Data Transformation to Enhance Explainability of Abnormality Diagnosis for Nuclear Power Plants

Ji Hyeon Shin, Jae Min Kim, and Seung Jun Lee*
Department of Nuclear Engineering, Ulsan National Institute of Science and Technology,
50, UNIST-gil, Ulsan 44919, Republic of Korea
**Corresponding author: sjlee420@unist.ac.kr*

1. Introduction

Due to the aging of nuclear power plants (NPPs), the possibility of abnormal situation occurrences is increasing. Numerous indicators and alarm systems exist in the NPP main control room. When an abnormal situation occurs at an NPP, the operator must perform the tasks described in advance by making a diagnosis as an appropriate thing among hundreds of abnormal situations within a given time based on such information on the NPP. For this, the operator tasks are accompanied by a high workload. Such task difficulty can increase the human error possibility. Recently, earlier research has been conducted to solve these safety problems with a diagnostic support system using artificial intelligence (AI) technology.

However, it is difficult to directly apply the diagnostic results of the AI model corresponding to the black-box to support the abnormal state diagnosis in NPPs. Recently, explanation techniques to validate these AI models have been studied. Operators can be provided support together with model diagnosis results and cause through the explainable AI. This increases the reliability of the operator support system and reduces operator confusion, increasing the applicability of AI technology to the actual nuclear industry. In this study, we tried the mapping of monitoring parameters to expect a better explanation of the model. The 3-channel convolutional neural network (CNN) trained on the imaged datasets was explained by Guided backpropagation + Gradient-weighted Class Activation Mapping (Guided Grad-CAM). Through this, it was confirmed that the model can provide a suitable diagnostic cause. In addition, we can expect to enhance operator understanding by post-processing the explanation results.

2. Methodology

The proposed experiment is performed as shown in the figure below. Parameters in the produced data set are arranged in the form of a nuclear power plant map. These arrayed image data sets maximize the advantage of the CNN model and model explanation techniques.

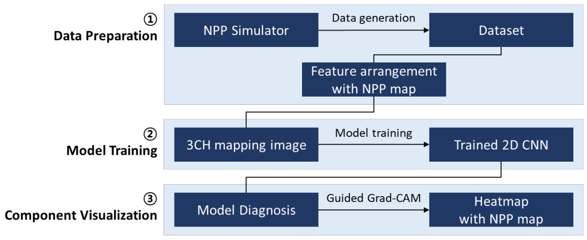

Fig. 1. Experimental Setup

2.1 Data Production

We used a 3KEYMASTER Simulator based on a 1400MWe Generic Pressurizer Water Reactor from Western Service Corporation (WSC) to produce the experimental data file [1].

15 abnormal situations as shown in Table I below were simulated in this simulator by injecting the malfunction. For each abnormal situation, it was produced 48 training data files to varying malfunction degrees. (720 data files in total) Each training data file contains monitoring parameter information while the 60s.

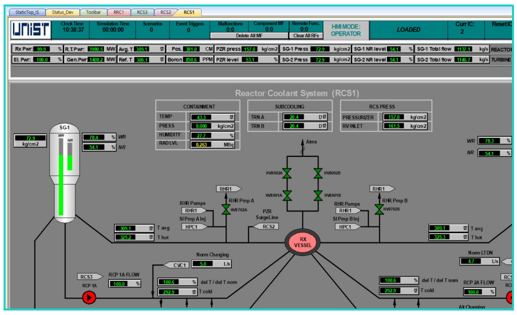

Fig. 2. WSC 3KEYMASTER simulator

Table I: Kinds of Abnormal Events for Datasets

State	Abnormal event
SGTL	Steam generator tube leakage
CHRG	Charging line break upstream
LTDN	Letdown line leakage inside the containment
CDS	Loss of condenser vacuum
POSRV	HV456A valve leakage
CWS	Circulating water tube leakage in LP condenser
MSIV	Main steam isolation valve abnormality
RCP	HV8351A valve abnormality

MSS	Main steam header steam leakage
PZR	PV455B valve leakage
CCW	CCW service loop header leakage
LFH	Feedwater heater 1A tube break
HFH	Feedwater heater 5A tube break
MFW	FV1B valve leakage
TCS	CV1 valve abnormality

2.2 Multi-channel Convolutional Neural Network

The earlier research have used 2-channel CNN to classify abnormal events [2]. As shown in Fig. 1 below, a 2-channel CNN uses a two-channel structure as input. The first channel has current information about parameter values, and the second channel has information about the amount of parameter change. This model has showed high accuracy compared to recurrent neural networks or 1-channel convolutional neural networks.

