Performance and Visual Explainability of Chromagram-based CNN-ANN as Acoustic Fault Classifier of Industrial Equipment

Lagazo, D.
Department of Information Systems and Computer Science
Ateneo de Manila University
Manila, Philippines
dlagazo@gmail.com

de Vera, J.
Department of Information Systems and Computer Science
Ateneo de Manila University
Manila, Philippines
jdevera@ateneo.edu

Coronel, A.
Department of Information Systems and Computer Science
Ateneo de Manila University
Manila, Philippines
acoronel@ateneo.edu

Jimenez, J.
Department of Information Systems and Computer Science
Ateneo de Manila University
Manila, Philippines

Gatmaitan, E.
Department of Information Systems and Computer Science
Ateneo de Manila University
Manila, Philippines

2021 IEEE Western New York Image and Signal Processing Workshop (WNYISPW) | 978-1-6654-3630-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/WNYISPW53194.2021.9661280

***Abstract*—The early and accurate diagnosis and prognosis of the defect of industrial equipment is crucial for predictive and prescriptive maintenance. Fault detection in airborne signals using acoustic sensors is a relatively new initiative in the field of predictive maintenance. Anomalies are detected in the sound using the reconstruction error of a CNN Autoencoder (CNN-AE) prior to downtime incidents. This has been demonstrated in prior research and is used as the foundation of this paper. By using the anomalies, downtime incidents can be predicted and maintenance work can be done proactively. The anomalies detected are extracted, visualized and used in building the fault classifier. The fault classifier is intended to help speed up the maintenance work by recommending the type of fault based on a probability distribution. In this paper, we discuss the design, performance, and visual explanation of the fault classification model. Visualization techniques, namely: intermediate activations, and gradient-weighted class activation mapping (Grad-CAM) are used to explain how the model classifies the input data.**

Keywords—fault classification, chromagram, industrial equipment, prognosis, diagnosis, sound pressure, convolutional neural network, intermediate activations, model explainability, class activation mapping

I. Introduction

The detection of anomalies prior to equipment breakdown incidents has been demonstrated using the reconstruction error of autoencoder[1]. The autoencoder is trained using the chromagram features of normal equipment sound. This paper builds a classification network on top of the anomaly detection network in prior research. After the detection of anomalies, diagnosis and classification of the fault using the anomalies is performed. Chromagram time-frequency representation is widely used in music processing applications that involves tonality of music, such as key or chord identification[2-5] and speech/music classification[6]. The use of autoencoders and time-frequency features of acoustic data has been demonstrated for leak detection[7], machine sound monitoring[8] and industrial anomalies[9]. Moreover, companies such as OneWatt, 3d Signals, and OtoSense by Analog Devices Inc., use acoustic sensing as one of their domains for predictive maintenance.

CNN-based classification of fault in industrial equipment[10] and food packaging[11] have been demonstrated to have high accuracy. The advantage of using deep learning is the absence of tedious hand-crafted features, and lack of need for domain expertise. Therefore, the use of raw time-series sensor data is possible for fault classification[12].

Deep Neural Networks are considered as black boxes due to their multi-layer non-linear structure and are often criticized to be non-transparent because it is hard to trace their predictions[13]. Gradient-weighted Class Activation Mapping (Grad-CAM) is a novel class-discriminative localization technique. It uses class-specific gradient information to localize important regions[14]. This technique is used to make CNN-based models more transparent by visualizing input regions that are important for predictions. Intermediate activations have been used to explain the automated optical inspection of welding using CNN[15,16].

II. Methodology

A. Chromagrams

Airborne signals such as sound can be represented as Chromagrams. The chroma vector is a 12-element(24 in our configuration) representation of short-time energy distribution of a music signal[19,20]. The concatenation of the chroma vectors across time is known as a chromagram[6]. Each of our chromagram samples (Fig. 1) has a time frame of one second.

