# Operational Anomaly Diagnosis Algorithm using Machine Learning Technology with MARS-KS Transient Simulation Database

S. M. Park [a*], J. Y. Lee [a], Y. K. Kwack [a], N. H. Hoang [a], S. K. Sim [a]
[a] *Environment and Energy Technology, Inc., 96 Gajeongbuk-ro, Yuseong-gu, Daejeon, Korean 34111*
[*] *Corresponding author: psm93@en2t.com*

## 1. Introduction

Recently, initiatives to apply artificial intelligence (AI) technology to nuclear power plant (NPP) are in progress globally [1-3]. Various deep neural networks have been applied to diagnose plant operating conditions to support operators for effective operator action to prevent and mitigate accidents of the NPP [4, 5]. Thus, various machine learning techniques are applied to diagnose and predict operational abnormal conditions in advance to support operators. This study aims to develop an operational anomaly diagnosis machine learning algorithm using MARS-KS simulated safety analysis database.

The operational anomaly diagnosis algorithm is a supervised learning model that can diagnose normal and abnormal conditions of the NPP using NPP transient database calculated by the MARS-KS [6] best estimate safety analysis code. For application to the anomaly diagnosis, Long-Short-Term-Memory (LSTM) neural networks based on Recurrent Neural Networks (RNNs), Light Gradient Boosting Model (LGBM) and Categorical Boosting (CatBoost) based on boosting type decision tree were used. This study presents a machine learning algorithm that can help the operator to determine for its action for the NPP by diagnosing the abnormal situation to prevent and mitigate the accident of the NPP.

## 2. Machine Learning Models for Operational Anomaly Diagnosis

In this study, machine learning models are used to diagnose the state of abnormal operational conditions. These models include LSTM, LGBM, and CatBoost.

### *2.1 LSTM*

LSTM is a type of RNN (Recurrent Neural Network), a model that predicts future data by considering the preceding and the past data more macroscopically [7]. It consists of a total of 6 parameters and 3 gates as shown in Fig. 1. The functions in Fig. 1 are computed by Equations (1) through (6).

$$i_t = \sigma(x_t U^i + h_{t-1} W^t) \quad (1)$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f) \quad (2)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o) \quad (3)$$

$$\tilde{C}_t = \tanh(x_t U^g + h_{t-1} W^g) \quad (4)$$

$$C_t = \sigma(f_t \cdot C_{t-1} + i_t \cdot \bar{C}_t) \quad (5)$$

$$h_t = \tanh(C_t) \cdot o_t \quad (6)$$

where, $\sigma$ is the sigmoid function as defined in Eq. (7).

$$\sigma(x) = \frac{1}{1+e^x} \quad (7)$$

$f$, o and $g$ are the input, output and current gate. $x$ and $h$ are the input and hidden state. $t$ is the time step. $U$ and $W$ are the weighting factors.

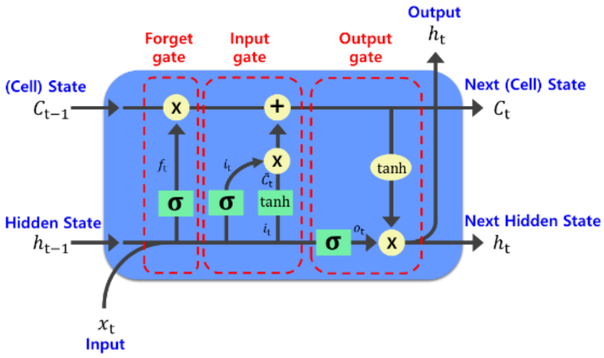

Fig. 1. Overview of LSTM RNN layer.

The cell state ($C_t$) is like a conveyor belt and it applies only for a small linear interaction and keeps the entire chain running. It is the part that keeps the information flowing without changing at all. In addition, the gradient spreads well even after a long period of the state. And the information is added or removed by a structure called gates, and the input data information is maintained or discarded through the training. The forget gate is the process of deciding whether to discard past information or not. The forget gate takes $h_{t-1}$ and $x_t$, and sends a value between 0 and 1 to $C_{t-1}$. If its value is 1, it saves all the information, and if it is 0, it drops all information. The input gate is a gate for storing current information and has the role of determining whether or not to add to the current cell state value. The update is the process of updating the old cell state to a new state. Users have to decide how much to throw away and how much to add at the input gate through the forget gate, and it will be calculated during the update process and update the input to the old cell state. The output gate determines which output value goes to the output and how much of the cell state value is subtracted.

