Positive Weak Supervision Quality Increase by Consolidation for Acoustic Defect Detection in Concrete Structures

Jun Younes Louhi Kasahara[1], Atsushi Yamashita[1] and Hajime Asama[1]

***Abstract*— The aging of concrete social infrastructures such as tunnels, bridges, and highways is a growing concern worldwide. Those require careful inspection to ensure their users' safety and traditional manual methods are not viable solutions due to the growing population of structures in need of testing and the manpower shortage. Among those inspection methods, the hammering test has been the focus of several previous works, including notably weakly supervised approaches. Those approaches query a human user on random audio sample pair similarity to transform the feature space into one suited for defect detection. However, the quality of the weak supervision obtained in such a way is often variable. Therefore, we propose a method to improve positive weak supervision quality by consolidating the dataset prior to the query process. Experiments conducted with concrete test blocks showed the effectiveness of our proposed method.**

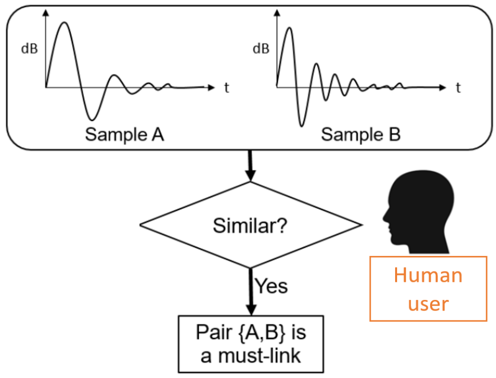


Fig. 1. Positive weak supervision, also known as must-links, is obtained by querying a human user on pairs of audio samples.

I. INTRODUCTION

Social infrastructures such as tunnels, bridges, and highways are primarily made of concrete. Such concrete structures are subject to deterioration due to aging and various environmental factors such as rains and vibrations. Therefore, regular inspection is paramount in order to ensure their users' safety [1]. This was underlined by several tragic events such as the collapse of the Morandi bridge in Italy [2] or the collapse of the Sasago tunnel in Japan [3]. While the population of structures in need of testing is increasing at an alarming rate, due to notably aging, the manpower required for inspection is decreasing. Therefore, the automation of inspection methods for concrete structure inspection is highly desirable.

Among inspection methods, the hammering test, an acoustic inspection method consisting of using a hammer to strike the surface of a structure and using the impact sound to assess the presence of defects, has been the focus of several previous works [4][5][6]. The hammering test is widespread in inspection sites, due to its effectiveness and ease of use.

Previous works dealing with the audio analysis of hammering sound have mostly employed machine learning approaches. The works in [7] and [8] employed supervised learning methods to classify hammering samples between defect and non-defect. Furthermore, defect samples were classified according to their depth from the surface. While achieving remarkable results, such supervised methods have their performance conditioned by the availability of appropriate training data. This is troublesome for concrete structures since each one is unique due to concrete mixing ratios, structure, and environmental conditions it was subjected to.

To bypass the issue of training data, unsupervised approaches were proposed in [9], [10], and [11], based on clustering approaches. Good results were obtained but since hammering position information was added in the analysis, those approaches required additional sensors to complement the audio hammering data. This involves the installation of the required sensors in the inspection apparatus. Furthermore, application to hammering data without position information is limited.

Weakly supervised methods, only requiring a human user to answer to queries on randomly selected sample pair similarity, as illustrated in Fig. 1, was proposed in [12]. This had the issue of gathering weak supervision of variable quality and therefore resulting in a variable defect detection performance. To tackle this issue, an expansion was proposed in [13], to include an active query scheme to reduce the variation in weak supervision quality. However, this active query scheme also relied on hammering sample position information, requiring additional sensors to record the hit position of a hammering sample.

Therefore, in this paper, a method to improve weak supervision quality in acoustic defect detection in concrete structures without the use of other sensors than a microphone is proposed. This is beneficial for the processing of previously gathered hammering data that only contain audio data and/or for inspection apparatus that can only record audio data due to their design. The improvement of weak supervision quality is conducted on positive weak supervision, i.e., positive answers to queries, also known as *must-links*. This

[1]Jun Younes Louhi Kasahara, Atsushi Yamashita and Hajime Asama are with the Department of Precision Engineering, The University of Tokyo, 113-8656 Tokyo, Japan. louhi@robot.t.u-tokyo.ac.jp

is achieved by consolidation of the hammering dataset prior to the query process. By reducing the query pool to only relevant members, more consistent positive weak supervision of higher quality can be collected.

