounds. Finally, the reconstruction error for an unknown hammering sound is calculated to distinguish between healthy and defective sounds.The following sections will explain the role and technical details of each processing step in this framework.

B. Acoustic Feature Processing

For the hammering sound data, the energy of each sound is first normalized to eliminate variations in impact force. Then, the Short-Time Fourier Transform (STFT) is applied to extract acoustic features in the time-frequency domain. The STFT for a time-series signal $x(t)$ is defined as follows:

$$\mathrm{STFT}\{x(t)\}(t,\omega) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j\omega\tau}d\tau. \quad (1)$$

Here, $w(\tau - t)$ is a window function used to locally segment the signal. By taking the square of the magnitude of the complex spectrum obtained by STFT, a spectrogram, which represents the energy distribution in the time-frequency domain, is obtained. The STFT is suitable for analyzing non-stationary signals like hammering sounds, as it can capture local time and frequency information of acoustic signals.

In this study, the frequency features obtained by STFT (31 frequency bins × 129 time frames, see Fig. 2) are used as input features.

C. Automatic Selection of Sound Data Using AEI

Prior to training the Autoencoder, it is necessary to collect sound data to be used as training data. As a criterion for this, we use AEI proposed by Shoda et al. [12].

AEI is defined as the ratio of the acoustic energy E_{acoustic} to the impact energy E_{impact} applied to the concrete. AEI is given by Equation (2):

$$\mathrm{AEI} = \frac{E_{\text{acoustic}}}{E_{\text{impact}}} \propto \frac{\int a(t)^2 dt}{(\int f(t)dt)^2}. \quad (2)$$

Here, $a(t)$ and $f(t)$ represent the amplitude of the sound pressure and the force at time t, respectively.

In this study, hammering sounds with an AEI below a certain threshold are considered sound data and are used as training data for the Autoencoder. The specific value of this threshold will be set based on the AEI distribution in Section III.C.

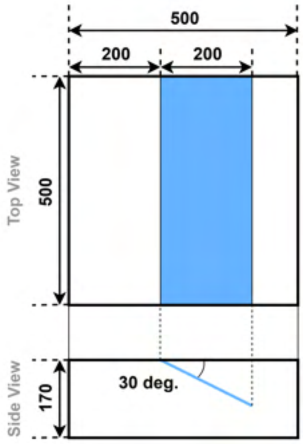


Fig. 3: Schematic diagram of the concrete specimen with a 30° crack. The crack is shown in blue.

D. Identification of Healthy and Defective Sounds using an Autoencoder

An Autoencoder [14] is a type of neural network that compresses (encodes) and reconstructs (decodes) input data. By setting the input and output to be the same and training the network to minimize the reconstruction error, it extracts the essential features of the data as a low-dimensional representation.

In this study, the frequency features of healthy sounds automatically extracted by AEI are used as training data, and an Autoencoder is trained with 31 × 129 dimensional feature vectors as input. The trained model can reconstruct healthy sounds with high accuracy, while the reconstruction error for unknown defective sounds tends to be large. Therefore, by using the Mean Squared Error (MSE) between the input and output, defect identification based on reconstruction error becomes possible.

In this study, the 99.99th percentile of the reconstruction error for healthy sounds obtained during training is set as the identification threshold, and hammering sounds with an error exceeding this threshold are determined to be defective.

III. EXPERIMENT

A. Experimental Objective and Overview

To validate the effectiveness of the proposed method, an experiment was conducted to compare its defect identification accuracy with an existing method [12]. An impact hammer was used to strike the healthy and defective areas of a concrete specimen, and AEI was calculated from the acquired acoustic and force signals. Since cracks are a typical defect in concrete structures, a concrete specimen with a 30° crack, as shown in Fig. 3, was used in this experiment.

