ICIC Express Letters
Volume **17**, Number **8**, August **2023**

ICIC International ©2023 ISSN 1881-803X
pp. **873–880**

ENSEMBLE METHOD FOR SHAPE DEFECT CLASSIFICATION OF RUBBER PRODUCTS

Hyemee Kim[1], Hyerim Bae[1,*], Jinhee Park[2] and Minho Seo[2]

[1]Industrial Data Science and Engineering, Department of Industrial Engineering
Pusan National University
2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 46241, Korea
khm0219@pusan.ac.kr; *Corresponding author: hrbae@pusan.ac.kr

[2]Convergence Automation Team
DRB Holding Co., Ltd.
28, Gongdandong-ro 55beon-gil, Geumjeong-gu, Busan 46329, Korea
{ park.jin.hee; seo.min.ho }@drbworld.com

Received November 2022; accepted February 2023

Abstract. *In most manufacturing industries, defect detection technologies using artificial intelligence (AI) have gained popularity. The usage of image-based AI for product shape defect inspection has significantly increased. In rubber manufacturing, the shape is closely related to quality, which is an important factor in production. This study investigates a classification model for quality control of the shape using measurement and image datasets. In contrast to previous studies that used only images, an ensemble model was used to reflect the shape size together. In addition, contrary to the general artificial neural network model, which can only be informed for a given label, we propose local labeling and multiple attention model structures that support the analysis of the deciding factor for classification. Using the proposed method, it was possible to examine the parts that need to be reviewed in the image, as well as important variables. This function makes it possible to classify and analyze the factors at a lower cost. Thus, this method can be used in early application fields with insufficient labeled data.*
Keywords: Defect detection, Classification, Neural network, Convolutional neural network, Attention, Shape defect, Localized labeling

1. **Introduction.** Quality management is an important task in rubber manufacturing. Among the quality measures, the cross-section of rubber products is essential. The cross-section is the cut side of the rubber product. The final rubber product is made by contacting two sub-products. Thus, the cross sections of the two sub-products should be the same, as this is an important measure of product quality. In particular, the shapes of the two cross-sections must be the same to manufacture high-quality products. However, the inspection of cross-section quality is a manual task. In the inspection task, some parts of the product are sampled and compared with an ideal cross-sectional shape to determine whether they are defective. Although the method is based on sampling, it cannot detect if a defect exists in products that are not sampled. The inspection results may vary depending on the skill level of the operator because the procedure is manual. In addition, the efficiency of work is not constant. Therefore, a system for mechanically inspecting defective products is required.

As smart factories develop, many studies on automatic defect classification have been conducted. Because most shape defect detection entails visual inspection, recent studies have focused on models that are classified based on images. These studies typically use a convolutional neural network (CNN) which is an appropriate model for the image dataset.

DOI: 10.24507/icicel.17.08.873

Most of the studies have applied vanilla CNN models [1-3]. Recently, new CNN structures have been proposed, such as the comprehensive attachment network (CAN) [4].

However, previous studies focused on methods for learning defective factors that have specific labels and finding particular labels in an entire image [4,5]. A study similar to ours, which detects defects in a pantograph slide, also uses labeled data for different types of shape defects [6]. Labeling this type of defect is costly. Therefore, it is inevitable to learn a classification model using data that are not sufficiently labeled in the early factories applying this technology. However, it may not be easy to provide sufficient information.

Therefore, the data used to determine the defect factor using an image must contain sufficient information, such as a local label. However, sufficient data are rarely obtained. This study solves this problem by obtaining additional information on the decision-making process using a model structure. The data used in this study only had labels for defectiveness; there was no localized labeling. In addition, unlike previous studies that used only images, this study used shape-measurement data to classify shape defects.

This study introduces a method for inspecting whether a rubber shape is defective based on neural network algorithms. A cross-sectional image and the measured shape size were used to determine the labels. In addition, this method involves identifying factors that are believed to be the cause of defects in the classification model.

2. **Related Works.** This study aims to combine two models for different data types and interpret the results of artificial neural networks (ANN), a black-box model. Thus, we introduce two ANN-based models that can obtain elements to explain: label pooling [8] for image classification and attention model [10] for measurement classification.

LabelPooling is a model for image classification and can be labeled partially. The model is an advanced one of CNN, which can be classified from problems of multiple labels to obtain each probability for multiple labels simultaneously by region.