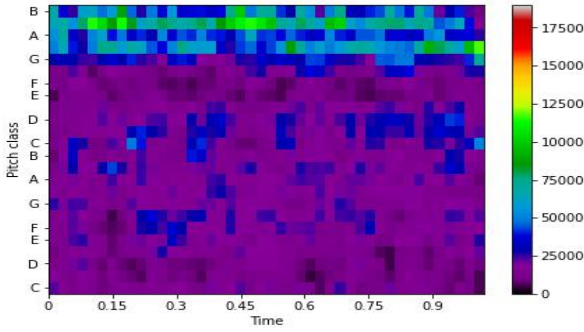

Fig. 1. Chromagram sample from the fault type "multiple defects" data set.

B. *Anomaly Detection and Extraction Using Reconstruction Error of Autoencoder*

In prior research, a CNN-Autoencoder (CNN-AE) was trained using only normal/healthy chromagrams[1]. This is intended to minimize the reconstruction error of the output image with respect to the input image. Reconstruction error is provided in (1).

$$\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \tag{1}$$

where x is the original input image array,

where y is the reconstructed output image array,

where i is the index of the image array

and n is the index array size

The effect of reducing the reconstruction error does not apply to abnormal/unhealthy chromagrams that are not included in the training of the CNN-AE. Fig. 2-b shows the output of the CNN-AE for normal or healthy equipment chromagram (Fig. 2-a) and Fig. 2-d shows the output of abnormal or unhealthy chromagrams (Fig. 2-c).

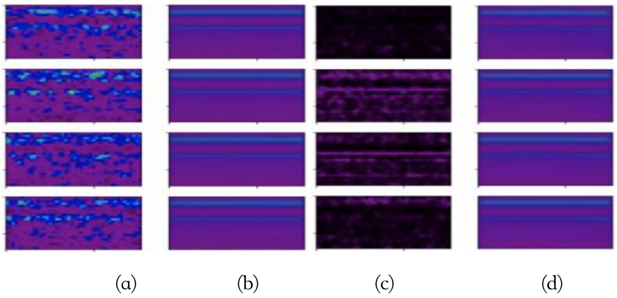
(a) (b) (c) (d)

Fig. 2. A) Four samples from the training data set. B) The reconstructed images using column A images as input. C) Column C illustrates four samples from the breakdown/anomaly data set. D) Column D illustrates the reconstructed images from column C.

A reconstruction error threshold is used to separate the healthy and unhealthy chromagrams. Using the threshold, a total of 30,771 chromagrams were flagged as anomalies. These anomalies were detected prior to downtime incidents. In the case of "wirebreak" incidents, anomalies were detected 2-4 hours before the incident. In the case of a "multiple defects" incident, the anomalies were detected starting at two days before the incident.

Table I shows the total amount of anomalies that were detected using prior research[1]. These anomalies are labeled according to the incident that was reported and are used to prepare the training and validation data sets for the fault classifier.

Fig. 3 illustrates the latent space representation of the CNN-AE that is used for anomaly detection. The blue data points represent the normal/healthy data. The red data points represent the downtimes (when the equipment is turned off for repairs or being restarted). Orange (January) and black (April) data points represent the unlabeled data. The yellow (multiple defects) and green (wirebreak) data points represent the anomalies that were detected prior to the downtime incidents. The anomalies are labeled according to the defects that were diagnosed during the downtime incidents. A third type of defect (brown) is artificially added to the normal (blue) data points. The sound of defective fan from the MIMII data set[18] is added to the normal equipment sound. Chromagrams of the fault type "defective fan" are represented by the brown data points.