B. Experimental Setup

In this study, the experimental setup was arranged as shown in Fig. 4. A microphone (PCB Piezotronics 377B02) was mounted on a stand, and the hammering sounds generated by an impact hammer (PCB Piezotronics 086C03)

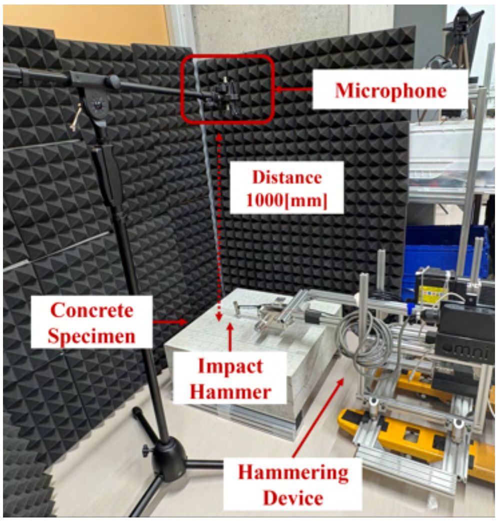


Fig. 4: Experimental setup.

TABLE I: Number of samples used in each class ratio

Class Ratio	Healthy Samples	Defective Samples
50:1	600	12
20:1	600	30
10:1	600	60

attached to an automatic hammering device were acquired simultaneously with the force signals by a data logger (Data Translation DT9837B).

C. Experimental Conditions

In this experiment, a total of 1,529 hammerings were made with the impact hammer on the healthy and defective regions. The breakdown is 916 points from the healthy region and 613 points from the defective region. The sampling frequency was set to 100 kHz, the maximum limit of the data logger.

For the experimental conditions of each class ratio, samples of healthy and defective sounds were randomly extracted as shown in Table I.

For this experiment, the AEI threshold was determined based on the distribution of AEI values from the entire collected dataset, comprising both healthy and defective samples. Furthermore, under each experimental condition, random sampling and identification processing were repeated five times on the extracted sound data to evaluate the average identification performance.

D. Comparison with Other Methods

To validate the effectiveness of the proposed method, its accuracy was compared with an existing method [12]. To investigate the validity of using an Autoencoder, the accuracy of the proposed method was also compared with the case where defect determination was made using only the AEI threshold.

IV. RESULTS

A. Setting the AEI threshold

Regarding the AEI threshold setting mentioned in Section III.C, based on the AEI distribution of the 1,529 hammerings acquired in this experiment, as shown in Fig. 5, the threshold was set to automatically extract sounds with an AEI of -17.25 dB or less as healthy sounds.

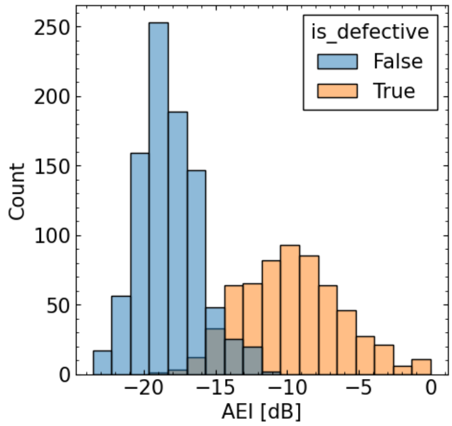


Fig. 5: AEI Distribution of Healthy and Defective Sounds in This Experiment

B. Identification Performance for Each Class Ratio

The identification performance of the proposed method was evaluated under three conditions with class ratios of 50:1, 20:1, and 10:1. Table II summarizes the classification accuracies under all conditions evaluated. Fig. 6 shows the reconstruction error distribution for healthy and defective sounds for each class ratio in the experiment, with the red dotted line indicating the identification threshold.

The identification performance of the proposed method was evaluated under three conditions with class ratios of 50:1, 20:1, and 10:1. Table II summarizes the classification metrics under all evaluated conditions, and Fig. 6 shows the reconstruction error distributions for each class ratio, with the red dotted line indicating the identification threshold. As shown in the ”Proposed Method” columns of Table II, the overall performance of the proposed method improves as the class imbalance becomes less severe. Specifically, the F1-score demonstrates a clear upward trend, increasing from 0.30 ± 0.13 at a 50:1 ratio to 0.55 ± 0.17 at 20:1, and further to 0.69 ± 0.10 at 10:1.

A detailed analysis of the metrics reveals that this trend is primarily driven by changes in precision, while recall remains consistently high. The recall score was maintained at 0.90 or above across all conditions, indicating that the proposed method can reliably detect defective sounds without significant omission. This high recall is visually supported by the error distributions in Fig. 6, where the vast majority of the defective sound distribution (orange) lies beyond the threshold.

