Multi-modal Classification Using Domain Adaptation for Automated Defect Detection Based on the Hammering Test

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

***Abstract*— Inspecting concrete structures such as tunnels is very important to keep them safe and durable. Due to the shortage of human inspectors, automated system for inspection is highly required. Hammering test is one of the popular inspection methods, and previous studies proposed automated systems for hammering test. Most works based on machine learning models to train a classifier to recognize hammering sounds suffer when the training data is not adequate for the considered data during deployment. This problem is also known as domain gap problem. In this paper, a methodology for concrete defect detection even when the available training data was collected from a tunnel that differs from the actually inspected tunnel is proposed. The proposed method selects part of the data from the inspection target tunnel, for which labels are not available, to use along traditional labeled training data in the training of a classifier within the semi-supervised support vector machine framework. This selection is conducted using the integration of visual information from an ordinary camera and acoustic information obtained using the hammering test. Experimental results showed that the proposed method yielded satisfying results in the laboratory conditions.**

Fig. 1. Hammering test by a human inspector

I. INTRODUCTION

Inspecting concrete structures such as tunnels regularly is very important to guarantee user safety. The importance of inspection was highlighted following tragic accidents such as the collapse of the Sasago tunnel in Japan [1]. The number of tunnels that should be inspected is increasing in the world. Figure 1 shows the hammering test, a very popular inspection method. Human inspectors hit the concrete surface with a hammer, listen to the sound, and identify whether a defect is present or not. Areas with visible cracks on the surface have a high priority for inspection, because those pose the threat of large concrete slabs breaking off and falling off the structure. Cracks on the surface are indicative of the presence of a defect propagating beneath the surface. Though hammering test is more efficient and simpler than other methods, there are some problems. Firstly, the number of skilled human inspectors is decreasing these days. Secondly, it costs a lot to train and maintain young human inspectors. Therefore, developing an automated method for hammering test is highly desirable.

Many studies about the robots for the automation have been done [2][3]. As for the algorithm, there are several studies that aimed at developing the automated system of the hammering test. In [4][5], authors investigated sound generation theoretically. In the paper it is indicated that the sound amplitude is useful for detecting defects, but it is difficult to apply to real, complicated environment. The sounds from hammering test in real situation vary widely due to the features of a tunnel like the compounds, degradation states, size and so on. Therefore, several approaches based on machine learning have been proposed. In [6][7], authors adopted an unsupervised learning based approach and proposed clustering methods. However, such methods can only classify sound data into several classes; they cannot identify if the hammering sound is defect or non-defect.

On the other hand, in [8], authors adopted a supervised learning based approach. Generally, the purpose of supervised learning is to predict labels of test data. For this purpose, a model is made from training data and their labels. The proposed method is based on the ensemble-learning method to improve the performance of the detection. However, the model suggested by the author would not work well if a domain gap between training data and test data exists; that is, when the test data from the inspected tunnel is different from the training data, which was collected from another tunnel, due to the difference of compounds, degradation status and so on. Therefore, in [9] authors proposed a method to improve environmental adaptivity of defect detection with a boosting algorithm. To improve environmental adaptivity, the method adopts onsite calibration where some of the test data is labeled by a human inspector, and those are used along with the training data to make a model. It is shown to be robust to the environmental noise but it needs human inspectors for

*This work was supported in part by JSPS KAKENHI Grant Number JP21K17829 and the Suzuki foundation.

[1]K. Ushiroda, J.Y. Louhi Kasahara, A. Yamashita and H. Asama are with the Department of Precision Engineering, Faculty of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ushiroda@robot.t.u-tokyo.ac.jp

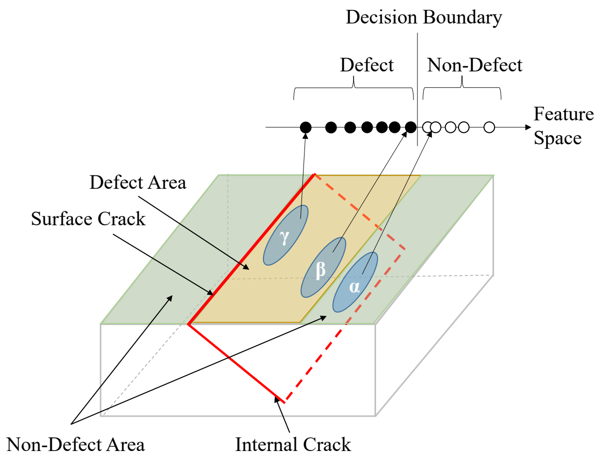

Fig. 2. A piece of concrete with a surface and internal crack. The yellow painted area is defect area, while the green painted area is non-defect area. The parts named: α, β and γ, indicate the corresponding location of the data in a feature space.

labeling yet. From that perspective, issues remains regarding the automation of the hammering test.