II. METHOD

A. Overview and Concept of Proposed Method

The motivation behind our proposed approach is that weak supervision on pairs of samples already considered similar by the clustering algorithm in the initial feature space does not contribute significantly to the search for a better feature space. Therefore, we limit the query pool by consolidation of the dataset.

There are two advantages to this process:

- Potential queries on pairs of samples already close together in the feature space prior to weakly supervised feature space transformation, which can be expected to have little contribution in the search of a better feature space, are removed.
- The overall size of the query pool can be greatly reduced. Assuming the pairs of samples to be queried to the human user are selected randomly, the probability of obtaining repetitions, i.e., pairs of samples with a common sample, is increased, allowing to generate more effective chunklets, or proto-clusters, for weakly supervised feature space transformation.

An overview of our proposed method is shown in Fig. 2. The hammering audio data collected from a microphone is converted to Mel-Frequency Cepstrum Coefficients (MFCC) feature vectors. Consolidation is used to reduce the query pool and the query process to gather weak supervision is conducted. Following this, Relevant Component Analysis (RCA) is used to transform the feature space. Finally, K-Means clustering is used to separate defect and non-defect samples.

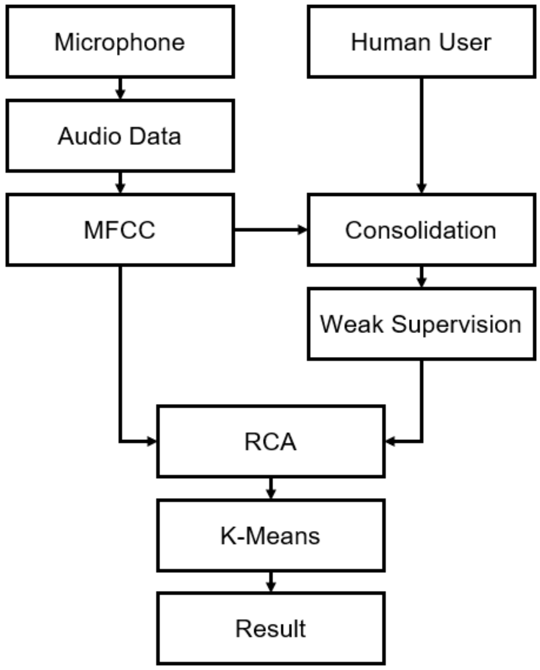


Fig. 2. Overview of the proposed method.

B. Pre-processing to MFCC Feature Vectors

Hammering samples are initially time-series audio data. They are first converted to Fourier Spectrum. Then, in order to account for the variations caused by irregular hammering force, a normalization to zero mean and unit variance as described in [13] is conducted. After this, MFCC feature vectors are computed. MFCC are feature vectors designed to mimic the human ear's perception of sounds and have been shown to be effective in discriminating defects and non-defect hammering samples [10]. In the remaining of this paper, the MFCC of a hammering sample will be noted as $\mathbf{x}_i$.

C. Consolidation to Improve Weak Supervision Quality

The proposed method to improve the quality of weak supervision consists of limiting the domain where the human user can be queried on. Given a dataset of N_{sample} hammering samples, the initial query pool, i.e., the set of possible queries, corresponds to all the unique possible pairs among N_{sample} samples. Among those, it can be expected that some do not contribute, or contribute to only a lesser extent, to the feature space transformation described in the following section. Therefore, since in most practical scenarios the amount of weak supervision that can be collected is limited, it is desirable to remove such pairs from the possible query pool.

Given a dataset of hammering samples D, K-Means is used to consolidate D, i.e., cluster D into a fine partition $P = \{P_1, ..., P_{K_c}\}$ such as $D = \cup_{i=0}^{K_c} P_i$ and $\cap_{i=0}^{K_c} P_i = \emptyset$. As opposed to the final clustering aiming at separating defect and non-defect hammering samples, this clustering at the consolidation step aims at *reducing* the dataset D to its meaningful members. Once partitioning has been conducted, one member of each partition P_i is randomly chosen as representative. Those representatives are used to establish the consolidated hammering dataset D', over which the query process is conducted.