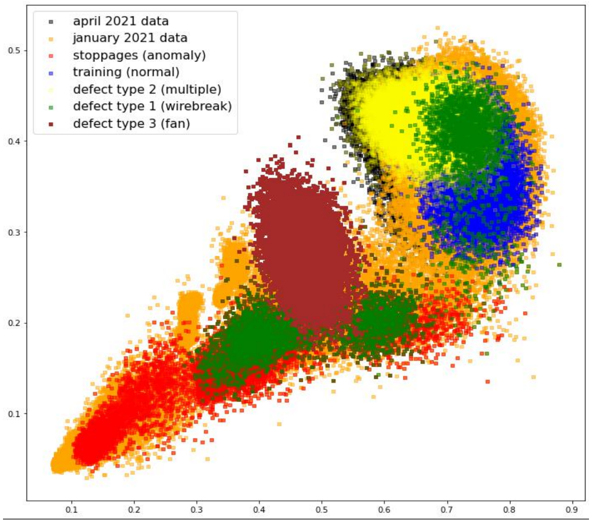

Fig. 3. The autoencoder's latent space representation of all the data set

C. *Dataset Preparation*

TABLE I. TRAINING AND VALIDATION DATA SETS

Class/Fault Type	Training Samples	Validation Samples
Defective Fan (artificial anomaly)	3,793	10,607
Multiple Defects	24,083	1,384
Wirebreak	4,065	1,239

Table I shows the breakdown of the samples for the training and validation data sets. The training samples were incremented to improve the performance of the final version of the model. The figures in Table I are the final values after the adjustments. The fault type "defective fan" was able to reach 100% accuracy and f1-score using a low number of training samples. For fault types "multiple defects" and "wirebreaks", it is necessary to increase the training samples to reach an acceptable accuracy and f1-score. As a disclosure, the data set is not publicly available and a property of the owner of the industrial equipment.

D. *Anomaly Detection - Fault Classification Architecture*

Using the architecture and results of a prior research[1], the CNN-AE anomaly detection serves as the initial system. Its task is to perform binary classification of the input. Inputs that are classified as anomalies are analyzed by the fault classifier. The architecture of the fault classifier involved several changes in its hyper-parameters to achieve the acceptable performance. To reduce the cost in training, the first version used the pre-trained encoder of the autoencoder with an added classifier layer. The encoder layers were frozen and only the classifier was trained. The second version involved the autoencoder with all of its layers set as trainable. The final version of the fault classifier is in Fig. 4. This version achieved the best performance and the highest cost in terms of training time.

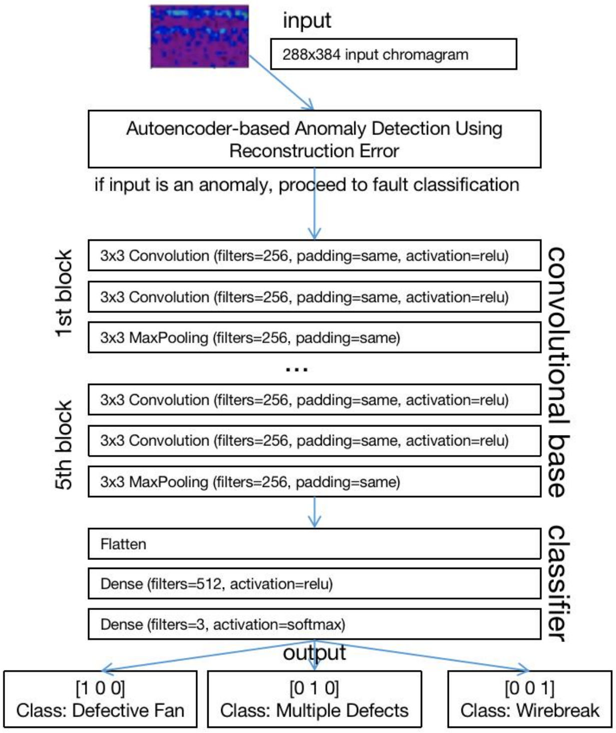
input
288x384 input chromagram
Autoencoder-based Anomaly Detection Using Reconstruction Error
if input is an anomaly, proceed to fault classification
1st block
3x3 Convolution (filters=256, padding=same, activation=relu)
3x3 Convolution (filters=256, padding=same, activation=relu)
3x3 MaxPooling (filters=256, padding=same)
...
5th block
3x3 Convolution (filters=256, padding=same, activation=relu)
3x3 Convolution (filters=256, padding=same, activation=relu)
3x3 MaxPooling (filters=256, padding=same)
convolutional base
Flatten
Dense (filters=512, activation=relu)
Dense (filters=3, activation=softmax)
classifier
output
[1 0 0]
Class: Defective Fan
[0 1 0]
Class: Multiple Defects
[0 0 1]
Class: Wirebreak