The measurement variables can be represented by a vector. Although this form typically exhibits high classification performance in MLP, MLP models cannot be interpreted, unlike machine learning-based models, SVM and XGBoost. Attention model is introduced as an MLP model that solves this problem. Attention is a method of generating a layer capable of obtaining weights for input data and reflecting the layer in input data. The weight layer is called the attention layer, and the results can be explained as the importance of each input variable.

3. **Methodology.** This study proposes an ensemble model in Figure 1 that uses two different data types. A CNN model for applying image datasets was a part of the model used as an image classifier. The other part, the measurement classifier, was the attention model for the measurements. The two predictions of these models were used as the final prediction value using a soft voting algorithm.

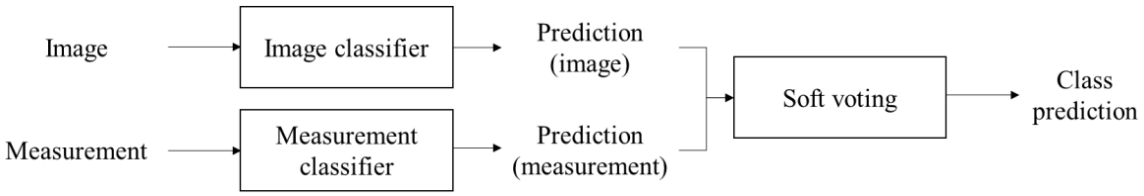

FIGURE 1. Illustration of the ensemble process

Furthermore, the architectures of the sub-models, such as images and measurement classifiers, included layers capable of representing partial causes of the classification results. A model architecture called local labeling in the image classifier identified local factors as the causes of classification. In addition, *multiple attention* was applied as a model structure to determining the importance of each measurement variable.

This section describes the two classifiers for determining the importance of these local factors and variables. Soft voting is an ensemble method that combines the predictions of the two classifiers.

3.1. **Image classifier with local labeling.** The CNN performed best in image classification problems [7]. Therefore, this study used a CNN as the primary image classifier model. In addition, a model structure known as the local labeling CNN was added to the basic model (vanilla CNN) for extracting local factors. Local labeling is a modification of LabelPooling [8] which is a method of locally labeling multiple classes.

In vanilla CNN, the feature map was obtained through the convolutional layers when the image was input. Furthermore, the obtained feature map was calculated as a vector through the global pooling layer, which predicted the class through the fully connected layer. The process of obtaining a feature map from an image in the proposed model (i.e., the local labeling CNN) was the same as that of the vanilla CNN, as shown in Figure 2. The feature map is then converted into a label map that can be used as a local label via a 1×1 convolutional layer. The label map was then flattened, converted into a vector, and passed through a fully connected layer to predict the class.

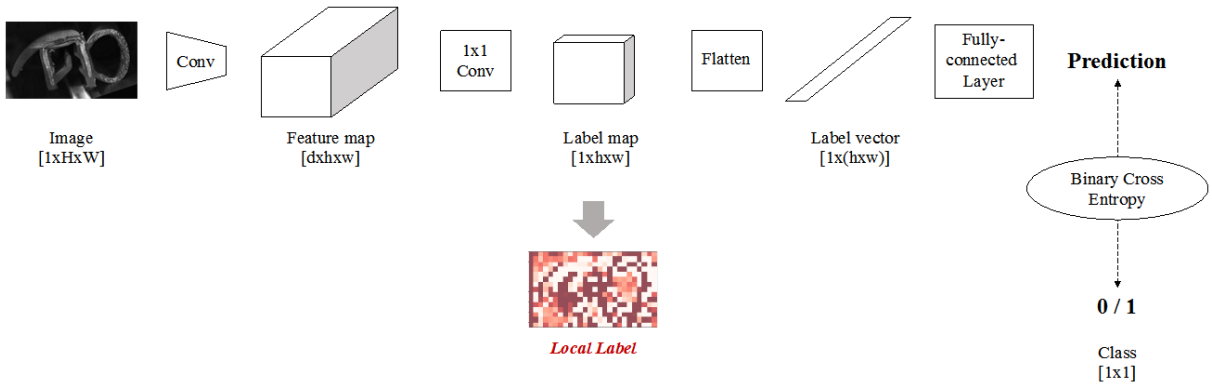

FIGURE 2. Process of local labeling CNN

3.2. **Measurement classifier with multiple attention.** The basic model of the measurement classifier was the multilayer perceptron (MLP) [9]. In addition, an attention block [10] was applied to the MLP input unit to select and weigh important variables from the measurement data. This attention block consisted of multiple parallel attention layers called multiple attention layers, as shown in Figure 3. This was to correct the biased attention rate, similar to the well-known method of multihead attention [11].