### *2.2 LGBM*

LGBM is based on the basic decision tree among various machine learning algorithm models and it is an algorithm used for Ranking or Classification [8]. In the process of increasing the branches of the tree, it calculates the gradient and divides the leaf at the part where the loss function can be reduced the most, and creates a tree that learns more intensively of the data that was not learned well in each previous step. Among machine learning algorithm models, tree models are built on a decision tree and the LGBM extends the tree vertically, unlike other tree algorithms that scale horizontally.

### *2.3 CatBoost*

CatBoost is an open-source software library developed by Yandex and provides a gradient boosting framework [9]. This framework of tree algorithms scales horizontally. If the existing boosting model computed the residuals for all the training data in batches, CatBoost calculates the residuals with only a portion of the data, creates a model with it, and then uses the value predicted by this model for the residual of the data. Among other features, CatBoost attempts to solve categorical features using permutation-driven alternatives compared with traditional algorithms.

## 3. Methodology of Operational Anomaly Diagnosis

### *3.1 Data Set*

Since it is difficult to collect abnormal operational data from actual nuclear power plants, the data were generated by calculating performance and safety analyses of the APR1400 using MARS-KS best estimate nuclear thermal-hydraulic safety code. Fig. 2 shows the MARS-KS nodalization of the APR1400, the reference NPP of this study.

Starting from a normal steady state with the best estimate input of the APR1400, transient analyses were performed and the changes of major parametric variables over time were calculated. The cause of the abnormal operational condition is classified by simulating the failures of the system and equipment, and the cause of the abnormality is diagnosed through supervised learning.

The abnormal operational scenarios were selected as IOSGADV (Inadvertent Opening S/G Atmospheric Dump Valve) and MSLB (Main Steam Line Break). Also included were normal conditions and performance related design basis event (PRDBE) for a 10% step-change increase of turbine power after 10 seconds of the steady state from a normal steady state at 90% core and turbine power. The input variables were set based on the APR1400 FSAR. Normal steady state including PRDBE, IOSGADV, and MSLB data were trained, and the break size was changed for training various abnormal scenarios of the IOSGADV and MSLB. The data were also tested as unlearned IOSGADV, MSLB abnormal data and PRDBE data.

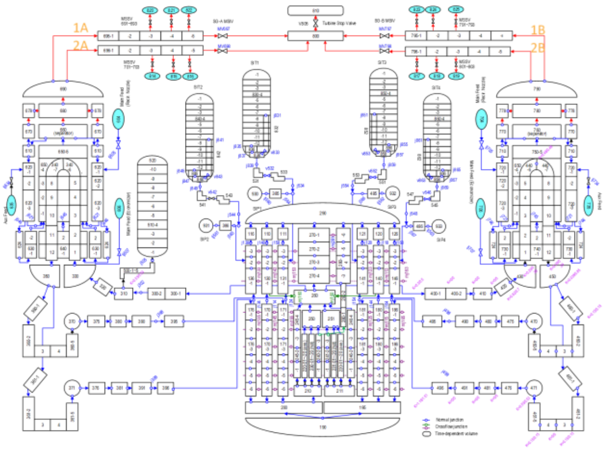

Fig. 2. APR1400 MARS-KS Nodalization

### *3.2 Selection of important parameters*

By comprehensively considering the variables of the MARS-KS analyses, important variables were selected according to the type and location of the detector signals and important safety related variables of the NPPs. Total of 95 variables were selected as important parameters from the MARS-KS performance and safety analysis database including 23 variables each for the pressure and temperature, 38 variables for mass flow rate, 4 variables for water level, 6 variables for void fraction, and 1 variable for reactor thermal power as shown in Table 1.

Table 1: Important parameters of MARS-KS analyses

| Variable | Component | Number |
|---|---|---|
| Thermal power | Reactor Core | 1 |
| Pressure | Reactor Pressure Vessel (RPV) | 2 |
| | Steam Generator (SG) | 2 |
| | Pressurizer (PZR) | 1 |
| | Cold Leg (C/L) | 4 |
| | Hot Leg (H/L) | 2 |
| | Reactor Coolant Pump (RCP) | 4 |
| | Main Steam Line (MSL) | 4 |
| | Safety Injection Tank (SIT) | 4 |
| Temp | Reactor Pressure Vessel (RPV) | 1 |
| | Pressurizer (PZR) | 2 |
| | Steam Generator (SG) | 2 |
| | Cold Leg (C/L) | 4 |
| | Hot Leg (H/L) | 2 |
| | Reactor Coolant Pump (RCP) | 4 |
| | Main Steam Line (MSL) | 4 |
| | Safety Injection Tank (SIT) | 4 |
| Mass flow rate | Core | 4 |
| | Steam Generator(SG) | 10 |
| | Cold Leg (C/L) | 4 |

