# Applying Supervised Learning Algorithm to Nuclear Power Plant Severe Accident Data Generated from MAAP Code

Seok Ho Song[a], Jeong Ik Lee[a]*

[a]*Department of Nuclear and Quantum Engineering N7-1 KAIST 291 Daehak-ro, Yuseong-gu, Daejeon, Republic of Korea 305-338, 1812wow@kaist.ac.kr*
**Corresponding author: jeongiklee@kaist.ac.kr*

## 1. Introduction

Since the Fukushima accident, the importance of research on nuclear power plant accidents and mitigation strategies has been emphasized globally. One of the research areas is evaluating the effectiveness of existing mitigation strategies for responding to various accident scenarios. The effectiveness of these strategies was measured with important indicators such as temperature and pressure inside a nuclear power plant [1-2]. The accident scenarios used in these analyses are generally produced from the probabilistic safety analysis (PSA), and pre-defined event trees. The accident sequences are evaluated by simplifying them to a level that can be evaluated at the level of currently available knowledge and tools. Even for accidents with the same sequence, analysis of detailed accident progress or unexpected complex events caused by differences in the entry time of a particular event is limited by conventional methods.

The PSA event tree, as seen in Figure 1, only identifies the accident sequence by branching a series of events that occur after a certain component fails. It should be noticed that the event tree does not include the time of component failure as variable. However, physically the propagation of an accident even for the same accident scenario, varies depending on the time of failure and sequence of failures [3]. It is anticipated that in the case of multiple failure accidents, both the timing of component failure and the beginning state of the accident will play a significant role in determining how the accident develops and manifests itself.

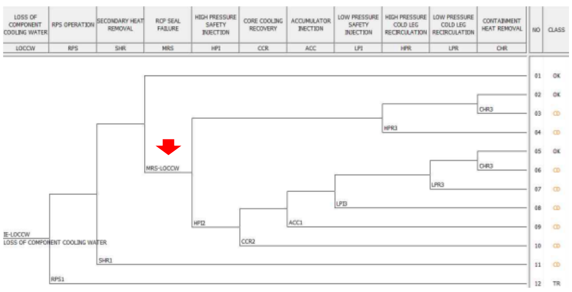


**Figure.1 PSA Event Tree in LOCCW Accident [4]**

In the previous study [5], the prediction of physical values in the loss of component cooling water (LOCCW) accident via an Artificial Neuron Network (ANN) is implemented. To simulate the progression of the whole accident, it should be possible to predict the status of the next state in time through the information of the initial state and to continuously update the state.

In this study, it is evaluated whether an accident scenario simulation is possible with continuous prediction using ANN. The accuracy of the trained network and the similarity of the accident prediction with the data generated from MAAP are evaluated.

## 2. Methods

This study uses LOCCW accident scenario data generated from MAAP as learning data. A double failure accident of reactor coolant pump (RCP) seal loss of coolant accident (LOCA) and high pressure injection (HPI) failure was simulated. The predictive ability of ANNs learned from accident data of RCP seal LOCA and random HPI failure in time following a log-normal distribution generated from MAAP was evaluated.

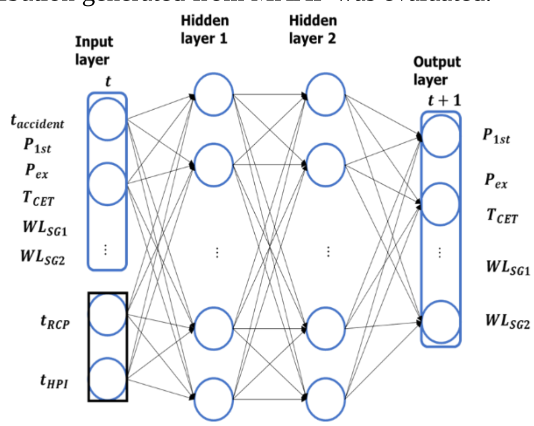



**Figure.2 Configuration Diagram of ANN**

The artificial neural network used in this study is a Multi-Layer Perceptron (MLP). The input layer that receives the input information and the output layer that produces the last information are connected via hidden layer. As can be seen in Figure.2, the input-layer consists of primary system pressure, ex-vessel pressure, Core Exit Temperature (CET), water level of Steam Generator (SG) 1, water level of SG2, accident progression time, RCP failure time, and HPI failure time. It is trained to predict the primary system pressure, ex-vessel pressure, Core Exit Temperature (CET), Steam Generator (SG) 1, and SG2 water levels after one hour with 8 input information. The above information is part of the information that can be checked at the nuclear power plant control facility in the event of an accident.