D. Relevant Component Analysis

RCA is a weakly supervised metric learning method initially proposed in [14]. RCA is essentially a biased form of Principal Component Analysis, computed on positive weak supervision.

First, using the transitive closure property of must-links, chunklets, i.e., sets of samples deduced to belong to the same cluster, are built. Given $N_{chunklet}$ chunklets $\{\mathcal{M}_l\}_{l \in [1 \dots N_{chunklet}]}$, with $\hat{\mathbf{m}}_l$ being the mean of elements in $\mathcal{M}_l$ and N_{total} being the total number of elements in all chunklets, the covariance matrix $\hat{\mathbf{C}}$ is computed as in (1). Then, the feature space transformation defined as in (2) is conducted.

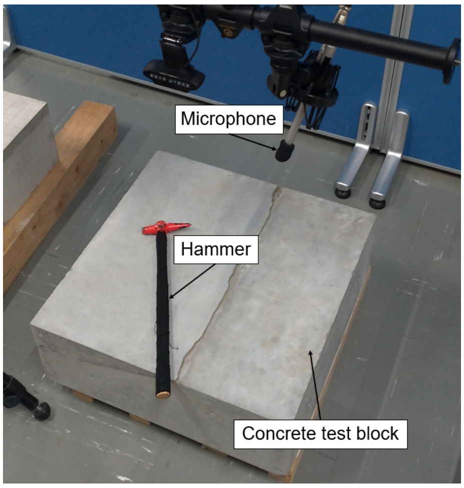


Fig. 3. Experimental setup.

$$\hat{\mathbf{C}} = \frac{1}{N_{total}} \sum_{j=1}^{N_{chunklet}} \sum_{\mathbf{x}_i \in \mathcal{M}_l} (\mathbf{x}_i - \hat{\mathbf{m}}_l)(\mathbf{x}_i - \hat{\mathbf{m}}_l)^T, \quad (1)$$

$$\mathbf{x}_i \rightarrow \hat{\mathbf{C}}^{-1/2}\mathbf{x}_i. \quad (2)$$

III. EXPERIMENTS

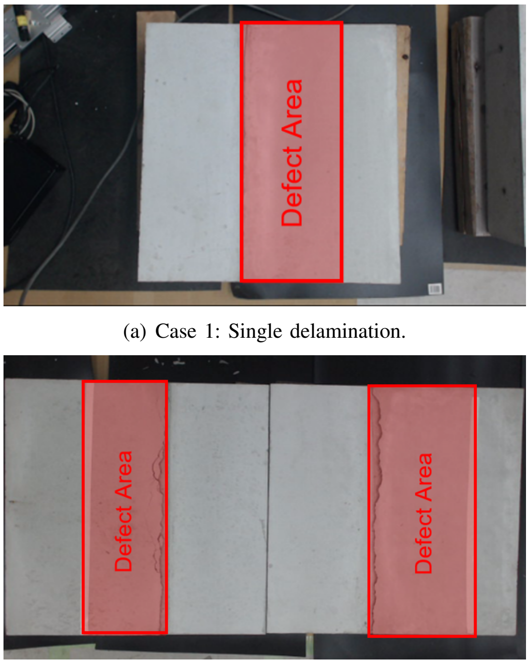
Experiments were conducted in laboratory conditions using concrete test blocks with the setup illustrated in Fig. 3. These concrete test blocks contain man-made defects of precisely known dimensions that simulate actual defects found on concrete structures. A KTC UDHT-2 hammer (head diameter 16 mm, length 380 mm, weight 160 g), commonly used in actual inspection sites, was used to hit the concrete blocks on several locations, once per location. A Behringer ECM8000 microphone was coupled with a Roland UA-25EX sound board for audio recording at 44.1 kHz, a common sampling frequency for multi-media. MFCC feature vectors were computed with 26 filters and 10 coefficients as in [13].

Two cases were considered: Case 1 and Case 2. Illustrations are provided in Fig. 4 and schematics are shown in Fig. 5. Those were also considered in [13] and K-Means was conducted with the number of cluster setting $K = 2$. For both cases, the consolidation process was conducted with the manual setting of $K_c = 150$.

