[UCI]I804:24011-200000359994

2021년 2월

석사학위 논문

Algorithm for Detection of Signal Stuck Error in Emergency Situation using Unsupervised Learning

조선대학교 대학원

원 자 력 공 학 과

윤 경 민

Algorithm for Detection of Signal Stuck Error in Emergency Situation using Unsupervised Learning

원자력발전소 비상 상황 시 비지도 학습을 사용한
고착 신호 탐지에 대한 알고리즘

2021년 2월 25일

조선대학교 대학원

원 자 력 공 학 과

윤 경 민

Algorithm for Detection of Signal Stuck Error in Emergency Situation using Unsupervised Learning

지도교수 김 종 현

이 논문을 공학 석사학위신청 논문으로 제출함

2020년 10월

조선대학교 대학원

원 자 력 공 학 과

윤 경 민

윤경민의 석사학위논문을 인준함

위원장　조선대학교　교수　나 만 균　(인)

위　원　조선대학교　교수　김 진 원　(인)

위　원　조선대학교　교수　김 종 현　(인)

2020년 11월

조선대학교 대학원

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

원자력발전소 비상 상황 시 비지도 학습을 사용한 고착 신호 탐지에 대한 알고리즘

윤 경 민
지도교수 : 김 종 현
원자력공학과
조선대학교 대학원

원전 비상 상황에서 신호 오류는 운전원을 혼란스럽게 하며, 이로 인한 운전원의 잘못된 조치는 중대한 인명 사고를 초래할 수 있습니다. 따라서, 비상 상황에서 신호의 변화가 사고에 의한 것인지 고장 신호에 의한 것인지 구별하는 것이 매우 중요합니다. 이전 연구에서는 데이터 기반 방법론을 사용하여 이상 신호를 탐지하였습니다. 하지만 방법론의 한계로 모든 상황에 대한 고장을 탐지하지는 못하였습니다. 본 연구는 이러한 한계점을 타개하기 위해 비지도 학습을 사용한 이상 신호 탐지 알고리즘 개발을 목표로 합니다. 제안하는 알고리즘은 프로세스 최적화와 신호 실패 결정 알고리즘, 두 단계로 구성됩니다. 프로세스 최적화는 알고리즘의 성능을 향상시키고, 신호 실패 결정 알고리즘은 이상 신호를 판별합니다. 프로세스 최적화에서 훈련이 어렵다는 비지도 학습의 단점을 보완하기 위해 피어슨 상관 계수 방법을 사용하여 모델의 입력을 결정하였습니다. 선정된 입력을 사용하여 모델을 학습하고, 학습된 모델은 문턱값을 정의하는데 사용되었습니다. Compact Nuclear Simulator (CNS)를 사용하여 냉각재상실사고(Loss of Coolant Accident; LOCA) 데이터를 수집한 후, 수집된 데이터를 사용하여 제안된 알고리즘의 훈련 및 검증을 수행하였습니다.

1. Introduction

For nuclear power plants (NPPs), signals are an important system that supports the operation by showing the status of the plant to the operator. At this time, faulty and conflicting sensor readings may often degrade the performance of the control system, confuse operators, and lead to actions that may compromise the safety of NPPs [1]. During the Fukushima Dai-Ichi accident, the NPPs lost most Instrumentation & Control (I&C) systems and some instruments provided erroneous signals to the operators [2]. For this reason, the operator made an inappropriate decision, which led to a major accident [2-4]. In this emergency situation, the operator's decision is likely to result in massive human casualties. Therefore, the integrity of the signal that is the basis for the operator's decision is very important. Recently, the interest in autonomous control is increasing and then the reliability of signal becomes more important for the success of it. For this reason, on-line monitoring (OLM) techniques of sensors and signals have been an active research area in NPPs [5].

Many researches have been suggested for the detection of anomaly signals so far. The approach for the detection of anomaly signals can be divided into model-based or data-driven approaches. Model-based approaches were applied early in the study by understanding the physical mechanisms of the system and presenting accurate models. Examples include Kalman Filter [6, 7, 13], Parity equations [8-10], and Parameter estimation [11, 12]. However, model-based approaches are not suitable for complex models, such as NPPs, because these models need to accurately understand physical mechanisms and present models. Data-driven approach are using historical operational data without accurate model presentations, such as the model-based approach. Typical methods include Principal Component Analysis (PCA) [14, 19-21], Auto Associative Kernel Regression (AAKR) [15, 16], Artificial Neural Networks (ANNs) [17, 18, 22-27].

Among the data-driven methodologies, various ANN methodologies were used [18, 22-27]. ANN methodologies can be divided into supervised learning and unsupervised learning. Supervised learning entails learning a mapping between a set of input variables X

and an output variable Y and applying this mapping to predict the outputs for unseen data [28]. Unsupervised learning studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns [29]. Most previous studies applied supervised learning for the detection of the anomaly signals [17, 22, 24-26]. This is because unsupervised learning is difficult in that it is necessary to learn patterns on its own, unlike supervised learning, which learns a given pattern of input/output variables. However, it is virtually impossible to train all signal failure types in NPPs consisting of thousands of signals and various types of failures, as supervised learning can only handle trained data. For this reason, this study used unsupervised learning that can handle untrained data.

Some of the previous studies focused on failure detection in the steady states [15, 16, 18-24, 26] in which the signal is not changing rapidly. However, unlike the steady state, in an emergency situation is thousands of signals change dramatically at the same time. Therefore, it is necessary to accurately distinguish whether the signal change is due to an accident or anomaly signal. This means that detection of anomaly signals in an emergency situation is more important than steady state.

This study aims to develop the algorithm for the detection of the anomaly signals in an emergency situation and used Variational Auto-encoder (VAE), a representative production model for unsupervised learning, and Long Short-Term Memory (LSTM) that processes time-series data such as NPP parameters. In order to compensate and optimize the proposed algorithm, which is difficult to learn in unsupervised learning, the determination input and determination hyper parameter steps were performed. In addition, to improve the performance of the algorithm, we performed a step of determination threshold for optimization. When the optimization process of the algorithm for the validation of the algorithm was completed, a failure was injected and the validation was performed, and the algorithm showed a result of detecting 97.39% of the total failures.

2. Signal failure in an emergency situation

Especially, in an emergency situation, if the sensor misrepresents the status of the device and the current situation to the operator, it can cause serious human accidents. This section describes the types of sensor failures and the reasons for the detection of the anomaly signals in case of an emergency situation.

2.1 Types of anomaly signals due to sensor failures

Sensor failures can occur in external or internal environmental causes such as environmental pollution, vibration, extreme temperature fluctuations, and aging of sensors. Many researches have been defined for the sensor failure types of far. As shown in Fig. 1, sensor failures are divided into fault such as bias error, drift error, stuck at a constant value [20, 26, 30-33].