Fig. 4. The architecture of the fault classification has five blocks of convolutional and maxpooling layers. Each block has the same parameters. The classifier has a final dense layer of 3 filters that represent each of the fault types. The autoencoder-based anomaly detection is part of a prior research.

E. *Model Explainability: Visualizing intermediate activations*

To visualize the output of the layers, the output of each filter in the first and final convolutional layers are presented in the results. Chollet[17] demonstrated this technique using the model "cats vs dogs" in the keras library.

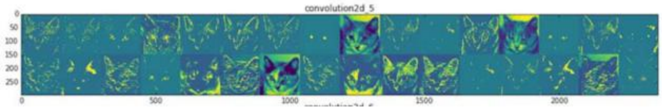

Fig. 5. Activations from 32 filters of the first convolutional layer using an image of a cat as input.

The first convolutional layer acts as a collection of various edge detectors and the activations (output of each filter) retain almost all of the information present in the input cat image.

The activations of the higher convolutional layers (Fig. 6) become increasingly abstract and less visually interpretable. Chollet[17] explains that the higher representations carry increasingly less information about the visual contents of the image, and increasingly more information related to the class of the image. Yang et. al[15] observed that their optical CNN-based inspector's first convolutional layer is a collection of various edge detectors. Their final convolutional layer are also less interpretable and sparse. Intermediate activations are also used to explain defects in welding[15,16].

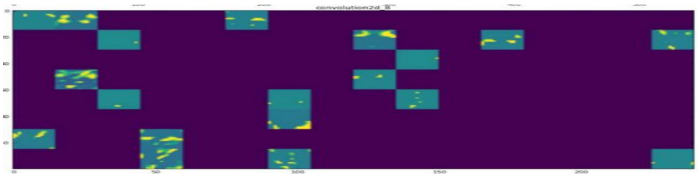

Fig. 6. Activations from 128 filters of the final convolutional layer using an image of a cat as input.

F. *Model Explainability: Gradient-weighted Class Activation Map*

Because Class Activation Mapping (CAM) cannot be used in networks that have multiple fully-connected (dense) layers before the output, Grad-CAM is used [14]. Chollet[17] demonstrated the use of the VGG16 model, initialized with imagenet weights, in generating a Grad-CAM heatmap (Fig. 7-a) for the model output #386 which has the gradients for the class "African elephant". The discussion on the mathematical[14] and algorithmic[17] implementation of Grad-CAM is not included in this paper.

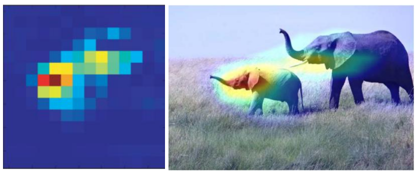

Fig. 7. Gradient-weighted class activation map of the class "African elephant" from the VGG16 imagenet model (left). Superimposed heat map over an input image of elephants (right).

After using opencv to superimpose the heat map onto the input image, the resulting image indicates the importance of each pixel in the input image for the activation of the class "African elephant". In this example, the head and ears of the elephant calf (red) are strongly activated while the edges of the head of the calf (yellow) are the second most important activation.

In the context of the input images, chromagrams, the Grad-CAM will be used to determine which pitch classes (Fig. 12) are strongly activated or have the most importance in activating a specific fault class.