To this end, the accident data with 5-minute intervals is obtained from MAAP and is processed to be used as training data for supervised learning. All data from the start of the accident to 72 hours are coupled as a pair with 1-hour interval and used as input and answer for the supervised learning. As shown in Figure.2, the supervised learning ANN model is trained to produce the physical values after 1 hour when the physical values at a specific point in time are provided as input. At this time, the root mean square error of the coupled MAAP data and the predicted output is set as the loss, and the training is conducted in the direction of decreasing the loss. For the optimization, Adam optimizer was used. The Relu function was used as the activation function of the input layer and the hidden layer. The activation function of the output layer was the sigmoid function. The output of the ANN is normalized to have a value between 0 and 1, so that the output can be predicted stably even if the output of the ANN is used as an input for the next prediction.

Additionally, the accident data was used for developing a front model when the data is before 36 hours after the accident, and the remaining data was used for developing a back model. Thus, two independent ANNs were trained to simulate the 72-hour accident process as a whole. This approach was useful, since the accident data before the secondary cooling water is exhausted has a big difference from the pattern after the water was depleted. Therefore, the two data sets were separated and used for training the front model and the back model, respectively.

To evaluate the learned ANN, accuracy evaluation and plane accuracy evaluation using cosine similarity were performed. In this study, the ANN was trained using 70% of the data set, and the accuracy was evaluated using the 30% data set not used for the training, which is commo m data split ration in the computer science field.

For 8 input information $x_i$ at a specific point in time existing in data set X having a total of N input sets, $y_{MAAP,i}$ , the correct answer of the prediction after 1 hour, already exists from MAAP simulation. In this case, the predicted value $y_{pred,i}$ can be obtained when using $x_i$ as input through ANN.

$$ANN(x_i) = y_{pred,i} \qquad (eq.1)$$

A new function *f* was defined that returns 1 when the predicted value $\mathbf{y_{pred,i}}$ has an actual error of $\mathbf{y_{MAAP,i}}$ less than or equal to 0.05, otherwise 0.

$$\boldsymbol{f(x_i) = \begin{cases} 1 & if \ \|y_{pred,i} - y_{MAAP,i}\| < 0.05 \\ 0 & else \end{cases}} \qquad (\boldsymbol{eq.2})$$

Accuracy is defined as the average value of $\mathbf{f(x_i)}$ for all $\mathbf{x_i}$ in the dataset.

$$\boldsymbol{Accuracy = \Sigma_{i=1}^{N}\left(\frac{f(x_i)}{N}\right) \times 100} \ (\%) \qquad (\boldsymbol{eq.3})$$

The accuracy test can show the predictive capability of surrogate model for unit time (i.e. 1 hour) simulation. However, the similarity between the 72-hour accident progression using the surrogate model and the calculation result of the MAAP code cannot be evaluated with the accuracy test. Therefore, the cosine similarity test was next used. The cosine similarity test is a method to evaluate the similarity of two vectors regardless of their sizes. It is mainly used to evaluate the similarity of character strings or image data.

As shown in Figure.3, if the accident is simulated with the surrogate model, a plane can be created for the accident time and HPI failure time for each output. MAAP data can also create a plane for the accident time and HPI failure time for the entire data set. At this time, the plane of the surrogate model is located in the normalized space, and the plane of the MAAP data is located in the non-normalized space. To evaluate the similarity between the surrogate model and the MAAP data, the relative position and pattern of this plane should be similar, not the absolute distance. Therefore, as shown in Eq. 4, cosine similarity was calculated by treating variables having the same accident time and HPI failure in both planes as vectors. The plane similarity of the entire model was calculated as the average of these cosine similarities (Eq. 5). Accuracy and plane accuracy were evaluated for ANN of 2, 3, 4, layered structures, respectively, and the difference between each ANN structure and its index was evaluated.