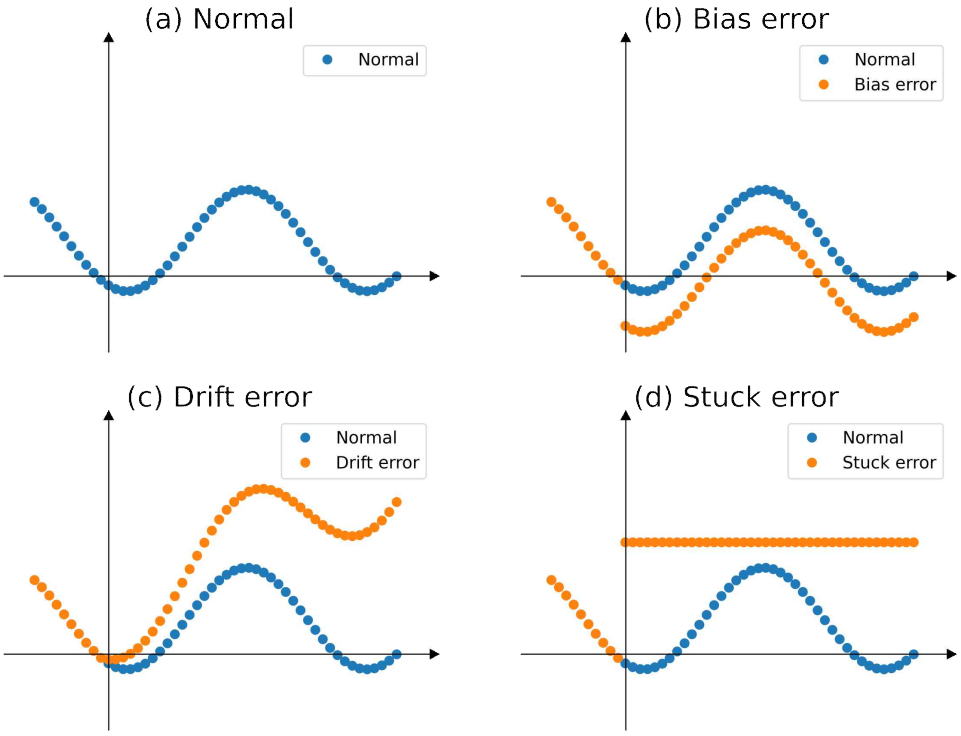

Fig. 1. Three typical types of sensor error

And, sensor failures can be parameterized in the following equations [31]:

$$y_i(t) = x_i(t), \text{ for } all\ t \geq 0 \tag{1}$$

$$y_i(t) = x_i(t) + d_i,\ \dot{d}(t) = 0,\ d_i(t_F) \neq 0 \tag{2}$$

$$y_i(t) = x_i(t) + d_i,\ \dot{d}(t) \neq 0, \text{ for } all\ t \geq t_F \tag{3}$$

$$y_i(t) = x_i(t_{F_i}) \text{ or } Constant, \text{ for } all\ t \geq t_F \tag{4}$$

Equations are normal (1), bias (2), drift (3), stuck at a constant (4). Where denotes the injected time of fault, and denotes the arbitrary constant.

The bias error in Eq. (2) is that constant values without variation are constantly added to the output. The drift errors in Eq. (3) indicate that the sensor output changes slowly regardless of the measurement function and, the stuck error in Eq. (4) is that the value is fixed at a constant value and is typically divided into high, low, and current errors. Additionally, noise errors can also be divided into a type of fault [27, 34-36], and noise errors are caused by electric fluctuations within components used in the measuring instrument. However, this study did not classify the noise into the type of failure because the actual signal value of the NPPs may contain noise. Besides, since the dataset through CNS is data without noise, 5% of Gaussian white noise was arbitrarily injected.

In the case of the Fukushima accident, the loss of most Instrumentation & Control (I&C) systems due to long-term Station Blackout (SBO), which resulted in the operator made an unsuitable decision, which led to a major accident [2-4]. Therefore, in this study, the stuck error in an emergency situation is classified as a failure in consideration of the case of such a signal stop. And, fault types are classified into three types: stuck at a high value, stuck at a low value, and stuck as is current.

2.2 Previous studies with data-driven method in NPP

In power plants, signal validation studies continued for decades. Table 1 shows the data-driven method applied to NPPs to detect anomaly signals. In order to detect drift errors during steady state, Li et al. [19] and Kaistha et al. [21], two studies, used PCA, and Di Maio et al. [16] applied AAKR. The PCA and AAKR methodologies used are limited in applying all the data because data must be grouped with similar data sets. Fantoni et al. [24] detected drift and stuck errors in steady state using ANN, one of supervised learning. Choi et al. [25] used LSTM to detect drift and stuck errors in emergency situations. Since the supervised learning method is a method of learning the correct answer Y to the problem X, it derives completely different answers to the unknown problems. Therefore, it is not suitable to be considered for all situations. Hines et al. [23] detected drift errors in steady state using AANN, an unsupervised learning method. In addition, Kim et al. [27] proposed a methodology to detect noise errors in emergency

situations using VAE, an advanced form of AANN. Unsupervised learning can overcome the limitations of supervised learning, but has a limitation in that learning is difficult.

Table 1. Data-driven method applied in NPP

Method	Reference	Technique	Situation	Error type	Limitation
PCA	Li et al. [19]		Normal	Drift	It is difficult to learn about all data because similar data is grouped and then model is created by group.
	Kaistha et al. [21]		Normal	Drift	
AAKR	Di Maio et al. [16]		Normal	Drift	It is difficult to learn about all data because similar data is grouped and then model is created by group.
ANN (Supervised learning)	Fantoni et al. [24]	ANN	Normal	Drift, Stuck	Anomaly detection is possible only for learned situations, and unlearned situations cannot be detected.
	Choi et al. [25]	LSTM	Emergency	Drift, Stuck	
ANN (Unsupervised learning)	Hines et al. [23]	AANN	Normal	Drift	Unsupervised learning is difficult to train because it has to learn patterns of data by itself.
	Kim et al. [27]	VAE	Emergency	Noise	

*FNN: Fuzzy Neural Network

*AANN: Auto Associative Neural Network

Unsupervised learning can be used to solve the aforementioned limitations of PCA, AAKR, and supervised learning. However, unsupervised learning has a problem that learning is difficult because it requires self-learning the pattern of training data. Therefore, to solve this problem, in this study, optimization steps were performed and applied to the

algorithm. In Kim's study, anomaly signal detection was performed by applying VAE, but since only noise error is applied, it cannot be applied to stuck errors that may occur in emergency situations. In order to overcome the limitations of supervised learning and develop the algorithm used in the optimization steps, the following section describes the VAE-LSTM methodology used in this study.