Developing a new method to be able to detect defects even though there is a domain gap between training data with labels and test data without labels is needed. In [10], authors adopted semi-supervised learning based approach to tackle this problem. They made a model using both training data with labels and test data without labels. However, in [11], it was pointed out that the result of prediction by semi-supervised approach may be worse than that by ordinary supervised based approach in some cases. This is because some unlabeled data can mislead the model. This problem is not considered well for the hammering test yet.

Therefore, in this study, we aim to propose a new automated method for defect detection using the hammering test that is effective through the domain gap. For this goal, we propose two concepts. Firstly, adopting semi-supervised approach and only selecting useful test data to be mixed with training data to make an improved model for detection. Secondly, we focus on the characteristics of concrete tunnels and define the ”usefulness” of test data.

II. CONCEPT

The problem setting is that there is a training data with labels from tunnel A, and there also is a test data without any labels from tunnel B. The data is hammering sound data and visual information of the surface. The labels on hammering sound samples are binary: Defect and Non-Defect. The purpose is to predict the labels of test data.

Semi-supervised support vector machine (S3VM) [12] is one of the semi-supervised learning methods. S3VM uses all the training data and all the test data without labels to make a decision boundary; thus, S3VM can deal with the domain gap problem between the training data and the test data in theory.

Our proposed method basically adopts S3VM, but only some selected test data is used to make a decision boundary. In order to select some data, we focus on a specific characteristic of concrete: cracks on the surface; we use only data collected close to surface cracks. As shown in Fig. 2, a location of a hammering point on the concrete surface is correlated to its relative location with other samples in the feature space. This is because the depth of the internal crack from the surface is a primary factor of the sound for the hammering test. When the depth of defect from the surface is shallow, the hammering sound on the location is treble and the concrete vibrates for a long time; while, when the depth is deep, the hammering sound is dull like the hammering sound of non-defect and the concrete vibrates for a short time [13]. Defect points close to a surface crack are far from the decision boundary in a feature space, while points that are far from the crack are close to the boundary. On the other hand, all non-defect points are on the opposite site of the boundary in the feature space. Even though which side is defect or non-defect is unknown after only collecting data in a target tunnel, in the proximity of a surface crack is present the most disparate set of hammering samples, i.e., the sample set with the largest distance across the boundary in the feature space can be acquired. We believe using such data collected close to a surface crack can contribute to make a better model for S3VM.

Therefore, the key concepts of our proposed method can be summarized as follows:

- Selecting only the helpful part of the test data to deal with a domain gap problem for hammering test
- Focusing on a surface crack, a characteristic of a concrete, and adopting multi-modal information; integrating acoustic and visual information

III. PROPOSED METHOD

A. Overview

An overview of our proposed method is shown in Fig. 3. First of all, the training data consist of acoustic data and their corresponding binary labels. They are made from tunnel A in advance. Acoustic and visual information of tunnel B, that is a target tunnel for inspection, is collected with a microphone and camera. Acoustic data is initially time-series data and is processed into Fourier spectrum using Fast Fourier Transform (FFT) and normalized. Visual information is initially RGB data. The crack locations and the locations of the hammering points are detected from the visual information. Then, acoustic data is divided into two groups according to the visual information. One is the group to be defined as useful points and used along with training data to train a model while the other group is just used as ”test” data that is predicted by the model. In order to train a model, S3VM approach is basically adopted. Finally, the predicted labels of the test data are obtained with the trained model.