III. Results

A. *Performance of Fault Classifier*

TABLE II. Classifier Performance on the Validation Data Set

Class/Fault Type	Precision	Recall	F1-score	Support
Defective Fan (artificial anomaly)	100%	100%	100%	10,607
Multiple Defects	94%	93%	94%	1,384
Wirebreak	90%	93%	92%	1,239

Table II shows the precision, recall, f1-score and support (number of validation samples). Due to the class imbalance due to the lower validation samples for "multiple defects" and "wirebreak", the F1-score is the recommended metric.

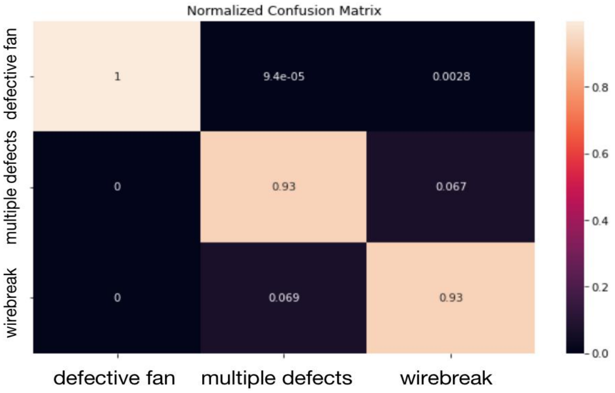

Fig. 8. Normalized confusion matrix visualized as a heat map of accuracy.

Fig. 8 shows the accuracy of the model in the form of a heap map. Due to the class imbalance, F1-score (Table II) is the recommended metric instead of accuracy (Fig. 8).

B. *Model Output per layer*

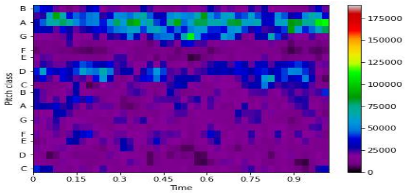

Fig. 9. Sample of fault type "wirebreak

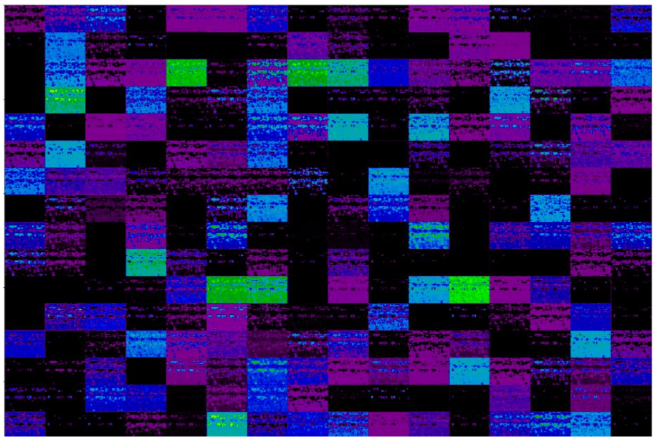

Fig. 10. Intermediate activations of the first convolutional layer.

To visualize the result of the filters in the bottom and top layers, a chromagram sample from the fault type "wirebreak" (Fig. 9) is evaluated by the model. The output of the first 256 filters in the first convolutional layer of the first block is shown in Fig. 10. The intermediate activation (Fig. 9-B) shows a similar behavior to Chollet[17] and Zang[16], with the layer being a collection of edge detectors that isolate the higher energy frequencies (pitch classes B, A, D). The intermediate activations of the final convolutional layer of the fifth block is shown in Fig. 11. The activations that represent the higher energy frequencies are no longer identifiable.