3. Methodology

This section describes VAE, which is a type of unsupervised learning, used to develop anomaly signals detection algorithms, and LSTM for processing time-series data.

3.1 Variational auto-encoder

VAE is a variant of an Auto-encoder (AE) rooted in Bayesian inference [37]. VAE is one of the ANNs, which is unsupervised learning, and is a model that learns to restore output values similar to input values.

The VAE consists of an encoder at the front and a decoder at the rear that are connected to each other. The encoder is made of an overall narrower shape with fewer nodes in subsequent layers than in previous layers. Conversely, the decoder has a wider overall pattern, with the later layers having more nodes than the previous layers. The encoder compresses the input data and performs dimension reduction, expressing a smaller number of parameters. And the encoder deduces probability distribution parameters of decoder inputs, instead of directly deducing inputs for the decoder (i.e. input of VAE's decoder is a random variable from continuous probability distribution). Accordingly, the decoder receives various inputs (probabilistic) even though the original input of the entire model is same. The decoder plays a role of restoring the compressed data back to the existing input data. The input of the decoder is derived through sampling from the corresponding probability distribution, and for this reason, it always produces different outputs for the same input [27, 38]. This allows VAE to be used not only as a model for dimension reduction, but also as a generation model that can generate new data. The structure of VAE is shown in Fig. 2.

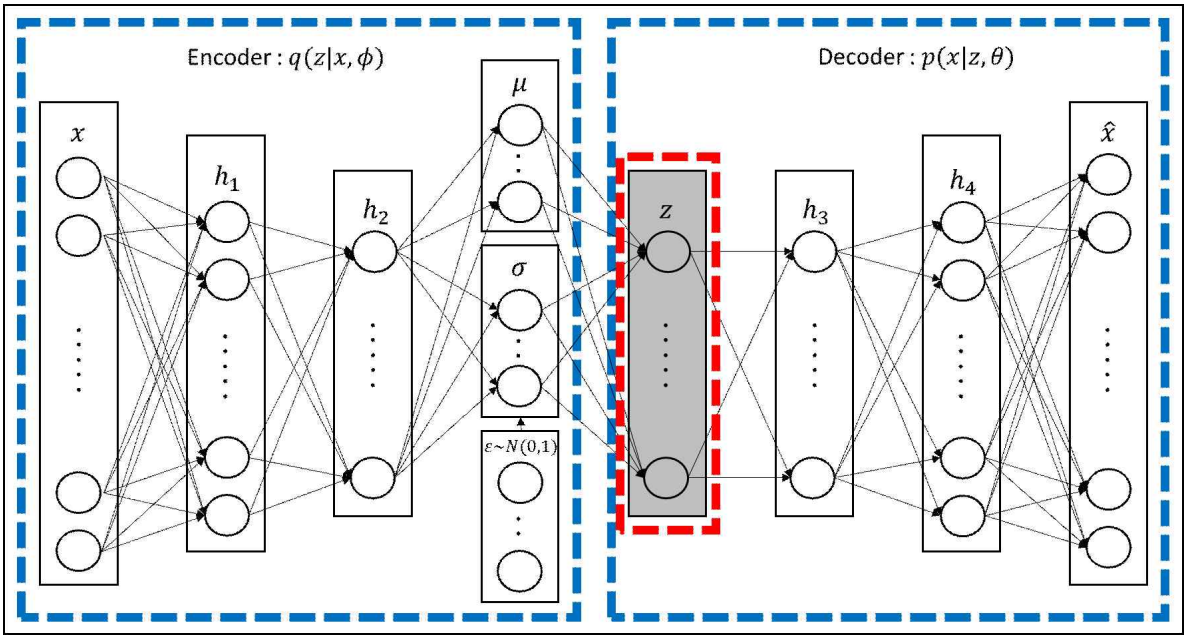

Fig. 2. The architecture of VAE

Goal of VAE is to model the distribution of observations $p(z)$ and generate new data by introducing latent random variables z. With the VAE, the posterior distribution is defined as $p(x) = \int p_\theta(z)p_\theta(x|z)dz$. Latent variable z is generated from a prior distribution $p(z)$. Φ and θ are parameters of the encoder and the decoder, respectively. Because the parameter θ and distribution for z are intractable, we can represent the marginal log-likelihood of an individual point as $\log p(x) = D_{KL}(q_\Phi(z|x)||p_\theta(z)) + L_{vae}(\Phi,\theta;x)$ notation from [39], where D_{KL} is Kullback-Leibler divergence from a prior $p_\theta(z)$ to the variational approximation $q_\theta(z|x)$ of $p(z|x)$ and L_{vae} is the variational lower bound of the data x by Jensen's inequality [38].

The VAE optimizes the parameters, Φ and θ, by maximizing the lower bound of the log likelihood, L_{vae},

$$L_{vae}(\Phi,\theta;x) = -D_{KL}(q_\Phi(zvertx)||p_\theta(z)) + E_{q_\Phi(z|x)}[\log p_\theta(x|z)] \quad (5)$$

The first term of Eq. (5) regularizes the latent variable z by minimizing the KL divergence between the approximated posterior and the prior of the latent variable. The second term of Eq. (5) is the reconstruction of x by maximizing the log likelihood $\log p_\theta(x|z)$ with sampling from $q_\Phi(z|x)$.

Anomaly detection through VAE is based on the probability of reconstruction. The high probability of reconstruction calculated through VAE means that the characteristics of the training data are similar, and the corresponding observation is data close to normal data. Conversely, if the reconstruction probability is low, it means that the characteristics are different from the training data, and the corresponding observation is data close to an anomaly data.

3.2 Lon Short-Term Memory

LSTM is a kind of recurrent neural network (RNN), capable of learning long-short term dependency in sequence data [40-42]. LSTM is designed to avoid the long-term dependency problem [43]. The structure of LSTM is a chain form of repeating a certain neural network (cell), which is same as RNN. The difference from RNN is that the each cell of the LSTM consists of three parts: forget gate, input gate and output gate. The structure of a LSTM cell is shown in Fig. 3.