In contrast, the precision score improved significantly from 0.22 at 50:1 to 0.65 at 20:1. Under the highly imbalanced 50:1 condition, the misclassification of even a small number of healthy sounds as defective (False Positives) has a disproportionately large negative impact on the precision score. Fig. 6 illustrates the source of these False Positives, showing the tail of the healthy sound distribution extending past the identification threshold. As the proportion of defective sounds increases, the relative impact of these misclassifications diminishes, leading to a higher precision and, consequently, a better F1-score.

C. Comparison with Other Methods

To evaluate the effectiveness of the proposed method, a comparison was made with an existing defect identification method based on AEI clustering [12], as well as with a simpler method using only the AEI value for classification. Table II shows the accuracy, precision, recall, and F1-score for each method. The proposed method achieved a higher F1-score than both comparative methods under all conditions, demonstrating its superior overall performance.

Compared with the existing method [12], the proposed method demonstrated superior results for defect detection under class imbalance. As shown in Table II, the recall of the existing method is near zero, which means it fails to identify most defects (i.e., it has a high number of False Negatives). Although the accuracy of the existing method (0.92) is higher than that of the proposed method (0.89) at the 50:1 ratio, this is a misleadingly high value achieved by defaulting to the majority "healthy" class. Considering the primary goal of defect detection, the F1-scores (0.02 for the existing method vs. 0.30 for the proposed) make the performance gap clear, confirming the fundamentally higher reliability of our approach.

The comparison with the AEI-only method clarifies the value of combining a physical indicator (AEI) with a machine learning model (the autoencoder). As mentioned in Section II.A, AEI alone cannot be a perfect classifier due to confounding factors such as hammer impact angle and noise. This concern was borne out by the experimental results, where the AEI-only method yielded extremely low precision, erroneously flagging numerous healthy sounds as defective (a high rate of False Positives). The proposed method compensates for this weakness using the autoencoder. By learning the spectral features of sounds pre-selected by AEI as "healthy," the autoencoder can correctly identify sounds that may have a high AEI value but still exhibit a "healthy-like" frequency structure. This effective reduction of False Positives is the direct contributor to the dramatic improvement in precision and, consequently, the superior F1-score, which is the core strength of our proposed framework.

V. DISCUSSION

In this study, we proposed a method to automatically select only healthy sounds using Acoustic Energy per Impact (AEI) as an indicator, in order to address the severe class imbalance between normal and defective sounds. An Autoencoder was then trained on the selected data to identify defective sounds.

The proposed method showed a higher F1-score than the existing method under all experimental conditions. In particular, the recall was consistently close to 1.00, confirming that defective sounds were detected without omission.

On the other hand, a trend was observed where the F1-score decreased as the class ratio became more imbalanced (i.e., as the number of defective samples decreased). However, this does not signify a decline in the model's fundamental ability to detect defects. In fact, the recall score was consistently maintained near 1.00 across all class ratios, indicating that the proposed method reliably identified the target defective sounds without omission.

The decrease in the F1-score is therefore primarily attributed to its impact on precision. Under conditions with an extremely small number of defective samples, the misclassification of even a few healthy sounds as defective (False Positives) can disproportionately lower the precision score. It is reasoned that the increased relative impact of these False Positives lowered the F1-score—the harmonic mean of precision and recall—even while a high recall was maintained. In practical inspections, missing defective parts is more critical than mistakenly flagging healthy ones, meaning recall has higher importance as a performance metric. Therefore, although the F1-score decreases in highly imbalanced cases due to a drop in precision, the proposed method still maintains reliable defect detection performance by consistently achieving near-perfect recall.

Furthermore, compared to defect identification using only the AEI value, the proposed method showed a high F1-score. This is thought to be because training the Autoencoder on the frequency features of healthy sounds allowed it to correctly identify even some healthy sounds whose AEI values exceeded the threshold, without misclassifying them as defective.

