Reliable Abnormality Diagnosis Model for Nuclear Power Plant Using Convolutional Neural Networks

Ji Hyeon Shin, Seung Jun Lee*
* *Ulsan National Institute of Science and Technology, 50, UNIST-gil, Ulsan, 44911*
**Corresponding author: sjlee420@unist.ac.kr*

1. Introduction

A nuclear power plant (NPP) is a large-scale power generation system consisting of thousands of components. In an NPP, the exacerbation of an abnormal situation caused by a failure of a specific component or function can make a serious situation [1]. This situation can cause enormous property damage due to interruptions in power generation. In addition, it can cause environmental damage due to radioactive leakage in an NPP. Therefore, an NPP belongs to a critical safety system that must prevent accidents. An operator in an NPP quickly recognizes the problem and take action to mitigate this situation. However, the information on plant parameters that the operator must accept to judge the situation is vast. It is required a heavy workload to accept and accurately diagnose the kinds of more than 200 abnormal events in a short time.

Recently, various artificial intelligence technologies are being applied to solve complex problems. In prior researches, several deep learning models have been developed to support an operator in diagnosing abnormal events. However, when the deep learning model is applied to the final event diagnosis of an NPP which is a critical safety system, the operator is dependent on this diagnostic system. In this case, the operator may not notice a failure of the model or his ability to recognize a situation may decrease [2-4].

To solve this problem, in this study, the contribution of each parameter to the diagnosis of 10 NPP states in the convolutional neural network (CNN) model is calculated by two interpretation methods. Through model training, we validate how useful the parameters identified by two methods are for classifying the NPP state by the model. Finally, we propose to provide the operator with parameters that are the basis for judgment, that is, the process of diagnosing the state of the CNN model.

2. Methods

In this study, the CNN model which can classify 10 labels corresponding to each 10 NPP status with high accuracy was used. One test data for each label that the trained CNN model can classify correctly is given. When the CNN model classified the given test data, which features have high relevance or high contribution was calculated by two interpretation methods. It was finally validated that the model can learn 10 labels with only the parameters for each label selected through the calculated heatmap. The overall experimental algorithm for this study is as following figure 1.

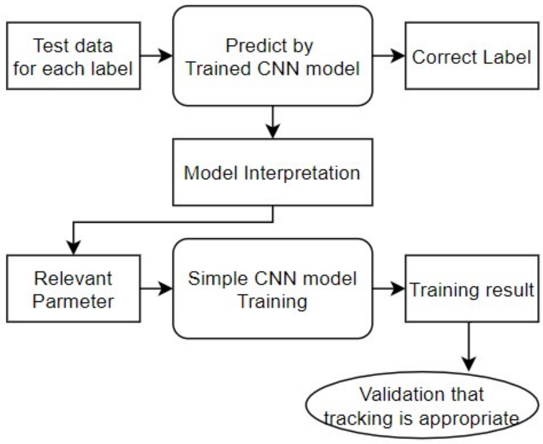

Fig. 1. Overall experimental algorithm

2.1. Base CNN Model Training

A CNN is a neural network that classifies images using patterns. Several convolutional layers are added in front of the existing dense layer. This convolutional layer extracts features about two-dimensional input. The extracted features are converted into one-dimensional data in a flatten layer. Then it is used for classification in a dense layer. [5]

In this study, a CNN model was used to diagnose abnormal conditions of an NPP. The training model hyperparameter is as follows.

- Number of convolutional layer : 3
- Kernel size of each convolutional layer : 3*3
- Filter number of each convolutional layer : 16
- Activation function of each convolutional layer: ReLU
- Activation function of Fully connected layer : softmax
- Optimizer : Adam (learning rate=0.01)

The labels of data used for training are 1 normal state and 9 abnormal states of an NPP: steam generator tube leakage (SGTL), charging line break (CHRG), letdown line leakage (LTDN), loss of condenser vacuum (CDS), circulating water tube leakage (CWS), reactor coolant pump abnormality (RCP), main steam line break (MSS), low-pressure feedwater heater abnormality (LFW), high-pressure feedwater heater abnormality (HFH). They are produced by 3KEYMASTER NPP simulator of Western Services Corporation, 50 data for each label [6]. One data is expressed as two-dimensional input data of 944 human machine interface parameters for 60 seconds. Figure 2 is the result of the CNN model trained for 100 epochs. The model classified 10 labels with a high performance of 99.33 % on the validation dataset.

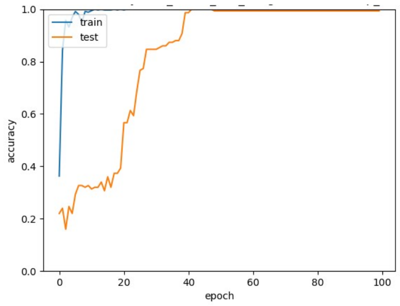

Fig. 2. Based CNN model accuracy

2.2. Model Interpretation Methods

When the above based CNN model classifies test data for 10 labels, relevance of each feature in the test data was calculated using the following two model interpretation methods.

- Saliency Map [7]

Saliency map explains the relevance of each feature by calculating the gradient of output to the input as following equation 1.

$$Saliency(x) = \nabla_x f(x) = \frac{\partial f(x)}{\partial x} = \frac{\partial output}{\partial input} \quad (1)$$

Gradient values are used to express how a small difference in the input feature changes the output. In other words, the Saliency(x) with the input feature x, which has the most influence of the output, has the highest value.