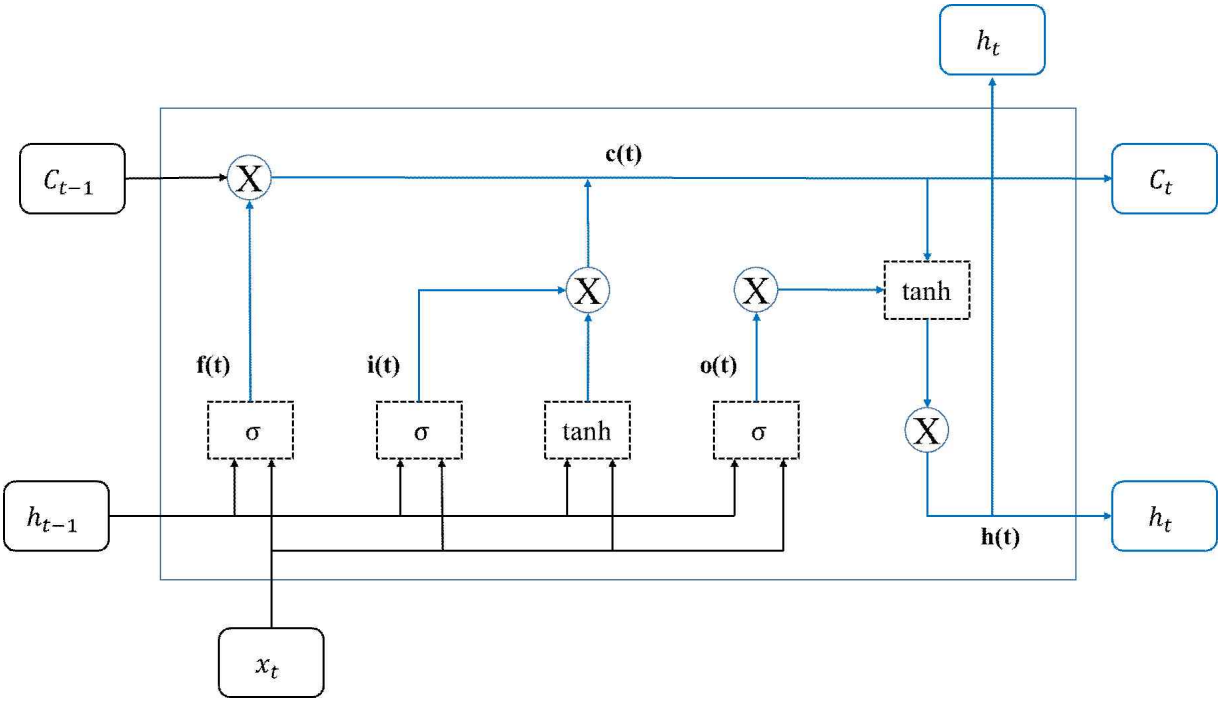

Fig. 3. The architecture of a LSTM cell

Eq. (6-9) describe the output from each gate unit in a LSTM cell:

$$i_t = \sigma(x_t W_{xi} + h_{t-1} W_{hi} + b_i) \tag{6}$$

$$f_t = \sigma(x_t W_{xf} + h_{t-1} W_{hf} + b_f) \tag{7}$$

$$o_t = \sigma(x_t W_{xo} + h_{t-1} W_{ho} + b_o) \tag{8}$$

$$c_t = f_t c_{t-1} + i_i \tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \tag{9}$$

where W is the weight matrix of each gate and b is the bias. The forget gate (f_t) reflects some of the previous cell state (C_{t-1}) for the cell state (C_t). It is remained or discarded according to the previous output and the present value. The input gate (i_t) modifies the value after the input data (x_t) has passed through the complete connection layer of tanh as an activation function. Finally, the input data (x_t) passes through the output gate. The output gate (o_t) considers past and modified input data, by adjusting the

input signal (x_t) to the tanh and making the output data. W_{xi}, W_{xf} and W_{xo} are respectively the weights between the input layer and the input gate, between the input layer and the forget gate, and between the input layer and the output gate. W_{hi}, W_{hf} and W_{ho} represent weights corresponding between each gate and hidden layer. W_{hi} is the weight between the hidden layer and the forget gate, W_{hf} is the weight between the hidden layer and the input gate, and W_{ho} is the weight between the hidden layer and the output gate. b_i, b_f and b_o are the additive biases of the input, forget and output gate, respectively [44].

4. Anomaly signal detection algorithm

An algorithm for the detection of the anomaly signals is proposed to detect anomaly signals using the difference between the input signal and the reconstructed signal for detection of the anomaly signals. The algorithm is divided into two steps: 1) Algorithm for the determination of the signal failures and 2) Optimization of Process. The algorithm for determining signal failure is the determination of signal failure. And the optimization of the process aims to compensate for the disadvantage of unsupervised learning and to improve the performance of the algorithm. And it uses a VAE model that produces an output similar to an input and an LSTM model for nonlinear data.