The multiple attention averages the n parallel attention rates A_i from the input data to calculate the final attention rate A_{fin}, as shown in Formula (1). The attention rate estimates the importance of each variable from the input value. Then, the final attention rate obtained through this process is multiplied by the input value in the same manner as in the basic attention algorithm in Formula (2). Finally, the obtained value is placed in the MLP to predict the class. This process consists of end-to-end learning, and the parameters of the attention layer that can output the attention rate suitable for the classifier MLP are learned. In this case, the final attention rate trained through this process can be used to analyze the importance of each variable.

$$A_{fin} = \frac{\sum_{i=1}^{n} A_i}{n} \tag{1}$$

$$\begin{aligned} Attentioned\ Layer &= A_{fin} \times Input \\ &= \{a_1, a_2, \ldots, a_d\} \times \{input_1, input_2, \ldots, input_d\} \\ &= \{a_1 \times input_1, a_2 \times input_2, \ldots, a_d \times input_d\} \end{aligned} \tag{2}$$

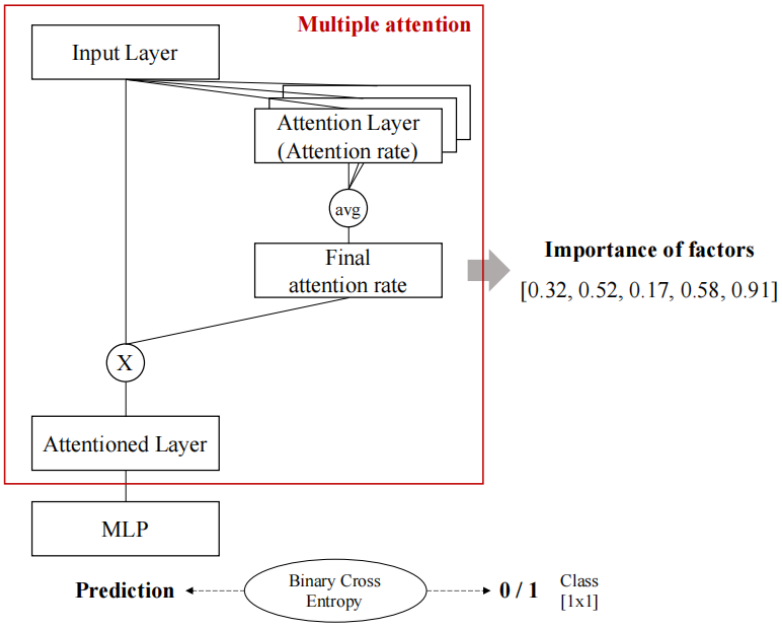

FIGURE 3. Process of multiple attention MLP

3.3. **Soft voting.** Voting is a well-known ensemble technique. This study used soft voting [12] for the two ensemble classifiers. The final predicted class was generated as a result of the ensemble model. The class is determined using (3) in a binary classification case. The condition of Formula (3) is the average of the image prediction result $Pred_{img}$ and the measurement prediction result $Pred_{measure}$.

$$Soft\ \ Voting = \begin{cases} 0, & \dfrac{Pred_{img} + Pred_{measure}}{2} < 0.5 \\ 1, & \dfrac{Pred_{img} + Pred_{measure}}{2} \geq 0.5 \end{cases} \tag{3}$$

4. **Experiments and Results.** This section explains the dataset used and experimental results. In addition, this section describes the data, experiments, performance, and explanatory factors. First, we deal with the current state and pre-process the data. Next, we list the models to be compared with the proposed model and briefly describe the validation method. Finally, we present the experimental results, including the performance and explanatory factors.