- Guided Gradient-weighted Class Activation Mapping (Guided Grad-CAM) [8]

Filters included in the last convolutional layer of the trained CNN model have information on the main feature map for classifying labels. Grad-CAM calculates the contribution of a feature using the weights given to filters in this convolutional layer. This calculated contribution can be considered the kind of label relevance in this study. Equation 2 below represents the importance of A, the feature map k for class c. As a result, Grad-CAM is calculated as following equation 3.

$$\alpha_k^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (2)$$

$$L^c_{Grad-CAM} = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right) \quad (3)$$

By multiplying this result and the result of Guided backpropagation, the resolution of a heatmap calculated by Grad-CAM can be increased.

3. Results

3.1. Relevant Parameters

Figure 3 below shows the heatmap expressing the relevance of each feature when the model classifies test data for the CDS label. They were calculated by Saliency map and Guided Grad-CAM.

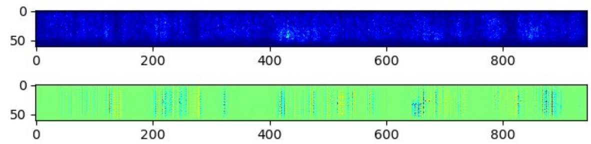

Fig. 3. Example of heatmap

As result, the parameter corresponding to the column with the highest mean value for each contributed feature value is expected to contribute the most for the CNN model to classifying sample data as CDS label. In this way, 10 sample data are classified into each label, and the most relevant parameter is tracked as shown in the following table 1.

Table. I: Relevant parameter name for each label

Label	Model explanation methods	
	Saliency map	Guided Grad-CAM
Normal	CONFV191	CVCTT216
SGTL	MFWTT129A	SPDIT4
CHRG	CVCFT142A	CVCLT149
LTDN	NBXIT6	CVCFT121
CDS	MGPTT17	TCS_15_B
CWS	SPDIT4	T701
RCP	RHRTT613	BDSTT73
MSS	CHSRT312	FPCFT80
LFH	CONPT25	T701
HFH	MFWST33	CCWLT2

3.2. Relevance Validation

Considering the overlapping parameter among the 10 calculated parameters using each explanation method in section 3.1, Saliency map selected a total of 10 and Guided Grad-CAM selected a total of 9 parameters. Table 2 below shows results training simple CNN models according to the kind of parameters used for training data.

Table. II: Training results based on the training data

Training data		Training result			
Param. Type*	Param. Num.	Acc*	Loss	Val. Acc.*	Val. Loss.
HMI*	944	1.0000	0.0003	0.9933	0.0119
Saliency map	10	1.0000	0.0038	1	0.0052
Guided Grad-CAM	9	1.0000	0.0253	0.96	0.0933

* Param. Type: parameter type; Acc.: accuracy; Val. Acc.: validation accuracy; HMI: human machine interface;

As a result, it showed high performance even only using about 10 parameters tracked by Saliency map and Guided Grad-CAM. Also, it showed an accuracy close to 100% similar to the model training with 944 parameter data. Through this, it can be validated that parameters with information that can classify 10 labels have a high relevance value. In other words, it can see that each method has well tracked the parameters relevant in classifying the test data with the correct label.

4. Conclusions

In this study, we explained a CNN model that classifies 10 labels which are NPP states to track parameters that are the basis of the model classification. When the CNN model correctly classifies a total of 10 test data for each label, relevance of each parameter was calculated by two model interpretation methods. Using only parameters with the highest relevance value for each label, the model was able to classify 10 labels with high accuracy. Through this, we validated that each method well selected the parameters that the model used to classify test data.

When a deep learning model is applied as an operator support system, we can provide the operator with information on which parameters the model diagnosed the NPP state as the main reason. This allows the operator to once again consider whether the model diagnosis is reliable. The model can be interpreted in various ways, and the result of the diagnostic relevance can be different. To apply this approach, evaluation and comparison of various methods for interpreting the model will be needed in a future study.

ACKNOWLEDGEMENTS

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20171510102040).

REFERENCES

[1] Knight, John C. "Safety-critical systems: challenges and directions." In Proceedings of the 24th international conference on software engineering, pp. 547-550. 2002.
[2] Endsley, Mica R., and Esin O. Kiris. "The out-of-the-loop performance problem and level of control in automation." Human factors 37, no. 2 (1995): 381-394.
[3] Doyle, Margery J., and K. Cole. "Impact of automation on decision making." (2008).
[4] Lee, Seung Jun, and Poong Hyun Seong. "Development of an integrated decision support system to aid cognitive activities of operators." Nuclear Engineering and Technology 39, no. 6 (2007): 703.
[5] Simard, Patrice Y., David Steinkraus, and John C. Platt. "Best practices for convolutional neural networks applied to visual document analysis." Icdar. Vol. 3. No. 2003. 2003.
[6] 3KEYMASTER simulator 2013, Western Service Corporation, Frederick, MD, USA.
[7] Simonyan, Karen, Andrea Vedaldi, and Andrew Zisserman. "Deep inside convolutional networks: Visualising image classification models and saliency maps." arXiv preprint arXiv:1312.6034 (2013).
[8] 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.