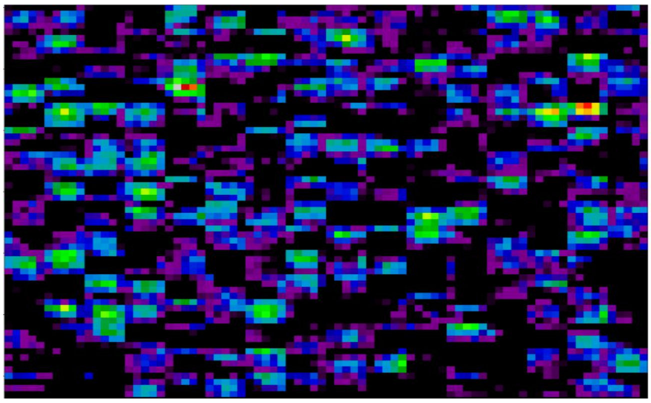

Fig. 11. Activations from 256 filters of the final convolutional layer (5th block).

C. *Results of Gradient-weighted Class Activation Map*

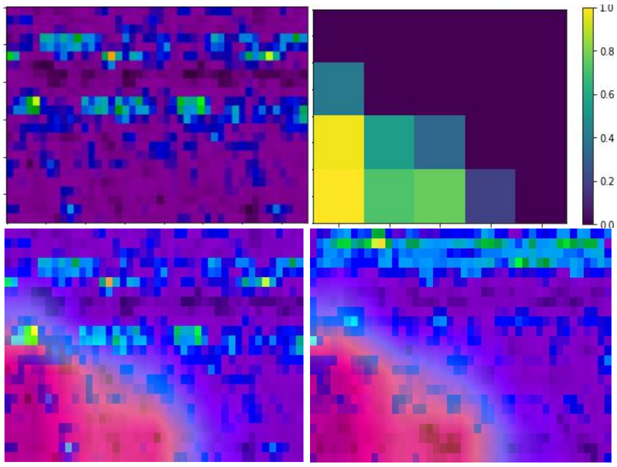

Fig. 12. A) A sample chromagram from the fault type "wirebreak" (top-left). B) The Grad-CAM heat map generated for the fault type "wirebreak" (top-right). C) The resulting image by superimposing the heat map onto the sample chromagram in 9-A (bottom-left). D) The resulting image by superimposing the heat map onto a sample chromagram of fault type "multiple defects".

Using the activations from the final convolutional layer (Fig. 11), a Grad-CAM heat map (Fig. 12-B) is generated for fault type "wirebreak". The Grad-CAM for fault type "multi" and "fan" are invariant to Fig. 12-B. Based on the heat map, the bottom 12 bins have the highest importance. Using opencv to add the heat map (at 0.4 alpha or transparency) to the sample chromagram, the resulting image is illustrated in Fig.12-C. The same steps are repeated using a sample chromagram of fault type "multi" and the resulting image is in Fig. 12-D.

ACKNOWLEDGMENT

This work is primarily funded by the PHILIPPINE COUNCIL FOR INDUSTRY, ENERGY AND EMERGING TECHNOLOGY RESEARCH AND DEVELOPMENT (DOST-PCIEERD) with Project No. 08241, 2020.

The authors acknowledge the support of Associated Wire Corporation of the Philippines, led by Mr. Jason Jose, Ms. Cecille Sayos and Mr. Joel Lastimosa.

The authors would like to thank Ateneo de Manila University's School of Science and Engineering, headed by Dr. Evangeline Bautista, for their support.

The authors would like to thank Dr. Jimson Ngeo for providing guidance and insights during the early stages of the research project. The authors would like to thank Mr. Aames Aguas of ACC Technology Consulting Inc. for his contributions in the proposal stage of the research.

REFERENCES

[1] D. Lagazo, J. de Vera, A. Coronel, J. Jimenez and E. Gatmaitan, "Condition-Based Monitoring and Anomaly Detection of Industrial Equipment using Autoencoder," *2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST)*, 2021, pp. 146-151, doi: 10.1109/ICAICST53116.2021.9497816.