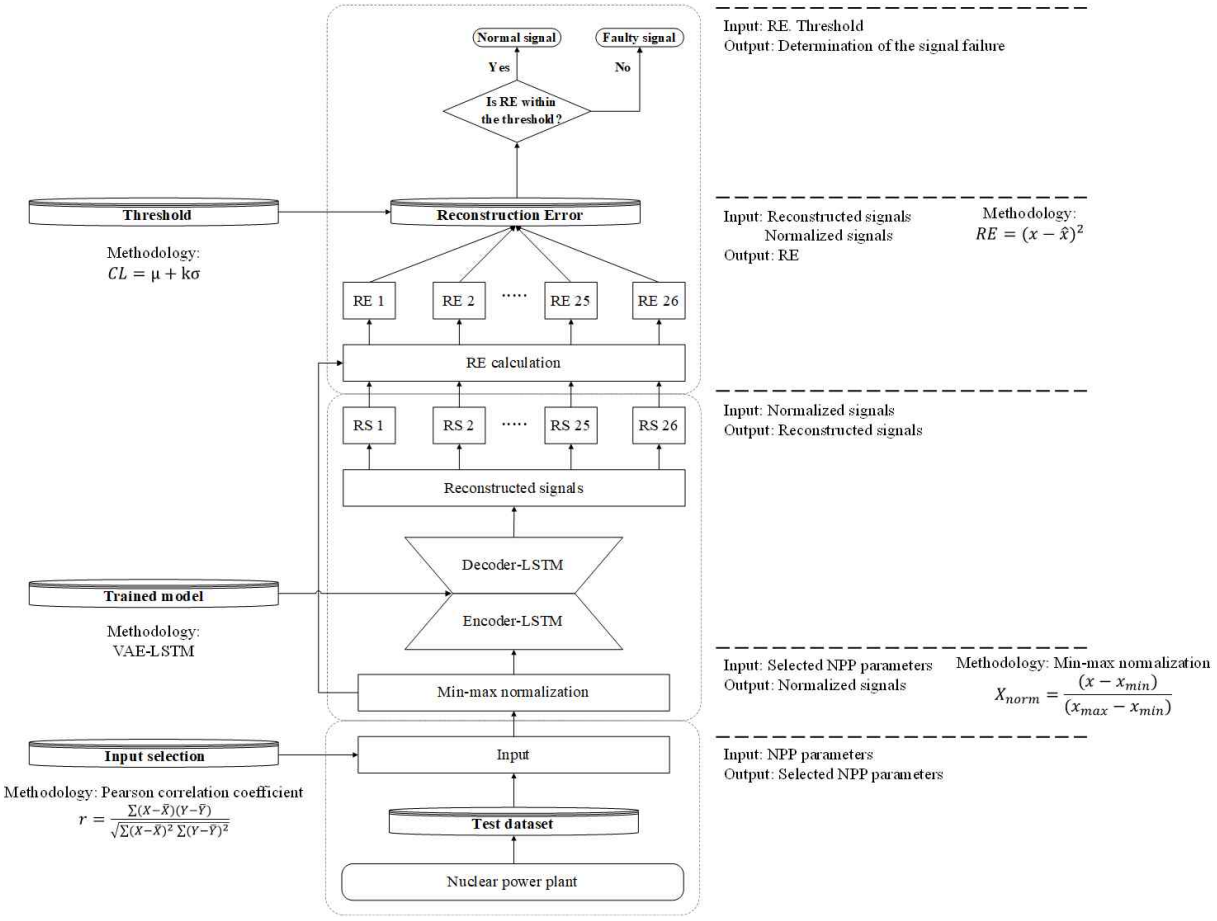

Fig. 4. The overview of the algorithm for the detection of the anomaly signals

And, Optimization of Process supports optimization to increase the failure detection rate of the signal failure determination algorithm. It consists of three steps: Determination of Input Selection, Determination of Hyper-parameters, and Determination of Threshold. Fig. 4 shows the structure of the suggested algorithm.

4.1 Training environment

Since there are few cases of emergency at actual NPPs, it is very difficult to collect data for algorithm training. Therefore, data obtained from the simulation were used to perform the process. The data used for training and validation of the algorithm were collected using a compact nuclear simulator (CNS) developed by the Korea Atomic Energy Research Institute (KAERI) [45]. The source plant of CNS is Westinghouse 3 loop 900MW Pressurized Water Reactor (PWR). Fig. 5 is the reactor coolant system (RCS) interface of CNS.

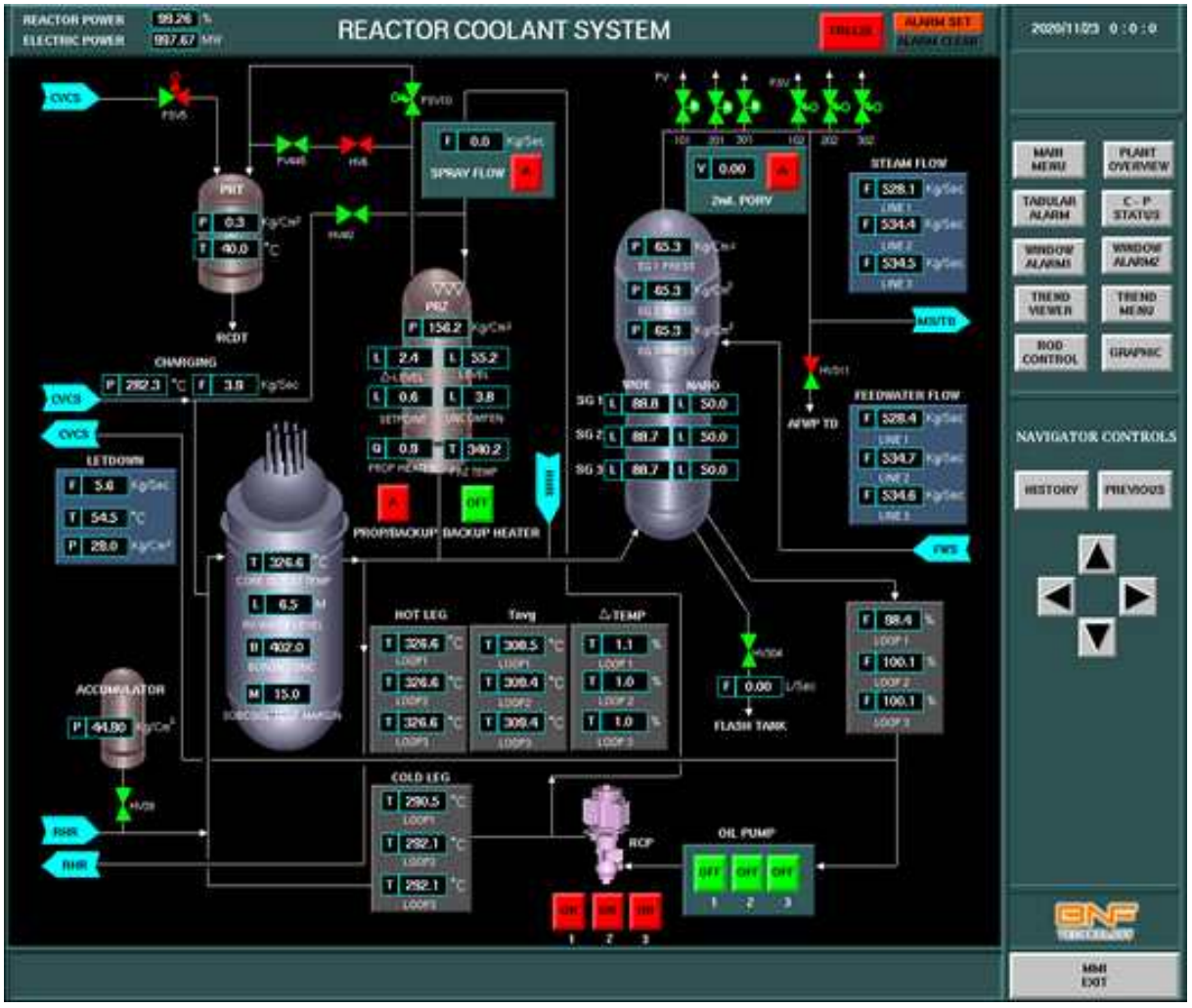

Fig. 5. Reactor Coolant System in CNS

To implement the anomaly signal detection algorithm, a desktop computer with the following hardware configurations is used: NVIDIA GeForce RTX 2080 Ti 11 GB GPU and Intel Core i7-8700 CPU at 3.20 GHz. The VAE-LSTM model was developed based on the Python programming language with the Keras machine learning libraries.