| | | |
|---|---|---|
| | Hot Leg (H/L) | 2 |
| | Reactor Coolant Pump (RCP) | 4 |
| | Main Steam Line (MSL) | 4 |
| | Safety Injection Tank (SIT) | 4 |
| | Main Feed Water Line (MFWL) | 4 |
| | Aux Feed Water Line (AFWL) | 2 |
| Water level | Pressurizer (PZR) | 1 |
| | Steam Generator (SG) | 2 |
| | Core | 1 |
| Void fraction | Pressurizer (PZR) | 1 |
| | Reactor Pressure Vessel (RPV) | 1 |
| | Cold Leg (C/L) | 4 |
| Total Number of Variables | | 95 |

*3.3 Data Preprocessing*

The data preprocessing sequence includes data extraction, data normalization, and data dimension change. First, the data is extracted from the MARS-KS calculation results, and the extracted data is first pre-processed according to the python program format. Second, the preprocessed data is then normalized using the MinMaxScaler method. Third, data dimension change is required for normalized data to a three-dimensional array in the case of LSTM and a two-dimensional array in the case of LGBM and CatBoost.

*3.4 Modeling*

In the case of LSTM, it was modeled based on the trial and error method, and, for the cases of LGBM and CatBoost, the optimal value was found using the grid search method. Tables 2 and 3 show the hyper-parameters of the LSTM, LGBM and CatBoost, respectively.

Table 2: Hyper-parameter for LSTM

| Layer | Time Step | Activation | Optimizer |
|---|---|---|---|
| 5 | 10 | Softmax | Adam |

Table 3: Hyper-parameter for LGBM and CatBoost

| | n_ estimator | learning rate | max_ depth | num_ leaves |
|---|---|---|---|---|
| LGBM | 300 | 0.03 | 5 | 16 |
| CatBoost | 400 | 0.04 | 12 | 5 |

**4. Results of the Operational Anomaly Diagnosis**

The results of the operational anomaly diagnosis analyses using hyper-parameters are shown in Fig. 3 - 5 for the IOSGADV, MSLB, PRDBE and NORMAL steady state cases. For each case, an anomaly diagnosis algorithm is verified using a test date set from a steady state of 0.0 s to a transient state of abnormal end point. The accuracy of operational anomaly diagnosis was tested by inserting even a normal case between the trained and untrained abnormal data cases.

Fig. 3 and 4 show the correct anomaly diagnosis result of 1.0, e.g., 100% accuracy and reliability for the LSTM and LGBM models, respectively. The results show that each abnormal condition is accurately and continuously diagnosed. Fig. 5 shows the result of the CatBoost model. In the case of CatBoost, different results were presented for the test data of the IOSGDAV and MSLB, NORMAL and PRDBE. The test of the data diagnosed about 90 - 100% for the IOSGADV and about 0 - 10% for the MSLB. NORMAL is dominant for the test of the NORMAL steady state data and about 0 - 11% of the data was diagnosed as the PRDBE.

Since the anomaly diagnosis model estimates the probability of the result through each feature, the accuracy of the model was quantitatively evaluated using evaluation metrics of the regression models. The evaluation metrics include mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination ($R^2$).

Table 4 shows the accuracy of each anomaly diagnosis model. From the results of this test set calculation, LGBM is the most accurate, followed by LSTM, and CatBoost is relatively inaccurate. In this study, it was shown that the accuracy of the anomaly diagnosis algorithm model greatly depends on the training conditions of the anomaly diagnosis algorithm model as well as the test data set calculated by the MARS-KS best estimate safety analysis code.