[2] M.A. Bartsch, G.H. Wakefield, "Audio thumbnailing of popular music using chroma-based representations", *IEEE Trans Multimed,* 2005

[3] M. Mueller, F. Kurth, M. Clausen, "Audio matching via chroma-based statistical features", *Proceedings of the 6th international conference on music information retrieval (ISMIR),* 2005, https://doi.org/10.5281/zenodo.1416800

[4] P. Dighe, P. Agarwal, H. Karnick, S. Thota, B. Raj, "Scale independent raga identification using chromagram patterns and swara based features", *IEEE international conference on multimedia and expo,* 2013

[5] Y. Lee, Y. Chiang, P. Lin, C. Lin, T. Tai, "Robust and efficient content-based music retrieval system", *APSIPA Trans Signal Inf Process,* 2016, https://doi.org/10.1017/ATSIP.2016.4

[6] G. Birajdar, M. Patil, "Speech/music classification using visual and spectral chromagram features", *Journal of Ambient Intelligence and Humanized Computing,* 2019, https://doi.org/10.1007/s12652-019-01303-4

[7] R. Cody, B. Tolson, J. Orchard, "Detecting Leaks in Water Distribution Pipes Using a Deep Autoencoder and Hydroacoustic Spectrograms", *Journal of Computing in Civil Engineering,* 2020, doi: 10.1061/(ASCE)CP.1943-5487.0000881

[8] D.Y. Oh, I.D. Yun, "Residual Error Based Anomaly Detection Using Auto-Encoder in SMD Machine Sound", *Sensors,* 2018, https://doi.org/10..3390/s18051308

[9] B. Bayram, T.B. Duman, G. Ince, "Real time detection of acoustic anomalies in industrial processes using sequential autoencoders", *Expert Systems,* 2021,

[10] S. Li, H. Wang, L. Song, P. Wang, L. Cui, T. Lin, "An Adaptive Data Fusion Strategy for Fault Diagnosis based on the Convolutional Neural Network", *Measurement,* 2020, https://doi.org/10.1016/j.measurement.2020.108122

[11] L. Medus, M. Saban, J. Francés-Víllora, M. Bataller-Mompeán, A. Rosado-Muñoz, "Hyperspectral image classification using CNN: Application to industrial food packaging", *Food Control Vol. 125,* 2021, https://doi.org/10.1016/j.foodcont.2021.107962.

[12] P. Rai, N. Londhe, R. Raj, "Fault classification in power system distribution network integrated with distributed generators using CNN", *Electric Power Systems Research Vol. 192,* 2021,

[13] V. Buhrmester, D. Muench, M. Arens, "Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey", arXiv preprint arXiv:1911.12116

[14] R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, "Grad-CAM: Why did you say that?", *NIPS 2016 Workshop on Interpretable Machine Learning in Complex Systems,* arXiv: 1611.07450

[15] Y. Yang, L. Pan, J. Ma, R. Yang, Y. Zhu, Y. Yang, L. Zhang, "A High-Performance Deep Learning Algorithm for the Automated Optical Inspection of Laser Welding" *Applied Sciences 10,* 2020, https://doi.org/10.3390/app10030933

[16] Z. Zhang, G. Wen, S. Chen, "Weld image deep learning-based on-line defects detection using convolutional neural networks for A1 alloy in robotic arc welding", *Journal of Manufacturing Processes 45,* 2019, https://doi.org/10.1016/j.jmapro.2019.06.023

[17] F. Chollet, "Deep Learning with Python", 1st edition, Manning Publications, 2017, pp. 162-176

[18] H. Purohit, R. Tanabe, K. Ichige, T. Endo, Y. Nikaido, K. Suefusa, Y. Kawaguchi, "MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection", *4th Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE), 2019,* arXiv:1909.09347

[19] T. Fujishima, "Real-time chord recognition of musical sound: a system using common lisp music", *International computer music conference,* 1999

[20] G. Wakefield, "Mathematical representation of joint-time chroma distributions", *SPIE,* https://doi.org/10.1117/12.367679