4.2 Algorithm for the determination of the signal failure

This is the step to determine signal failure using the data (i.e. input selection, model hyper-parameters, threshold) determined by the optimization module of the process. It consists of three steps: Input Pre-processing, Signal Reconstruction, Determination of Signal

Failures. Fig. 6 shows the structure of the determination of the signal failures step.

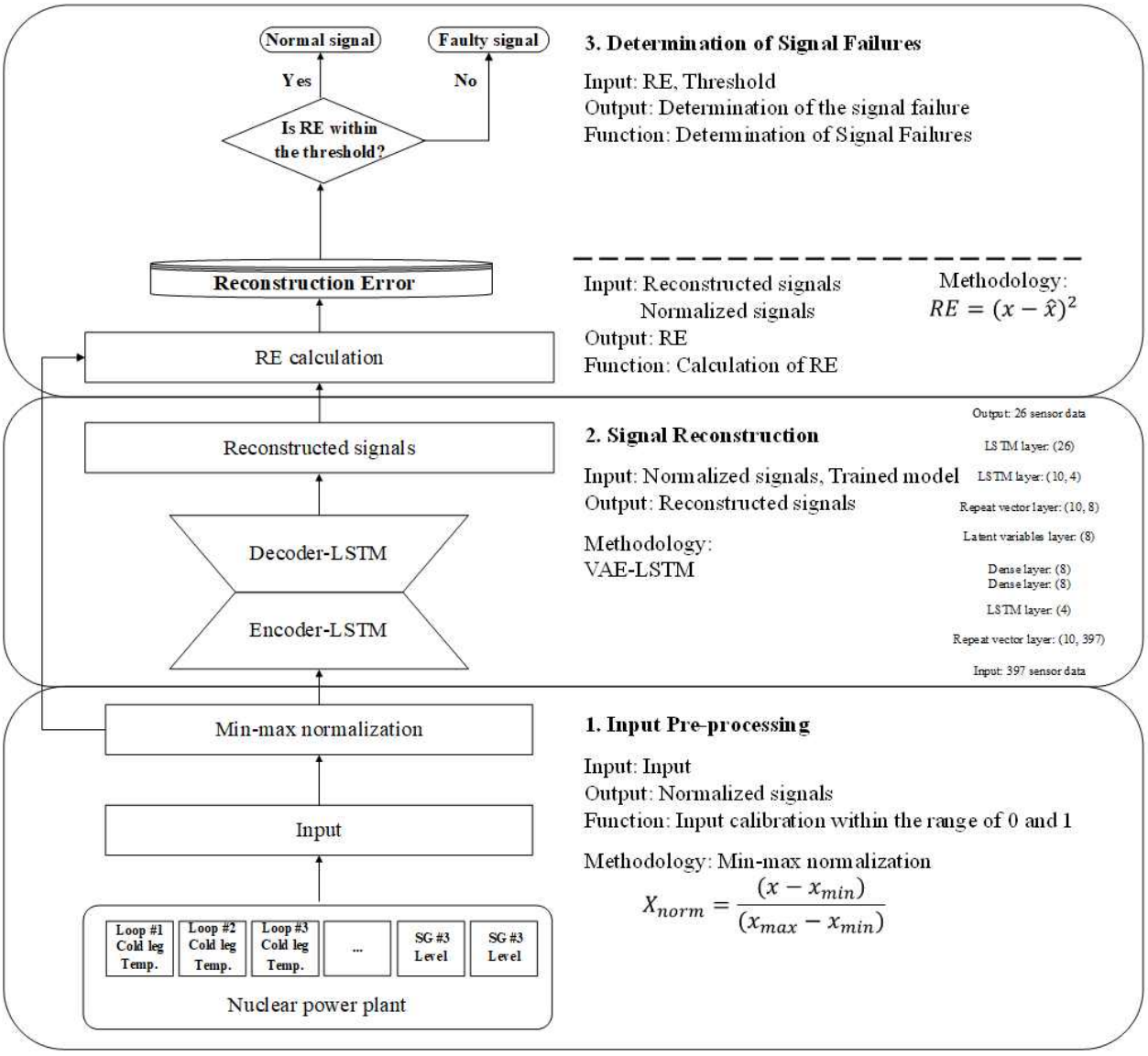

Fig. 6. The structure of the determination of the signal failures step

In the Input Pre-processing step, normalization of all input values was performed to improve the performance by converting the input values determined at Optimal of process module. As the signal values have different scales, normalization can prevent convergence at the local maximum and minimum. The min-max normalization method is applied and maximum and minimum values are determined from the training data and the input is

calibrated within the range of 0 to 1 through Eq. (10).

$$X_{norm} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} \tag{10}$$

The Signal Reconstruction step is the step of using the VAE-LSTM model and reconstructing the signal about the input value after applying the hyper-parameters defined in the determina-tion of hyper-parameters step of an Optimization module. In this step, VAE, a representative production model of unsupervised learning, was used to produce output similar to input, and LSTM, which can process time-series data, was used because the NPPs produces time-series data. Besides, if the output is produced very similar to the input, signal reconstruction is improved, allowing even minor errors to detect faults quickly and easily.

The Determination of Signal Failure step is the step of comparing the calculated Reconstruction Error (RE) with each predefined threshold for the target signal. This step consists of a calculation module for reconstruction errors and a determination module of signal failure. The RE calculation module calculates RE by comparing the normalized input signal with the reconstruction signal. RE is Eq. (11).

$$Reconstruction\ error = (x - \hat{x})^2 \tag{11}$$

Where x denotes normalized input signal, and $\hat{x}$ denotes the reconstruction signal. The better the reconstruction ability of the model, the lower the RE value. And, the Determination module of Signal Failures determines signal failures by comparing the respective threshold and REs for predefined target signals in the optimization module. At this time, if the RE exceeds the threshold, it will be determined as a signal failure.

4.3 Optimization of process

Optimization of Process is a step to improve the optimization and failure detection

performance of the proposed algorithm for the detection of signal failures. It consists of three steps: Determination of Input Selection, Determination of Hyper-parameters, and Determination of Threshold. Determination of Input Selection and the Determination of Hyper-parameters are steps to increase the number of inputs to the target signal and to experimentally determine the optimal hyper-parameters for the reconstruction model to compensate for the disadvantages of unsupervised learning. And, Determination of Threshold is a step in determining each threshold for the target signal to enhance the algorithm's ability to detect failures. When this Optimization of Process is complete, the determined data (i.e., determined input, determined hyper-parameters, determined thresholds) is fixed and applied Algorithm for the determination of the signal failures.