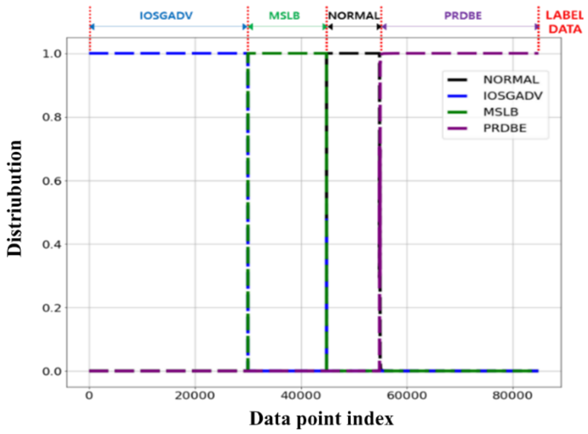


Fig. 3. Anomaly Diagnosis Result of LSTM

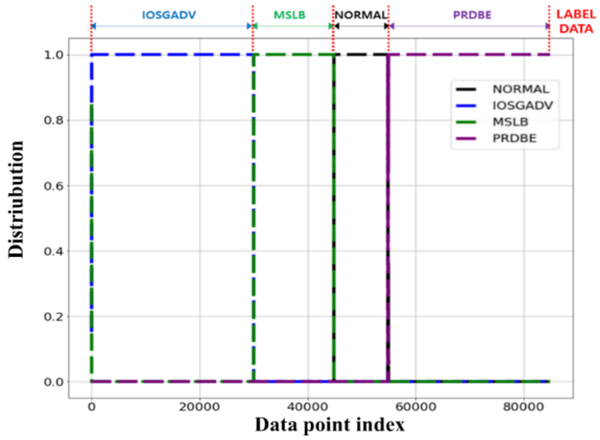


Fig. 4. Anomaly Diagnosis Result of LGBM

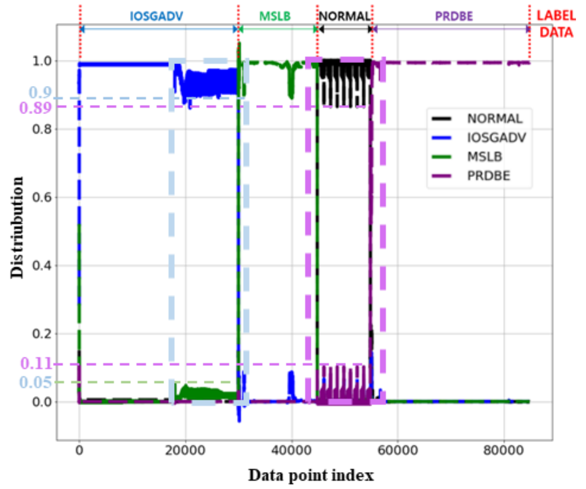


Fig. 5. Anomaly Diagnosis Result of CatBoost

Table 4. Performance of the Models

| Model | R-squared | MAE | MSE | RMSE |
|---|---|---|---|---|
| LSTM | 0.99773 | 0.00049 | 0.00032 | 0.01784 |
| LGBM | 0.99969 | 0.00009 | 0.00005 | 0.00673 |
| CatBoost | 0.99786 | 0.00643 | 0.00042 | 0.02040 |

### 5. Conclusions

In this study, operational anomaly diagnosis was performed using LSTM, LGBM, and CatBoost, which are widely used AI deep learning diagnosis algorithm models in anomaly data diagnosis and prediction. It was confirmed that the operational anomaly diagnosis was performed well. And, this result can be used to support operators to make decisions for their actions in case of the operational abnormal conditions and can assist the operator in search for the variables that have a major cause of the abnormal condition and its impact. In future work, we plan to develop operational anomaly prediction algorithm using the operational anomaly diagnosis algorithm of this study. Also, this operational anomaly diagnosis algorithm will be coupled with the anomaly detection model using unsupervised learning to distinguish between the trained and untrained abnormal cases.

### ACKNOWLEDGMENTS

This study has been funded and supported by the Nuclear Safety Research Program through the Korea Foundation of Nuclear Safety (KoFoNS) using financial resources granted by the Nuclear Safety and Security Commission(NSSC) of the Republic of Korea (No. 2103087).

### REFERENCES

[1] No, Y, G., Kim, J, H., Na, M, G., Lim, D, H., and Ahn, K, I., 2012, Monitoring Severe Accidents Using AI Techniques, Nuclear Engineering and Technology, Vol 44, 4, 393-404.
[2] Hanna, B., Son, T, C., and Dinh, N., 2019, An Artificial Intelligence-Guided Decision Support System for The Nuclear Power Plant Management, 18th International Topical Meeting on Nuclear Reactor Thermal Hydraulics Portland, OR, August, 394-406
[3] Hanna, B., Son, T, C., and Dinh, N., AI-Guided Reasoning-Based Operator Support System for the Nuclear Power Plant Management, Annals of Nuclear Energy ,154, 108079.
[5] Yang, J., Kim, J., 2018, An accident diagnosis algorithm using long short-term memory, Nuclear Engineering and Technology, 50, 582-588.
[6] Yang, J., Kim, J., 2020, Accident diagnosis algorithm with untrained accident identification during power-increasing operation, Reliability Engineering and System Safety, 202, 107032.
[7] Korea Institute of Nuclear Safety (KINS), 2018, MARS-KS code manual Volume II: Input Requirements. KINS/RR-1822 Vol. 2.
[8] Sak, H., Senior, A., Beaufays, F., 2014, Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. Google
[9] Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021, A comparative analysis of gradient boosting algorithms, Artificial Intelligence Review, volume 54, 1937–1967.