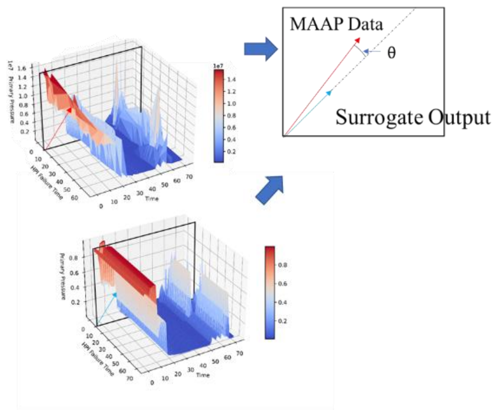



**Figure.3 Conceptual Diagram of Cosine Similarity**

$$\cos(\theta) = \frac{\overrightarrow{MAAP} \cdot \overrightarrow{Surrogate}}{\|\overrightarrow{MAAP}\|\|\overrightarrow{Surrogate}\|} \qquad (\text{eq.}4)$$

$$\text{Plane accuracy} = \frac{\Sigma_{t=0}^{T}\Sigma_{i=0}^{N}(\cos(\theta_{i,t}))}{NT} \qquad (\text{eq.}5)$$

### 3. Results & Discussions

The accuracy of ANN is shown in Table.1. Overall, the accuracy of the test set shows that the accuracies of the ANNs for training set are high. This is expected by training the front model and the back model by dividing the data in half while maintaining the size of the ANN. Since the accuracy of the test set does not differ significantly from the accuracy of the training set, it is difficult to determine if an overfitting had occurred. All types of ANN succeeded in stably predicting 72 hours continuously, and based on this, it was possible to obtain an output surface with respect to HPI failure time and accident progression time.

**Table.1 Accuracy of ANN**

| Accuracy | | | |
|---|---|---|---|
| Layer number | Front /Back | Test set | Training Set |
| 2L | Front | 97.16 | 98.17 |
| | Back | 95.87 | 98.13 |
| 3L | Front | 96.60 | 97.30 |
| | Back | 95.39 | 96.76 |
| 4L | Front | 96.88 | 98.22 |
| | Back | 95.22 | 96.69 |

In order to evaluate the accuracy of predicting the entire accident scenario of the MAAP data and the entire accident scenario through ANN, the plane accuracy test was defined previously, and the results are shown in Table 2.

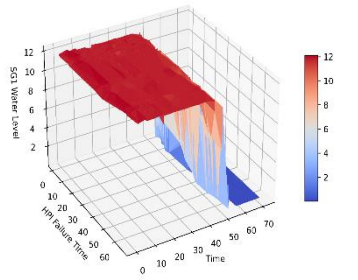
**Table.2 Plane Accuracy of ANN**

| Target | 2L model | 3L model | 4L model |
|---|---|---|---|
| Primary Pressure | 0.3672 | 0.3007 | 0.4624 |
| Ex-vessel Presure | 0.7779 | 0.7842 | 0.7315 |
| CET | 0.5667 | 0.5623 | 0.5983 |
| SG1 Water Level | 0.8533 | 0.8605 | 0.8789 |
| SG2 Water Level | 0.8542 | 0.8604 | 0.8780 |
| Total | 0.6839 | 0.6736 | 0.7098 |

The water levels of SGs have a relatively simple surface shape and therefore showed high plane accuracy. The top of Fig.4 is the water level of SG1 from MAAP data, and the bottom of Fig.4 is the water level of SG1 of the 4 layered ANN model with the highest plane accuracy. It can be seen that the predicted values of the surrogate model are very similar with the MAAP data.

Also, the top of Fig. 5 is the ex-vessel pressure of MAAP data, and the bottom of Fig. 5 is the ex-vessel pressure of the 3 layered ANN model. It was confirmed that the similarity was high when the plane accuracy was high even when the plane has more complicated shape than the SGs' water level.