B. Preprocessing

Training data is obtained by inspection of a tunnel by a human inspector. The training data is expressed by $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$. $\mathbf{x}_i$ describes an acoustic information of the $i-th$ hammering

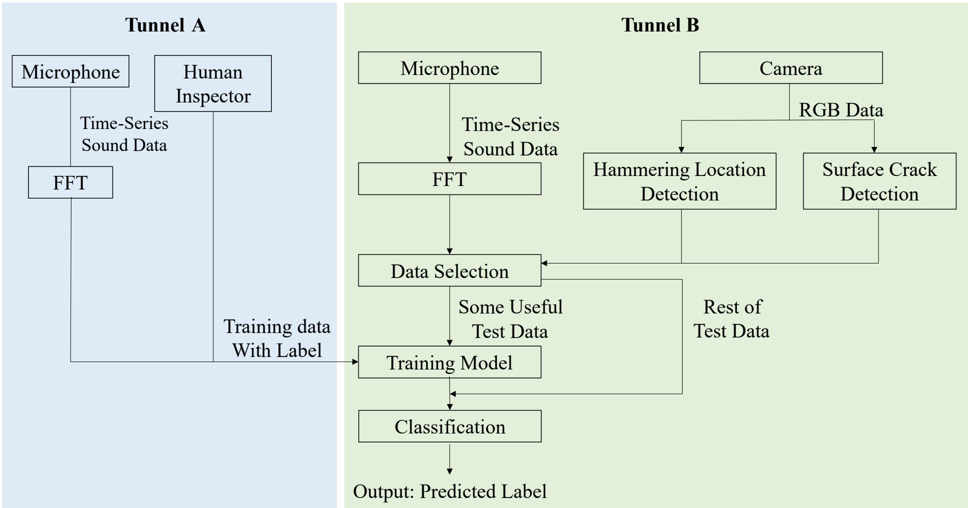

Fig. 3. Overview of our proposed method

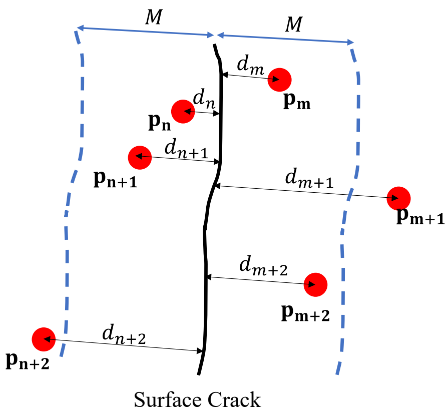

Fig. 4. Test data selection in terms of distance between the locations of the surface crack and the locations of the hammering points

point. y_i describes a binary label expressing defect or non-defect. When the hammering test is carried out for another tunnel, test data is obtained. The test data is expressed by $\{\mathbf{x}_i\}_{i=l+1}^{l+u}$.

C. Data Selection of Test Data with Visual Information

In order to detect the locations of a surface crack, visual information from a camera and a computer vision technique [7] are used. The RGB image of the concrete surface is converted to a gray scale image and the crack is expressed as the set of the lower brightness pixels. The locations of hammering points are detected with a image processing technique: the center of the hammer head is detected using its color. $\{\mathbf{p}_i\}_{i=l+1}^{l+u}$ describes a hammering location in the 2D coordinates of the image. The locations of the surface crack is also expressed by $\{\mathbf{q}_i\}_{i=1}^{n}$. $\mathbf{q}_i$ are the two dimensional coordinates of the pixel detected as a surface crack.

Positional information is used in order to select some data from whole test data for the model training. As Fig. 4 shows, the hammering samples located in close proximity of the surface crack are useful to deal with the domain gap. A distance between a location of each hammering sample and the crack is calculated as below:

$$d_i = |\mathbf{p}_i - \mathbf{q}_{closest}|_2^2, \quad (1)$$

where $\mathbf{q}_{closest}$ denotes the closest pixel among all pixels detected as crack to the hammering point $\mathbf{p}_i$.

With a parameter M, manually set for each test data set, the set Ψ of selected points are defined as below:

$$\Psi = \{\mathbf{x}_i | d_i < M, i = l+1, \ldots, l+u\}. \quad (2)$$

D. Training Model and Prediction

A classification model is trained from training data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$ and the selected data set Ψ to predict the labels of the test data.

The formulation of S3VM is described as below:

$$\min_{f,\mathbf{y}'} \frac{1}{\mathbf{l}} \sum_{i=1}^{l} \phi(y_i, f(\mathbf{x}_i)) + \frac{\lambda'}{|\Psi|} \sum_{\mathbf{x}_j \in \Psi} \phi'(y'_j, f(\mathbf{x}_j)) + \lambda ||f||_{\mathscr{H}}^2, \quad (3)$$

where λ and λ' are parameters, ϕ and ϕ' are loss functions, f is an objective function in a reproducing kernel Hilbert space.