- Case 1: Single delamination. Illustrated in Fig 4(a), this dataset contains a single delamination, running at an angle of 30 degrees from the surface. It is composed of 462 samples, with 272 non-defects and 190 defects.
- Case 2: Dual delaminations. Illustrated in Fig 4(b), this dataset contains two distinct delaminations, both running at an angle of 15 degrees from the surface. It is composed of 270 samples, with 155 non-defects and 115 defects.

The following methods were compared in our experiments:

- (A) K-Means clustering on MFCC feature vectors of hammering samples.



(a) Case 1: Single delamination.

(b) Case 2: Dual delaminations.

Fig. 4. Picture of the considered cases in laboratory conditions. Red areas indicate defect areas.

- (B) K-Means clustering on the feature space defined by RCA with 20 must-links using random query, as proposed in [12].
- (C) K-Means clustering on the feature space defined by RCA with 20 must-links using the active query scheme using position information proposed in [13].
- (D) The proposed method consisting of K-Means clustering on the feature space defined by RCA with 20 must-links using random query after consolidation.

In all our experiments, the Rand index was used to evaluate performance [15]. The Rand index is a measure of clustering performance based on pairs of samples, ranging between 0 and 1. The closer the value is to 1, the better is the clustering.

IV. RESULTS AND DISCUSSIONS

In Fig. 6 are reported the average performance obtained for the considered methods over 20 sets of weak supervision.

For Case 1, it can be seen that K-Means on the initial feature space defined by MFCC feature vectors obtains a good result performance, with no variations in performance over 20 runs. This indicates that, as reported in [10], MFCCs are already quite suited for discrimination of defect hammering samples. Furthermore, the lack of variations even with the random seeding nature of K-Means hints at the convexity of the fitness function computed in this feature space. In contrast, the method of [12] shows large variations in performance, resulting in an average performance slightly under the one of K-Means on MFCC. This is due to the random nature in the selection of weak supervision, ultimately resulting in variations in the quality of the final feature space for discrimination of defects: depending on the quality of the

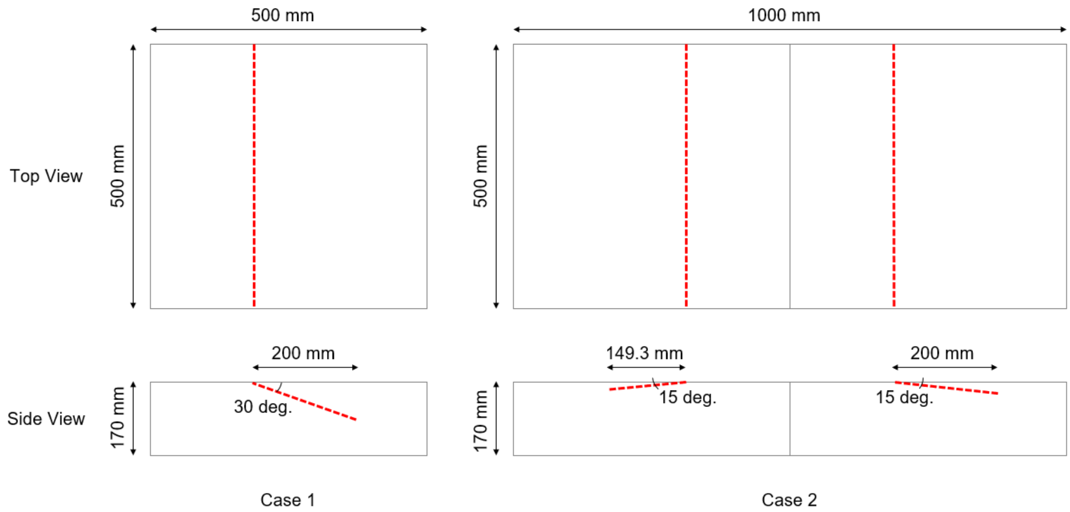
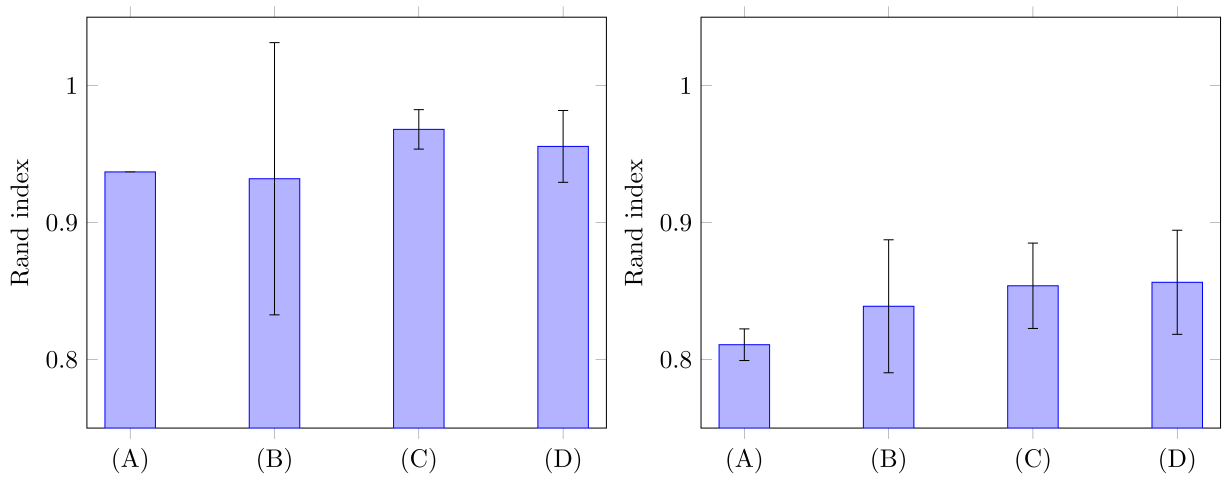