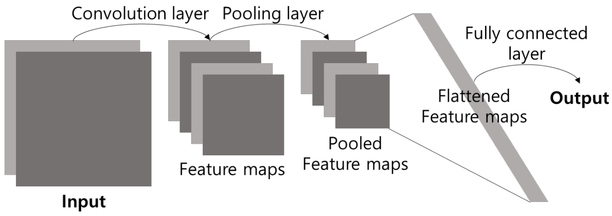

Fig. 3. 2-CH CNN Structure

In this study, a 3-channel CNN was used by reflecting the previous research model. Each model has a total of 3 channels: current information, change information from 10 seconds ago and change information from 15 seconds ago.

2.3 Pre-processing with Data Structure

Unlike data structure with the simple stacking and square shapes performed in the earlier study mentioned, we try to take advantage of the feature that the CNN trains information from neighboring pixels for a position through filters (3*3). In this study, pixels representing 332 monitoring parameters in datasets are arranged in the form of the NPP. Each parameter is placed at 20 system locations. Some parameters are used redundantly to express the shape of the system. In addition, it was considered the pump and line locations to arrange. Input image with mapped parameter looks like below Fig. 4.

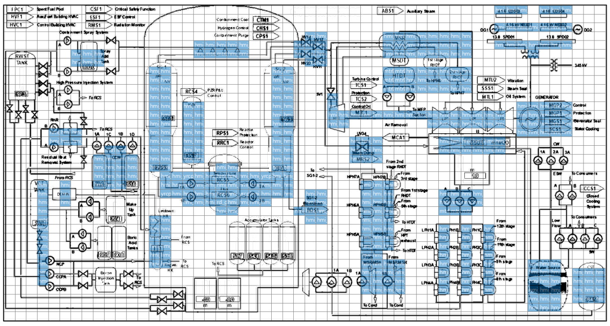

Fig. 4. Monitoring Parameters on the NPP Map

2.4 Guided Grad-CAM

Grad-CAM, a representative model analysis technique based on CNN, was used [3]. The equation of the Grad-CAM technique is as follows.

$$L^{c}_{Grad-CAM} = ReLU\left(\sum_{k} \alpha^{c}_{k} A^{k}\right)$$

Fig. 5. Equation for Grad-CAM

Guided Grad-CAM can be expressed by multiplying the existing Grad-CAM with the Guided Backpropagation result as shown in Fig. 6 below. This shows a higher resolution than the existing technique.

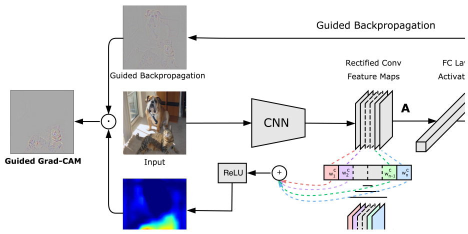

Fig. 6. Guided Grad-CAM [3]

2.5 Post-processing with Explained Results

It is necessary to consider how to provide diagnostic explanation results that change every second. In addition, we tried to provide the operator with a visualization of the parameter change score as it was judged to be important as a diagnostic cause. In this regard, the following two criteria were established.

(1) *(Relevance score · |Change score while 15s|)*^{2}
(2) *Normalized relevance score > 0.5*

3. Results

3.1 Model Training

The model was trained with very high performance as following Table II. Also, the model accurately diagnosed all 15 test data files given for each scenario.

Table II: Training Result

Training dataset		Validation dataset	
Accuracy	Loss	Accuracy	Loss
1.0000	2.1711e-6	1.0000	7.6496e-6

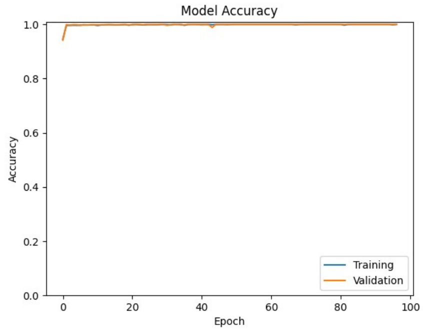

Fig. 7. Learning Curve with Accuracy

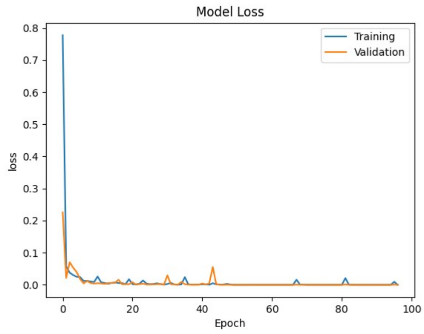

Fig. 8. Learning Curve with Loss

3.2 Comparison with Post-processing

We explain when the model diagnoses the test scenario cases. The explanation result is immediately shown in Fig. 9-12 for examples. It was visualized so that the cause of the diagnosis could be known.

Fig. 9 and Fig. 10 below are visualization results by averaging the explained relevance score for 60 seconds immediately after the abnormal diagnosis and 20 seconds immediately after the model diagnosis to the LFH event. It was confirmed that it is possible to express even the cause component by visualizing the explained relevance score in a relatively short time. In Fig. 10, the direct causative component, near the feedwater heater 1A tube, was visualized with a high relevance score.