Optimizing Eq. (3) is difficult because both objective function f and predicted labels $\mathbf{y}'$ are estimated at the same time. The loss function ϕ' is also non-convex function due to the absence of the labels of test data; thus the optimizing technique using Quasi-Newton method [14] is adopted.

After the optimization, the prediction function f is obtained and labels of the test data are predicted. The final output is $\hat{y}_i (i = l+1, \ldots, l+u)$. Each $\hat{y}_i$ denotes if the $i-th$ hitting point is defect or non-defect.

IV. EXPERIMENT

A. Experimental Setting

The experimental setting is illustrated in Fig. 5. We used three concrete test pieces that imitate real tunnel, designed

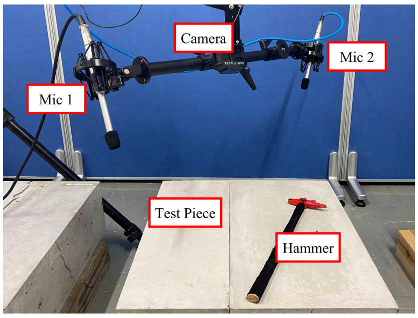

Fig. 5. Experimental setting

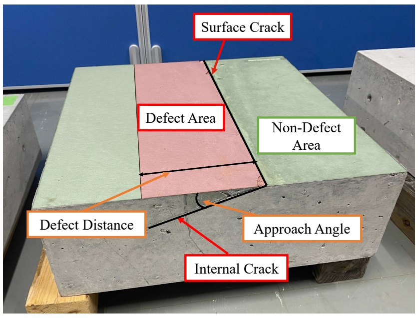

Fig. 6. Concrete test piece

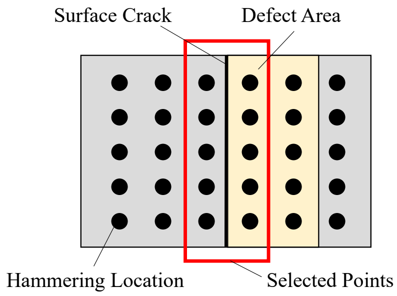

Fig. 7. Illustration of data selection with our proposed method

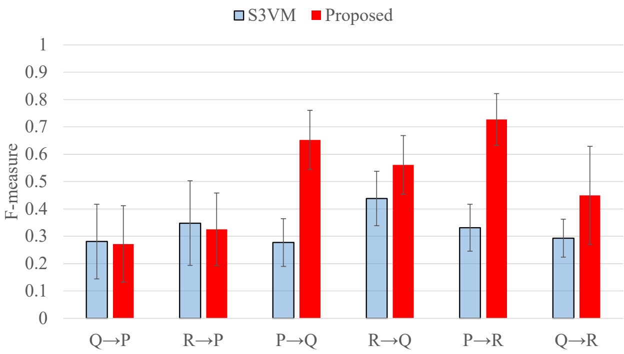

Fig. 8. Performance evaluation of S3VM and Proposed method for the hammering test with three test pieces: ”X $\rightarrow$ Y” denotes ”X” was the test piece used for training data and ”Y” was the test piece used for test data. The averages of 100 iterations were reported. Error bars correspond to one standard deviation.

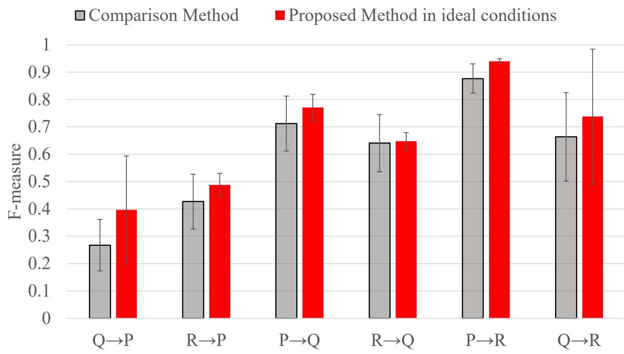

Fig. 9. Performance evaluation of *Comparison Method* and *Proposed Method in ideal conditions* in the additional experiment. The averages of 100 iterations were reported. Error bars correspond to one standard deviation.

as P, Q and R. As shown in Fig. 6, a test piece has a defect that runs slantingly to the surface. Therefore, when the defect area is hit with a hammer, the defect sound can be collected. The characteristics of the three pieces are shown in Table I. Each test piece has a different kind of defect.