4.3.1 Determination of input selection

This step aims to find the optimal input variables in designing the model. The input selection step selects inputs through correlation analysis between the selected target signal and the CNS total parameters (A total of 2200). In this study, target signals are 26 signals that indicate the main process parameters in the Loss of Coolant Accident (LOCA) scenario, as shown in Table 2. This step is to compensate for the difficulty of learning unsupervised, and Pearson correlation analysis was used as the correlation analysis method.

Table 2. Selected target signal from the CNS

Parameter	Units
FEEDWATER PUMP OUTLET PRESS	kg/cm^2
FEEDWATER LINE 1, 2, 3 FLOW	kg/sec
FEEDWATER TEMP	℃
MAIN STEAM FLOW	kg/sec
STEAM LINE 1, 2, 3 FLOW	kg/sec
MAIN STEAM HEADER PRESSURE	kg/cm^2
CHARGING LINE OUTLET TEMPERATURE	℃
LOOP 1, 2, 3 COLDLEG TEMPERATURE	℃
PRZ TEMPERATURE	℃
CORE OUTLET TEMPERATURE	℃
NET LETDOWN FLOW	kg/sec
PRZ LEVEL	%
PRZ PRESSURE(WIDE RANGE)	kg/cm^2
LOOP 1, 2, 3 FLOW	kg/sec
SG 1, 3 LEVEL(WIDE)	%
SG 1, 3 PRESSURE	kg/cm^2

*PRZ: Pressurizer

The Pearson correlation coefficient analysis is a linear correlation coefficient, which is used to reflect the linear correlation of two normal continuous variables [46]. The Pearson correlation coefficient analysis is one of the most widely used relationship measures. The analysis of two variables and is defined as the covariance of the two variables divided by the product of standard deviations and it can be defined by:

$$r = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum (X_i - \overline{X})^2 \sum (Y_i - \overline{Y})^2}} \tag{12}$$

$$* \overline{X} = \frac{1}{n} \sum_{i=1}^{N} X_i \ denotes \ the \ mean \ of \ X$$

$$* \overline{Y} = \frac{1}{n} \sum_{i=1}^{N} Y_i \ denotes \ the \ mean \ of \ Y$$

Pearson's value is between [−1, 1]. When equals 1, it becomes a completely positive correlation. When it equals −1, it becomes a completely negative correlation. When it equals 0, the linear correlation between and is not obvious. It means that the greater the absolute value of the Pearson's , the stronger the correlation.

Fig. 7 and 8 are the results of Pearson's correlation of the pressurizer water level and core temperature for all CNS variables, and r is more than 0.995 in the calculation results. As shown in Fig. 7 and 8, each calculation result represents different values, so all the results for the target signal will be output differently. Therefore, the input was selected as the union of the results.

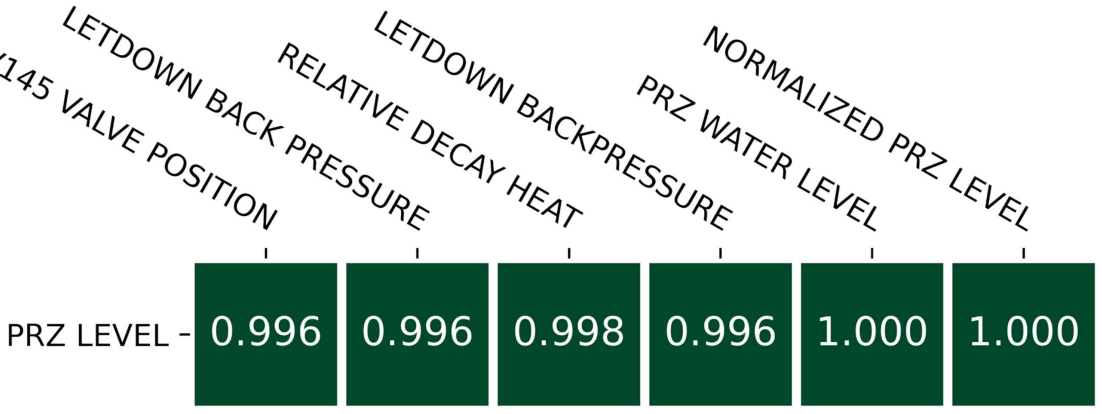

Fig. 7. Pearson's calculation result of PRZ level

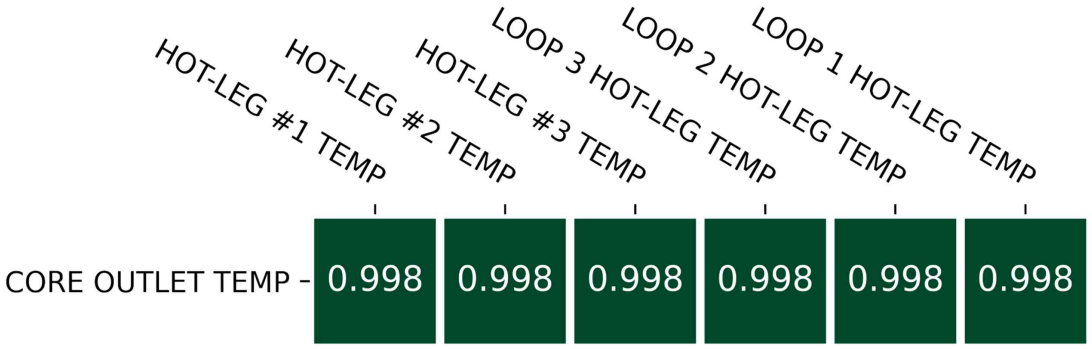

Fig. 8. Pearson's calculation result of core outlet temperature

Table 3 shows the number of selected inputs and the performance of the model according to the selected inputs when r is 0.995, 0.985, and 0.975. The reconstruction rate was defined by learning the VAE-LSTM model based on the selected input and confirming the reconstruction value for the normal signal. When r is 0.995, 157 inputs were selected, but the model performance according to the selected inputs did not show good results. And, the performance of the model was checked for the case where r is 0.975 and 0.985, and finally 157 inputs when r is 0.985 were selected. The performance check of the model used here is detailed in the next subsection.

Table 3. Number of inputs according to Pearson's r

r	# of Input	Target signal reconstruction rate (%)
0.995	157	94.2%
0.975	604	97.5%
0.985	397	99.8%

4.3.2 Determination of hyper-parameter

The second step is to determine the optimal hyper-parameter of the VAE-LSTM based on the input selected in the previous step. Data collection was done in advance before determining the model's hyper-parameters. In this study, the dataset for determination and

validation of hyper-parameters was focused on the LOCA scenario of an emergency situation. For each scenario, simulations were performed for cold-leg and hot-leg by varying the fracture size in 5 from 10 to 50 . The simulation was performed for about 15 minutes after the reactor shutdown. A total of 80,697 data were collected for 54 scenarios in the LOCA, as shown in Table 4. Ninety percent (90%) and ten percent (10%) of collected data are used for training and testing, respectively. The actual signal value of NPPs can include noise, but the dataset through CNS is noiseless data, so 5% of Gaussian white noise is arbitrarily injected.