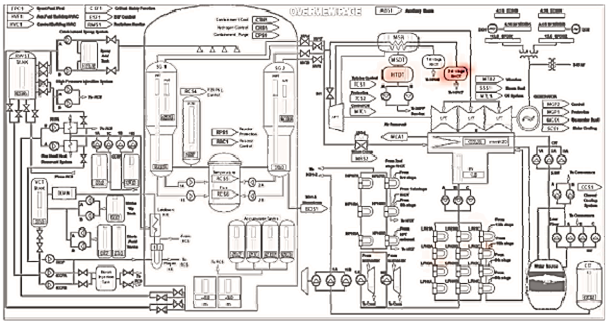

Fig. 9. Explanation Result while 60s at LFH Event

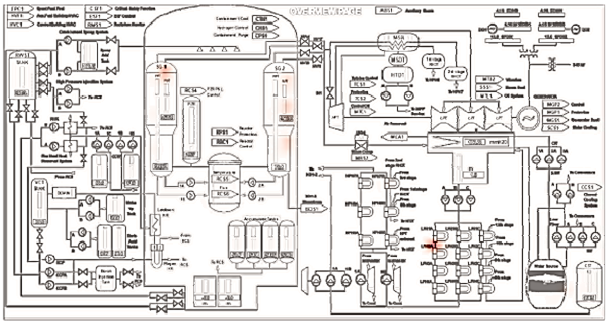

Fig. 10. Explanation Result while 20s at LFH Event

Fig. 11 and Fig. 12 below respectively show the visualization results according to whether parameter change information is reflected in the explanation results when the model diagnoses MSIV. In Fig. 11, the six causative systems are visualized. As a result of reflecting the parameter change information and transforming with 0.5 thresholds in Fig. 12, only four causative systems were derived, and a neat visualization image can be confirmed. However, it can be seen that the score near the main steam isolation valve, which is the main cause, is lowered.

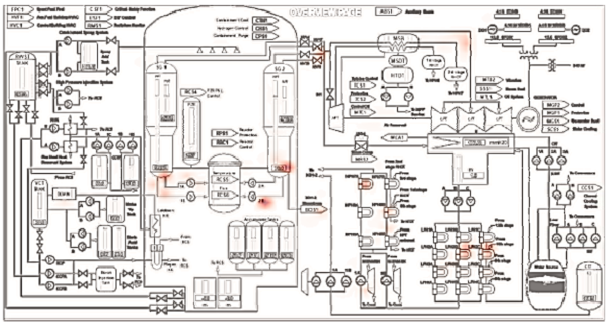

Fig. 11. Explanation Result at MSIV Event

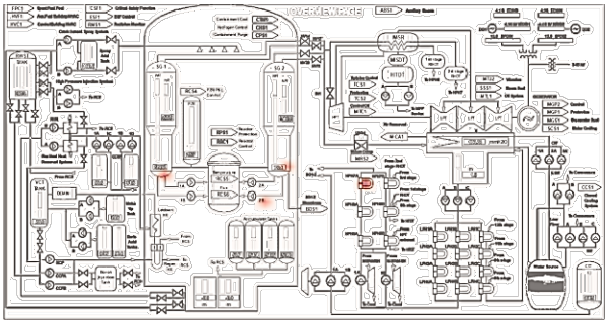

Fig. 12. Explanation Result with Change at MSIV Event

4. Conclusions

To increase the operator task efficiency in the abnormal situation of an NPP, we proposed to support the operator using explainable AI. The approach was able to provide diagnostic results along with the immediately visualized diagnostic causes by explaining the CNN model trained using parameters arranged to the NPP map through Guided Grad-CAM. By mapping parameters to the NPP map, the preprocessed image trained the model with high learning performance and was advantageous in terms of visualization. In addition, it was possible to provide the diagnostic cause at an understandable level to the operator through appropriate processing of the explanation result. In future studies, it is required to validate that the visualized causative system to become practical support information for operators to diagnose abnormal events in NPPs.

ACKNOWLEDGEMENT

This work was supported by Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MOTIE) (No. 20211510100020, Development of operation supporting technology based on artificial intelligence for nuclear power plant start up and shutdown operation).

REFERENCES

[1] 3KEYMASTER Simulator 2013, Western Service Corporation, Frederick, MD, USA.
[2] Lee, Gyumin, Seung Jun Lee, and Changyong Lee. "A convolutional neural network model for abnormality diagnosis in a nuclear power plant." Applied Soft Computing 99 (2021): 106874.
[3] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-cam: Visual explanations from deep networks via gradient-based localization." In Proceedings of the IEEE international conference on computer vision, pp. 618-626. 2017.