The flow of the experiment is as below: firstly, gathering acoustic and visual information from each from the three test pieces. Several samples of each test piece were hit with a hammer usually used in inspection sites. The hammer was a KTC UDHT-2 (length 380 mm, weight 160g and head diameter 16 mm). The sound was recorded at 44.1 kHz using two BEHRINGER ECM8000 microphones and a Roland EDIROL UA-25EX audio interface. The visual information was taken with a logicool Carl Zeiss Tessar HD 1080p webcam. The defect and non-defect ground truth was obtained here for each hitting point using the schematics of each test piece. Secondly, pairs of pieces were chosen in the experiments: one used as source of training data and the other used as source of test data. Therefore, 6 experiments were conducted in total. Lastly, our proposed method and another method for comparison were used to predict the label of test data.

Figure 7 shows the illustration of the test piece and data selection for our proposed method. We hit the surface of the test piece at regular intervals. The parameter M was manually configured for each test piece to encompass the first rows on both sides of the surface crack.

To verify the effectiveness of our proposed method, S3VM [12][14] method was used. S3VM and our proposed method have optimization process; thus, we iterated for 100 times to deal with the dispersion in each experiment. F-measure was used to evaluate the performance.

TABLE I
THE CHARACTERISTICS OF THREE CONCRETE TEST PIECE

Name	P	Q	R
Defect Distance [mm]	40.0	149.3	200.0
Approach Angle [deg]	45	15	30
Number of hitting Points (Non-Defect, Defect)	54, 9	105, 65	91, 78

B. *Result And Discussion*

As shown in Fig. 8, proposed method performed better than S3VM in four cases: ”P $\rightarrow$ Q”, ”R $\rightarrow$ Q”, ”P $\rightarrow$ R” and ”Q $\rightarrow$ R”. Especially in the case of ”P $\rightarrow$ R”, the performance of proposed method was the best and the F-measure was almost 0.73. On the other hand, S3VM performed better than proposed method in the cases of ”Q $\rightarrow$ P” and ”R $\rightarrow$ P”.

There are three major factors for these results. Firstly, the ratio of the number of samples of the training data to test data can be very important on the model training process. Both S3VM and proposed method used test data to train a model so that the model can deal with the domain gap between training data and test data. But if the amount of test data is much smaller than that of training data, the performance would be worse due to the small effect in the model training process. As shown in Table I, the dataset from test piece P was much smaller than from Q and R, and thus much smaller amount of data was selected from test piece P. Therefore, the result of proposed method could be worse than that of S3VM. Secondly, the ratio of the amount of defect labels to non-defect labels, called class balance, affected to the results. As shown in Fig. 6 and 7, each test piece was hit with a hammer in regular rows parallel to the line of the surface crack. So proposed method selected an equal amount of defect and non-defect samples across the crack for the model training, and the class balance was 1:1. On the other hand, S3VM used all points, and the class balance did not change. The S3VM method [14] makes an assumption that the class balances of both train data and test data are same. However, the class balance assumption was not fulfilled in our experimental setting due to the impossibility of knowing any label of test data, so the class imbalance problem may have occurred. Lastly, the key concept of our study can influence the result positively. Especially the results of proposed method were much better than the S3VM method in the cases of "P $\rightarrow$ Q", "R $\rightarrow$ Q" and "P $\rightarrow$ R", and these results were enabled by selecting part of the test data to be used for training the model.

C. Additional Experiment

To verify the effectiveness of proposed method's concept regarding the proximity of samples selected for the domain gap problem with the surface crack, an additional experiment in ideal conditions was conducted. Proposed method was the basis of both methods, but among the previously mentioned factors, two were made equal; that is, the ratio of the number of samples of the training data to test data and the class balance of the training data and selected data were set to the same value in each example. The difference between the two methods is that; while *Proposed Method in ideal conditions* selects samples near to a surface crack, *Comparison Method* selects samples randomly from all test data. The two methods were also iterated for 100 times due to the dispersion in each experiment.

As shown in Fig. 9, *Proposed Method in ideal conditions* performed better than *Comparison Method* in the all cases. The results show that the samples close to a surface crack work well for dealing with a domain gap if the ratio of the number of samples of training data to test data and the class balance of training data to selected data in the model training process were fulfilled. Comparing the proposed method in Fig. 8 and *Proposed Method in ideal conditions* in Fig. 9, the two factors: ratio of training data to test data and class balance, had a pretty high influence on the F-measure. While information about the ratio of training data to test data can be obtained, obtaining information about the class balance in the scenario of this study is difficult. Focusing on these two factors could lead to significant improvements of the proposed method.