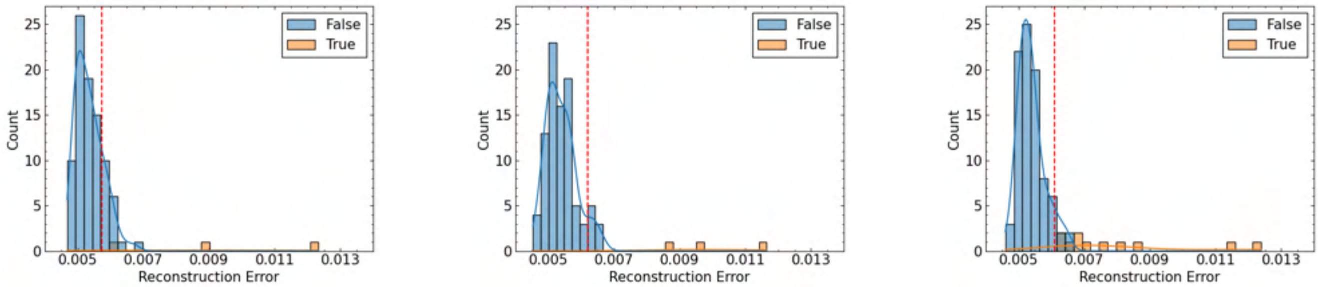
However, under all experimental conditions, there were some healthy sounds that exceeded the identification threshold based on the Autoencoder's reconstruction error. One possible reason for this is that healthy sounds with an AEI value above -17.25 dB were excluded from the training data. As a result, these healthy sounds were not part of the learning target, which may have led to a higher reconstruction error. This suggests that if the threshold for AEI-based healthy sound extraction is inadequately set, it can adversely affect identification accuracy.

Additionally, some defective sounds were observed to have a lower reconstruction error than healthy sounds. A possible reason is that the frequency structure of the corresponding defective sounds was similar to that of healthy sounds. Moreover, the possibility that defective sounds with an AEI value of -17.25 dB or less were mistakenly included in the training data as healthy sounds cannot be ruled out. In such cases, it is presumed that this mislearning suppressed the reconstruction error for those defective sounds.

Although the proposed framework achieved high recall and improved F1-scores, it still relies on a fixed AEI threshold when extracting healthy sounds. An inappropriate threshold may exclude healthy sounds or include defective ones, reducing generalization in different materials or field environments. Future work will therefore introduce statistical or adaptive threshold optimization and conduct experiments on specimens with different crack types and noisy conditions to improve robustness in practical inspections.

TABLE II: Comparison of accuracy with the conventional method for each class ratio.

	50:1			20:1			10:1		
	[12]	Only AEI	Proposed Method	[12]	Only AEI	Proposed Method	[12]	Only AEI	Proposed Method
Accuracy	**0.92 ± 0.06**	0.71 ± 0.01	0.89 ± 0.04	**0.95±0.00**	0.76±0.11	0.91±0.03	**0.91±0.00**	0.73 ± 0.01	0.88 ± 0.06
Precision	0.01 ± 0.01	0.06 ± 0.00	**0.22 ± 0.12**	0.00±0.00	0.14±0.00	**0.65±0.35**	0.00±0.00	0.25 ± 0.00	**0.57 ± 0.11**
Recall	0.20 ± 0.20	**1.00 ± 0.00**	0.90 ± 0.10	0.00±0.00	**0.99±0.01**	0.98±0.04	0.00±0.00	**0.98 ± 0.01**	**0.98 ± 0.02**
F1 Score	0.02 ± 0.02	0.12 ± 0.00	**0.30 ± 0.13**	0.00±0.00	0.24±0.01	**0.55±0.17**	0.00±0.00	0.40 ± 0.01	**0.69 ± 0.10**



(a) 50:1 (b) 20:1 (c) 10:1

Fig. 6: Distribution of reconstruction error for normal and defective sound identification for each class ratio. The blue region indicates the distribution of normal sounds, the orange region indicates the distribution of defective sounds, and the red dashed line represents the threshold for defect detection.

VI. CONCLUSION

In this study, to solve the dual challenges of "labeling cost" and "class imbalance" in the hammering test, we proposed a novel framework that uses the physical index AEI to automatically select training data for an Autoencoder and identify defects. Through experiments using concrete specimens with simulated cracks, we confirmed that the proposed method achieves defect identification performance superior to that of conventional methods, even in a severely imbalanced data environment and without any manual labeling.