Fig. 5. Schematics for Case 1 and Case 2. Red dotted lines indicate delaminations.



(a) Case 1.

(b) Case 2.

Fig. 6. Performance evaluation of several weakly supervised methods for the hammering test in laboratory conditions: (A) K-Means on MFCC, (B) method of [12], (C) RCA with the active query scheme of [13] and (D) proposed method. Average of 20 runs with sets of 20 must-links each are reported. Error bars corresponds to one standard deviation.

obtained weak supervision, defect detection performance can be very good or lacking. Using the active query scheme based on position information proposed in [13] obtained the best results overall on average on Case 1, with also the smallest standard variation among the considered weakly supervised methods. This shows that the quality of the selected weak supervision by the active query scheme is good and is consistently good, thanks to the introduction of position information. Finally, our proposed method, only using audio data, achieved the second-best performance on average, only losing to the active query scheme using position information and showing a significant improvement over the method of [12].

Case 2 is a more difficult dataset, resulting in performances being overall lower than for Case 1. This is due to the presence of two distinct defects and the fact that the dataset is composed of two concrete test blocks: the clusters for defect and non-defects are less well defined, as indicated by the variations in the performance of K-Means on MFCC. In comparison with Case 1, the method of [12] shows a much smaller value of standard deviation. While this seems contradictory, especially with the results obtained in Case 1, this is explained by the lower number of samples in Case 2. This resulted in the random query process being much less penalizing. Still, this method still shows the highest value of standard deviation among the considered methods. The active query with position information and the proposed method obtained similar average performance, with our proposed method having a slight edge. The values of standard deviations are so comparable, with the active query

TABLE I

AVERAGE NUMBER±STANDARD DEVIATION OF CHUNKLETS OVER 20 RUNS OF 20 MUST-LINKS EACH.

	Case 1	**Case 2**
Random query [12]	18.3±1.5	16.8±1.8
Active query using position information [13]	18.5±1.1	17.3±1.2
Proposed method	15.0±1.6	15.4±1.7

using position information having a little more consistency. This is also certainly due to the smaller dataset size, naturally resulting in a smaller query pool over which the query selection process can exert its influence.

In Table I are reported the average number of chunklets obtained by each of the weakly supervised methods under the same conditions as the results in Fig. 6. With the same number of must-links, a lower number of chunklets indicates that more groupings using transitive closure were found and that the chunklets are bigger and closer to effective proto-clusters.

It can be seen that the active query scheme using position information has the highest number of chunklets among the considered methods for both Case 1 and Case 2. This is because this query scheme actively aims to spread the query process over the whole tested concrete area using position information. While this is done to increase the informativeness of each of the obtained individual must-link, it does also lower the chance of having common elements in pairs. Our proposed method achieves a much lower number of chunklets in both Case 1 and Case 2. This shows that the consolidation process of our proposed method contributes to obtaining repetitive elements from the query and therefore build fewer and bigger chunklets.