4.1. **Dataset.** In this study, we used two types of data: an image dataset and a measurement dataset. The image data were photographs of an object collected using a camera during the manufacturing process. The measurement dataset was a measurement of the shape size using the segmentation model. The measurement dataset had seven input variables: width of part radius (L1), height of part radius (L2), excess specification dimension for part size (L_S), outside width (Q1), inside width (Q2), total height (H), and overall width (W).

For these experiments, the data volume was 27,252 rows. The normal data had 26,532 rows, whereas the abnormal data had only 720 rows. In particular, the abnormal data were too few compared to normal data for the model to be trained. To solve this problem, undersampling was used for model training. Specifically, we constructed a dataset such that the amount of abnormal data for training was 80% of the total abnormal data and the amount of normal data for training was 120% of the abnormal training data. In this case, the training dataset was randomly selected.

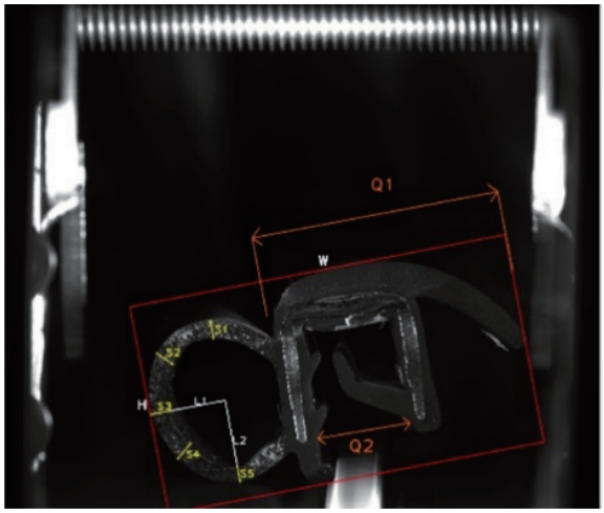

FIGURE 4. Example of image data and measurement data

The image contained additional objects that were not required. The target object was cropped from the original image to remove other objects. In addition, the target object and the background were similar in color, making them difficult to distinguish. Contrast-limited adaptive histogram equalization (CLAHE) was used to resolve this issue. CLAHE is a filter that equalizes the histogram from the color distribution of an image [13].

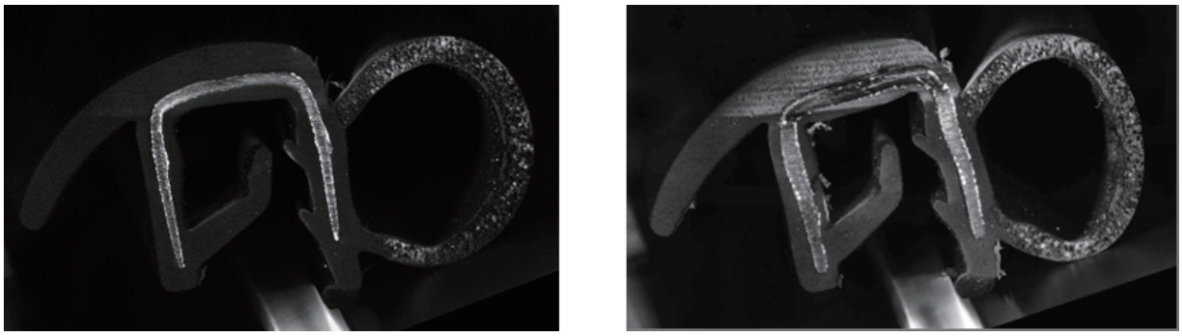

FIGURE 5. Example of filtered image (Left: Original, Right: CLAHE)

4.2. **Experiments.** We compared other models to verify the performance of the proposed model. Before validating the ensemble model, we validated it in a model where only a single data type was available, that is, a partial model of the ensemble. First, a CNN with local labeling (CNN_{local}) was validated using a vanilla CNN model with an image data model. Next, an MLP with multiple attention (MLP_{att}) was compared to support vector machine (SVM) [14], extreme gradient boosting (XGBoost) [15], and vanilla MLP as models of measurement data. Finally, an ensemble model utilizing both data types, that is soft voting, was validated. Here, only models containing explanatory elements were used as verification targets; CNN was used for images, while MLP and XGBoost were used for measurement data.

For a reliable performance estimation, the experiment in this study used k-fold validation. Five folds were used in the validation. The training dataset used only data randomly selected as undersampling, and the verification dataset utilized the entire data within the other four folds.