Future work will involve optimizing the criteria for selection healthy sounds based on AEI and the final identification threshold to further improve accuracy and enhance generalization performance.

ACKNOWLEDGMENT

This work was supported in part by the Suzuki foundation.

REFERENCES

[1] Ministry of Land, Infrastructure, Transport and Tourism (MLIT), "Current state and future of aging social infrastructure - infrastructure maintenance information." `https://www.mlit.go.jp/sogoseisaku/maintenance/02research/02_01.html`. Accessed on July 6, 2025.

[2] Ministry of Land, Infrastructure, Transport and Tourism (MLIT), "Mlit notification no. 282 of fy2008," 2007. Notification by the Ministry of Land, Infrastructure, Transport and Tourism.

[3] P. Chun, J. Dang, S. Hamasaki, R. Yajima, T. Kameda, H. Wada, T. Yamane, S. Izumi, and K. Nagatani, "Utilization of unmanned aerial vehicle, artificial intelligence, and remote measurement technology for bridge inspections," *Journal of Robotics and Mechatronics*, vol. 32, no. 6, pp. 1244–1258, 2020.

[4] Y. Nishimura, S. Takahashi, H. Mochiyama, and T. Yamaguchi, "Automated hammering inspection system with multi-copter type mobile robot for concrete structures," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9993–10000, 2022.

[5] H. Fujii, A. Yamashita, and H. Asama, "Automated diagnosis of material condition in hammering test using a boosting algorithm," in *2014 IEEE International Workshop on Advanced Robotics and its Social Impacts*, pp. 101–107, 2014.

[6] B. Li, K. Ushiroda, L. Yang, Q. Song, and J. Xiao, "Wall-climbing robot for non-destructive evaluation using impact-echo and metric learning svm," *International Journal of Intelligent Robotics and Applications*, vol. 1, no. 3, pp. 255–270, 2017.

[7] Łukasz Sadowski, "Non-destructive identification of pull-off adhesion between concrete layers," *Automation in Construction*, vol. 57, pp. 146–155, 2015.

[8] T. Fukumura, H. Aratame, A. Ito, M. Koike, K. Hibino, and Y. Kawamura, "A study on hammering test using deep learning," in *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)*, pp. 261–264, 2020.

[9] T. Fukumura, H. Aratame, A. Ito, M. Koike, K. Hibino, and Y. Kawamura, "An efficient learning method for sound classification using transfer learning for hammering test," in *2020 IEEE SENSORS*, pp. 1–4, 2020.

[10] J. Igual, A. Salazar, G. Safont, and L. Vergara, "Semi-supervised bayesian classification of materials with impact-echo signals," *Sensors*, vol. 15, no. 5, pp. 11528–11550, 2015.

[11] J. Y. L. Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Fuzzy clustering of spatially relevant acoustic data for defect detection," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2616–2623, 2018.

[12] K. Shoda, J. Y. Louhi Kasahara, H. Asama, Q. An, and A. Yamashita, "Clustering of hammering sounds and identification of defect clusters based on acoustic energy per impact for detection of concrete defects," in *2024 24th International Conference on Control, Automation and Systems (ICCAS)*, pp. 614–619, 2024.

[13] K. Y. Moghadam, M. Noori, A. Silik, and W. A. Altabey, "Damage detection in structures by using imbalanced classification algorithms," *Mathematics*, vol. 12, no. 3, 2024.

[14] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[15] H. Emoto, N. Fukui, Y. Iitaka, and S.-i. Kanazawa, "Identification of spalling in concrete structures by a hammering test using autoencoder," in *Life-Cycle of Structures and Infrastructure Systems*, pp. 570–577, CRC Press, 2023.

[16] K. Ikeda and A. Kamimura, "Hammering acoustic analysis using machine learning techniques for piping inspection," *Journal of Robotics and Mechatronics*, vol. 32, no. 4, pp. 789–797, 2020.

[17] N. Yasuda, N. Misaki, Y. Shimada, and D. Yamaoka, "Detection and characteristics estimation of defects in concrete structures using laser ablation-induced vibration," *Tunnelling and Underground Space Technology*, vol. 103, p. 103460, 2020.