V. CONCLUSION

A method using semi-supervised support vector machine and selecting useful data from test data to deal with the domain gap problem by using multi-modal information and focusing on the concrete characteristics was proposed. The visual information is helpful to select the closest hammering points from a surface crack, and using such data is effective to train a model with training data. Experiments with three different concrete test pieces showed the effectiveness and potential of our proposed method.

As future work, there are three points to be considered. Firstly, the class balance problem and the ratio of the amount of training data and test data are highly important and are to be tackled to improve the performance. Generally, the class balance of test data cannot be known in the situation of automation; however, focusing on dividing the data into several groups based on their similarity will be useful to deal with the class balance problem. Secondly, adopting other kinds of information may also be helpful. Lastly, the idea that each side of the crack has a different label could be also very useful to improve our proposed method, so it can be used as a constraint in the optimization process to improve the prediction accuracy.

REFERENCES

[1] Martin Fackler: "Nine Killed When Highway Tunnel Collapses in Japan", New York Times, December 2, 2012, http://www.nytimes.com (accessed September 1, 2021).

[2] Roberto Montero, Juan G. Victores, Santiago Martínez de la Casa Díaz, Alberto Jardón Huete and Carlos Balaguer: "Past, present and future of robotic tunnel inspection", Automation in Construction, vol. 59, pp. 99-112, 2015.

[3] Yusuke Takahashi, Satoshi Maehara, Yasuichi Ogawa and Tomoya Satoh: "Concrete Inspection Systems Using Hammering Robot Imitating Sounds of Workers", Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC 2018), pp. 206-210, 2018.

[4] Yoshimi Sonoda and Yuki Fukui: "A Basic Study on Hammering Tests of Deteriorated Concrete Structures", Proceedings of the 35th Our World in Concrete and Structures, 2010.

[5] Masanori Asano, Toshiro Kamada, Minoru Kunieda and Keitetsu Rokug: "Impact acoustics methods for defect evaluation in concrete", Non-Destructive Testing in Civil Engineering, BB 85-CD, 2003.

[6] Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama: "Complementarity of Sensors and Weak Supervision for Defect Detection in Concrete Structures", Proceedings of the 2020 IEEE/SICE International Symposium on System Integration (SII2020), pp. 1-6, 2020.

[7] Jun Younes Louhi Kasahara, Shinji Minato, Alessandro Moro, Hanwool Woo Atsushi Yamashita and Hajime Asama: "Defect Detection in Concrete Structures Using Sensor Fusion of Force Sensor and Camera", Journal of the Japan Society for Precision Engineering, vol. 86, no. 12, pp. 975-981, 2020.

[8] Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama: "Defect Detection with Estimation of Material Condition Using Ensemble Learning for Hammering Test", Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA2016), pp. 3847-3854, 2016.

[9] Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama: ”Improvement of Environmental Adaptivity of Defect Detector for Hammering Test Using Boosting Algorithm”, Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2015), pp. 6507-6514, 2015.

[10] Ye Janxing, Takumi Kobayashi, Hiroshi Tsuda and Masahiro Murakawa: “Robust Hammering Echo Analysis for Concrete Assessment with Transfer Learning”, Structural Health Monitoring-an International Journal, pp. 943-949, 2017.

[11] Yu-Feng Li and Zhi-Hua Zhou: ”Towards Making Unlabeled Data Never Hurt”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, no. 1, pp. 175-188, 2015.

[12] Kristin P. Bennett and Ayhan Demiriz: ”Semi-supervised support vector machines”, Advances in Neural Information Processing Systems, vol. 11, pp 368-374, 1998.

[13] Jun Younes Louhi Kasahara, Hiromitsu Fujii, Atsushi Yamashita and Hajime Asama: ”Unsupervised learning approach to automation of hammering test using topological information”, ROBOMECH Journal, vol. 4, 13, pp.1-10, 2017.

[14] Fabian Gieseke, Antti Airola, Tapio Pahikkala and Oliver Kramer: ”Fast and Simple Gradient-Based Optimization for Semi-Supervised Support Vector Machines”, Neurocomputing (ICPRAM 2012 Special Issue), vol. 123, pp.23-32, 2014.