4.3. **Performance.** There were four measures for performance evaluation: accuracy, sensitivity, specificity, and F1-score. In this case, sensitivity refers to how well the normal is

identified out of the total normal and specificity refers to how well the abnormal is classified out of the total abnormal data. The F1-score is a classification performance measure of data imbalance. Therefore, the most critical indicator in this experiment, which was verified with unbalanced data, is the F1-score.

Table 1 presents the performance of the model with a single data type. Based on the F1-score, the image classification model has the highest performance of the CNN and the measurement classification model has the highest performance of the MLP_{att}. However, the two models did not differ significantly from the improved or base models.

TABLE 1. Experimental results for image and measurement dataset

Data type	Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
Image	CNN	**94.17**	**94.23**	91.94	**96.92**
	$\text{CNN}_{\text{local}}$	94.14	94.20	**92.08**	96.90
Measurement	SVM	79.79	79.96	73.19	88.51
	XGBoost	86.77	86.85	84.03	92.74
	MLP	86.80	86.97	**84.60**	92.76
	MLP_{att}	**86.89**	**86.98**	83.75	**92.82**

The result was an experiment that used an ensemble model. As described above, a CNN was used for the image model, and XGBoost and MLP were used for the measurement model. Accordingly, two ensemble models were verified, and the results are listed in Table 2. The results indicate that the MLP ensemble performed better than the CNN, which performed best on a single data type, while the XGBoost ensemble performed less than the CNN.

TABLE 2. Experimental results of the ensemble model

Method		Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)
$\text{CNN}_{\text{local}}$	XGBoost	89.78	89.06	83.33	94.49
$\text{CNN}_{\text{local}}$	MLP_{att}	**94.63**	**94.84**	**87.22**	**97.18**

4.4. **Explainable factor.** In addition to this model's predictions, other explanatory factors include local labels and the attention rate. Local labels are elements that allow one to view the labels of partial images as a result of an image classification model. The attention rate is the result of a measurement classification model that indicates the importance of each input variable.

The local labels are shown in Figure 6. The plot was divided into units, and each unit displayed the class in the form of probabilities between 0 and 1. This allows us to determine which part of each image is defective. For example, the dark part of Figure 6 is a unit with a high probability of defects. Since it was supported to determine whether it was defective by that part, it can be interpreted that there are defect factors in that unit.

The attention rate is the final attention obtained from multiple attentions. These values can be interpreted in conjunction with the variables. At this point, the attention rate changes the importance of the variable, depending on the input value. The changing importance is the most significant difference compared to the fixed importance of variables in other interpretable models, such as XGBoost. In addition, as shown in Figure 7, the attention rate can be expressed as a distribution to interpret fluctuations. In addition to the representative value of the attention rate, extreme values, such as outliers, can be interpreted.