The top of Fig. 6 is the surface of primary pressure from the MAAP data, and the bottom of Fig. 6 is the surface of the primary pressure predicted with the 3 layered ANN model. From the output with the lowest plane accuracy, it was confirmed that the pressure in the early 0 to 20 hours of the accident was predicted to be higher than the MAAP data. However, the pressure peak that occurs with RPV failure was predicted correctly between the low-pressure section in the midst of the accident between 50 and 60 hours. It succeeded in predicting the peak occurrence but failed to predict the exact value. In particular, concerning HPI failure time, the pressure of 0 to 20 hours in the first half of the accident is not very sensitive, so this needs to be further




refined.

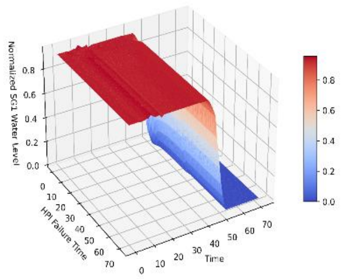
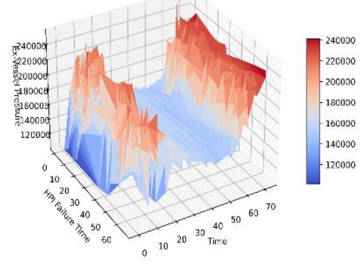



**Figure.4 Steam Generator 1 Water Level Surface of MAAP(Top), 4-L ANN (Bottom)**

**Figure.5 Ex-vessel Pressure Surface of MAAP(Top), 4-L ANN (Bottom)**



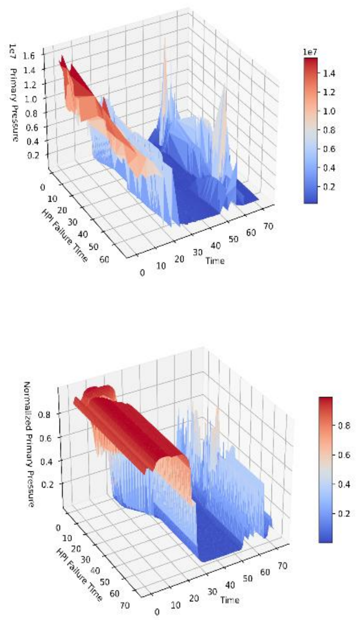



**Figure.6 Primary Pressure Surface of MAAP(Top), 4-L ANN (Bottom)**

## 4. Conclusions & Further Works

From this study, it was possible to show the possibility of simulating large-scale long-term accidents such as severe accidents in a nuclear power plant with ANN. The existing MAAP code is capable of simulating a severe accident but it takes a long time to simulate a single accident, thus it requires a lot of computational resources for repeated calculations. However, the simplified accident progression can be modeled with high speed if the supervised learning is applied.

However, it is necessary to improve the accuracy of accident simulation with ANN. It is expected that such accuracy can be improved further by using ANN specialized for time series data such as LSTM algorithm. Additionally, adjusting the time step of the surrogate model can also improve the results. The surrogate model used in this study simplified the 72-hour accident scenario and predicted the parameters for every 1-hour interval. However, as this time step becomes smaller, the authors predict that the simulation model can become more accurate.

## Acknowledgment

This work was supported by KOREA HYDRO & NUCLEAR POWER CO., LTD (No. 2020-Tech-01).

## REFERENCES

[1] Lee, Seongnyeon, et al. "Validation of RCS depressurization strategy and core coolability map for independent scenarios of SBLOCA, SBO, and TLOFW." Journal of Nuclear Science and Technology 51.2 (2014): 181-195.

[2] Jeon, Joongoo, et al. "Numerical investigation of in-vessel core coolability of PWR through an effective safety injection flow model using MELCOR simulation." Annals of Nuclear Energy 121 (2018): 350-360.

[3] Xiong, Qingwen, et al. "Global sensitivity analysis for nuclear reactor LBLOCA with time-dependent outputs." Reliability Engineering & System Safety 221 (2022): 108337.

[4] Korea Hydro & Nuclear Power Co. Ltd. “PSA-MLOCA- Accident Circumstance Analysis”

[5] Song, Seok Ho et al. “Feasibility Estimation of Development of Real-time Interactive Nuclear Power Plant Accident Simulation Using Supervised Learning.” Korean Nuclear Society Spring Meeting. 2022.