V. CONCLUSION

In this paper, a method for improving the quality of weak supervision for acoustic defect detection of concrete structures without the use of additional sensors was proposed. This was achieved by consolidating the initial dataset of hammering samples and conducting the query process on the consolidated dataset, allowing to avoid queries with little informative value as well as increasing the chance of obtaining bigger chunklets for the weakly supervised feature space transformation. Experiments conducted using concrete test blocks showed that the proposed method allows better and more consistent performance with a given amount of weak supervision, i.e., allows the gathering of weak supervision of better quality.

As future work, we would like to investigate the effects of the consolidation parameters, such as the number of clusters in the partitioning process, on the quality of weak supervision. Indeed, in this paper, this setting was manually tweaked for Case 1 and kept for Case 2 but it can be expected to have a negative influence on the quality of the obtained weak supervision for some value ranges. Furthermore, this paper focused mainly on positive weak supervision and we would like to extend our proposed approach to incorporate negative weak supervision as well.

ACKNOWLEDGMENT

This work was supported in part by the Japan Construction Information Center (JACIC) foundation. The authors thank Prof. Hiromitsu Fujii, Department of Advanced Robotics, Chiba Institute of Technology for his valuable help. The authors also thank the Institute of Technology, Tokyu Construction Co., Ltd. for their precious help and cooperation.

REFERENCES

[1] K. Nagatani and Y. Fujino, "Research and development on robotic technologies for infrastructure maintenance," in *Journal of Robotics and Mechatronics*, vol. 31, no. 6. Fuji Technology Press Ltd., 2019, pp. 744–751.

[2] BBC News, "Italy bridge collapse: Genoa death toll rises to 43. [web page]," https://www.bbc.com/news/world-europe-45241842, 2018, accessed: 2019-04-11.

[3] BBC News, "Japan Sasago tunnel collapse killed nine. [web page]," https://www.bbc.com/news/world-asia-20576492, 2012, accessed: 2019-04-11.

[4] Y. Takahashi, S. Nakamura, Y. Ogawa, and T. Satoh, "Velocity control mechanism of the under-actuated hammering robot for gravity compensation," in *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, vol. 34. IAARC, 2017.

[5] S. Nakamura, Y. Takahashi, D. Inoue, and T. Ueno, "The variable guide frame vehicle for tunnel inspection," in *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, vol. 34. IAARC, 2017.

[6] Y. Takahashi, S. Maehara, Y. Ogawa, and T. Satoh, "Concrete inspection systems using hammering robot imitating sounds of workers," in *Proceedings of the International Symposium on Automation and Robotics in Construction (ISARC)*, vol. 35. IAARC, 2018.

[7] H. Fujii, A. Yamashita, and H. Asama, "Defect detection with estimation of material condition using ensemble learning for hammering test," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3847–3854.

[8] J. Im, H. Fujii, A. Yamashita, and H. Asama, "Multi-modal diagnostic method for detection of concrete crack direction using light-section method and hammering test," in *Proceedings of the Ubiquitous Robots and Ambient Intelligence (URAI)*. IEEE, 2017, pp. 922–927.

[9] J. Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Unsupervised learning approach to automation of hammering test using topological information," in *ROBOMECH Journal*, vol. 4, no. 1. Springer, 2017, pp. 13–23.

[10] J. Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Clustering of spatially relevant audio data using Mel-frequency cepstrum for diagnosis of concrete structure by hammering test," in *Proceedings of the International Symposium on System Integration*. IEEE, 2017, pp. 787–792.

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

[12] J. Y. Louhi Kasahara, H. Fujii, A. Yamashita, and H. Asama, "Weakly supervised approach to defect detection in concrete structures using hammering test," in *Proceedings of the International Conference on Consumer Electronics*, 2019, pp. 997–998.

[13] J. Y. Louhi Kasahara, A. Yamashita, and H. Asama, "Acoustic inspection of concrete structures using active weak supervision and visual information," in *Sensors*, vol. 20, no. 3. Multidisciplinary Digital Publishing Institute, 2020, p. 629.

[14] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning distance functions using equivalence relations," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2003, pp. 11–18.

[15] L. Hubert and P. Arabie, "Comparing partitions," in *Journal of classification*, vol. 2, no. 1. Springer, 1985, pp. 193–218.