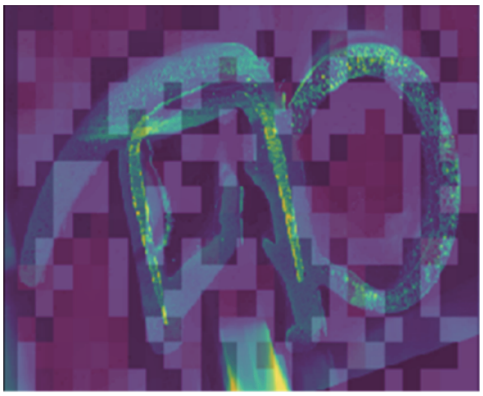

FIGURE 6. Examples of local labeling result

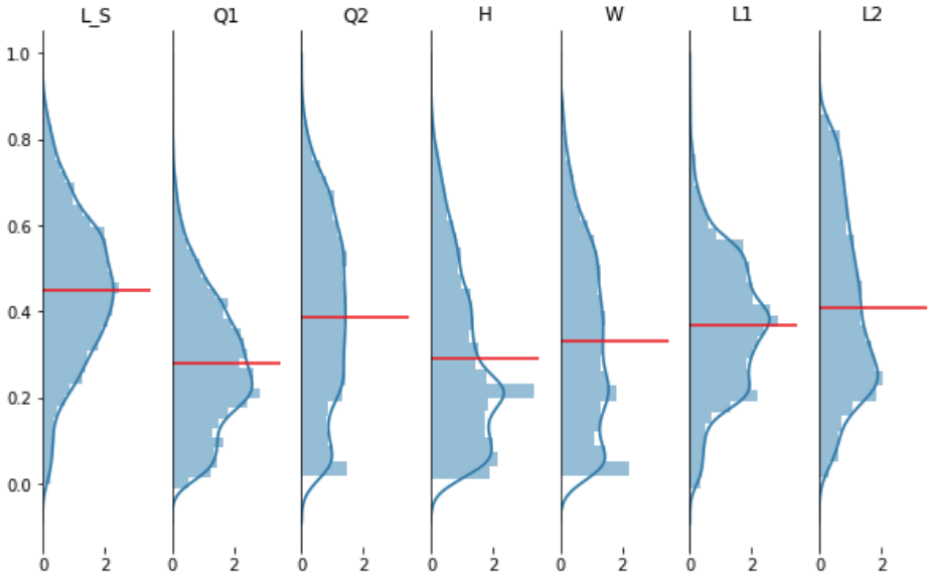

FIGURE 7. Factor importance in MLP with multiple attention

5. **Conclusions.** This study uses image data and measurements of cross-sectional shapes to classify defects in shapes. An ensemble method was used for both types of data. In addition, an artificial neural network model that can derive the factors by assuming the classification cause was proposed. In this case, a value that can be interpreted based on the input data was obtained. In particular, factors derived from measurements can be used immediately in decision-making. This model exhibited only a slight difference in performance.

In the case of localized labels in image data, it is difficult to analyze the cause because localized factors occur only in the divided unit without being classified by shape. These challenges are expected to be solved through research on labeling methods based on unsupervised learning.

Acknowledgment. This research was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program (IITP-2022-2016-0-00318) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (No. 2020R1A2C1102294).

REFERENCES

[1] M. Garg and G. Dhiman, Deep convolution neural network approach for defect inspection of textured surfaces, *Journal of the Institute of Electronics and Computer*, vol.2, no.1, pp.28-38, 2020.

[2] T. He et al., Application of deep convolutional neural network on feature extraction and detection of wood defects, *Measurement*, vol.152, 107357, 2020.

[3] L. Zhang et al., Convolutional neural network based multi-label classification of PCB defects, *The Journal of Engineering*, vol.2018, no.16, pp.1612-1616, 2018.

[4] B. Su et al., Deep learning-based solar-cell manufacturing defect detection with complementary attention network, *IEEE Transactions on Industrial Informatics*, vol.17, no.6, pp.4084-4095, 2020.

[5] M. Mundt et al., Meta-learning convolutional neural architectures for multi-target concrete defect classification with the concrete defect bridge image dataset, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[6] X. Wei et al., Defect detection of pantograph slide based on deep learning and image processing technology, *IEEE Transactions on Intelligent Transportation Systems*, vol.21, no.3, pp.947-958, 2019.

[7] S. Albawi, T. A. Mohammed and S. Al-Zawi, Understanding of a convolutional neural network, *2017 International Conference on Engineering and Technology (ICET)*, 2017.

[8] S. Yun et al., Re-labeling imagenet: From single to multi-labels, from global to localized labels, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[9] H. Ramchoun et al., Multilayer perceptron: Architecture optimization and training, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol.4, no.1, pp.26-30, 2016.

[10] Y. Kim et al., Structured attention networks, *arXiv Preprint*, arXiv: 1702.00887, 2017.

[11] A. Vaswani et al., Attention is all you need, *Proc. of the 31st International Conference on Neural Information Processing Systems*, pp.6000-6010, 2017.

[12] H. Wang et al., Soft-voting clustering ensemble, in *Multiple Classifier Systems. MCS 2013. Lecture Notes in Computer Science*, Z. H. Zhou, F. Roli and J. Kittler (eds.), Berlin, Heidelberg, Springer, 2013.

[13] A. M. Reza, Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement, *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol.38, no.1, pp.35-44, 2004.

[14] W. S. Noble, What is a support vector machine?, *Nature Biotechnology*, vol.24, no.12, pp.1565-1567, 2006.

[15] T. Chen et al., *XGBoost: Extreme Gradient Boosting*, R Package Version 0.4-2 1.4 1-4., 2015.