Table 4. The database used for algorithm training and validation

Initiating situation	Number of scenarios	Number of training sets
Cold/hot leg loss of coolant accident (LOCA)	54	80,697

To select the optimal hyper-parameter of the model, the VAE-LSTM model was trained by changing the conditions as shown in table 5 below. Training was carried out with the aim of minimizing RE. The training was performed with 300 epochs each, and the best performance was achieved when the batch was 32 and the LSTM node was 4 and the latent node was 8. Fig. 9 shows the results of the model to which the selected hyper-parameter is applied. The blue line is the CNS data injected with Gaussian noise, and the red line is the signal reconstructed by the model. Even if the noise was injected through the figure, it was confirmed that the model reconstructs signal with good performance.

Table 5. Hyper-parameter

Num	Batch	LSTM node	Latent node	RE
1	32	16	8	1.748E-3
2	32	8	8	1.961E-3
3	32	8	4	1.251E-3
4	32	4	8	**1.074E-3**
5	64	16	8	2.259E-3
6	64	8	8	4.392E-3
7	64	8	4	1.319E-3
8	64	4	8	1.310E-3

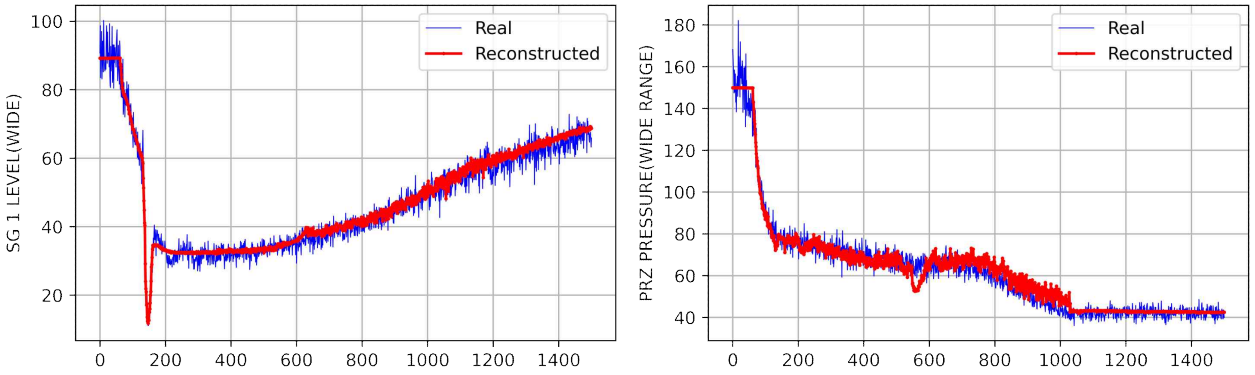

Fig. 9. Reconstruction result of selected model

4.3.3 Determination of threshold

The Detection of Threshold is the step in determining each optimal threshold for target signals. This is calculated using the residuals between normal and reconstructed data, it is to improve the algorithm's ability to detect anomaly signals. However, there are two issues with determining thresholds. First, if the threshold is calculated high, the signal failure can

be determined to be a normal signal. Second, if the threshold is calculated low, the normal signal can be determines to be a failure signal. For these reasons, it is important to determine the optimal threshold.

The threshold is calculated by the method suggested by Shewhart [47]. The threshold value is calculated as follows:

$$CL = \mu + k\sigma \tag{13}$$

the μ and σ refer to the mean value and standard deviation of reconstruction errors. The CL is used as the criteria to determine the signal failure. And, k is constant.

To determine the optimal threshold, the threshold value for the target signal was calculated by setting the k value of the Eq. (13). at 1, 3 and a single defect was injected to compare the failure detection rate according to the k value. From this, each failure detection ratio was calculated for the target signal with a k value of 1 and 3. Each failure detection ratio for this is shown in Table 6.

Table 6. The ratio of each failure detection to the target signal when k values are 1 and 3

Parameter	Signal failure detection probability when k = 1 (%)	Signal failure detection probability when k = 3 (%)
FEEDWATER PUMP OUTLET PRESS	100	100
FEEDWATER LINE 1, 2, 3 FLOW	100	100
FEEDWATER TEMP	100	89.51
MAIN STEAM FLOW	100	98.15
STEAM LINE 1, 2, 3 FLOW	100	100
MAIN STEAM HEADER PRESSURE	100	100
CHARGING LINE OUTLET TEMPERATURE	98.77	72.84
LOOP 1, 2, 3 COLDLEG TEMPERATURE.	98.35	84.57
PRZ TEMPERATURE	100	90.74
CORE OUTLET TEMPERATURE	92.59	79.01
NET LETDOWN FLOW	100	100
PRZ LEVEL	87.04	72.22
PRZ PRESSURE(WIDE RANGE)	89.51	35.19
LOOP 1, 2, 3 FLOW	100	100
SG 1, 3 LEVEL(WIDE)	100	94.75
SG 1, 3 PRESSURE	90.74	66.98
Total	97.86%	89.03 %

As shown in Table 6, when k is 1 it detects the failure better than when k is 3. Therefore, in this step, the threshold value for the target signal was determined so that more failures could be detected by setting the k value to 1, and through this, the failure detection performance of the anomaly signal detection algorithm was improved.

5. Validation

In this section, we verify the proposed algorithm to detect the anomaly signals. The data used for verification totaled 80,697 data and was collected for 54 scenarios in LOCA accident. Additionally, verification data were collected using CNS, and Gaussian noise (1, 0.05) was injected to imitate actual NPPs data. The target signals are 26 pre-selected, and the respectively signal fault is injected for 54 scenarios to verify the fault detection performance. A stuck fault was used as a type of single fault, which is classified as stuck at the higher value, stuck at the lower value, and stuck as is the current value.

5.1 Process of validation for the proposed algorithm

As shown Table 4, to verify the anomaly signals detection algorithm in an emergency situation, data on emergency situations for LOCA accident were collected using CNS. Also, because the actual signal value of the NPPs is the value that contains noise, the data collected to imitate it was injected with Gaussian noise (1, 0.05). The anomaly signal detection algorithm uses the data (i.e., input values, hyper-parameters for the model, and each threshold for the target signal) determined in the Optimal of Process. Fig. 10 shows a signal reconstruction using the data determined in the Optimal of Process for one of the 54 LOCA scenarios. From the Fig. 10, it can be seen that the trained VAE model has undergone denoising due to structural features.

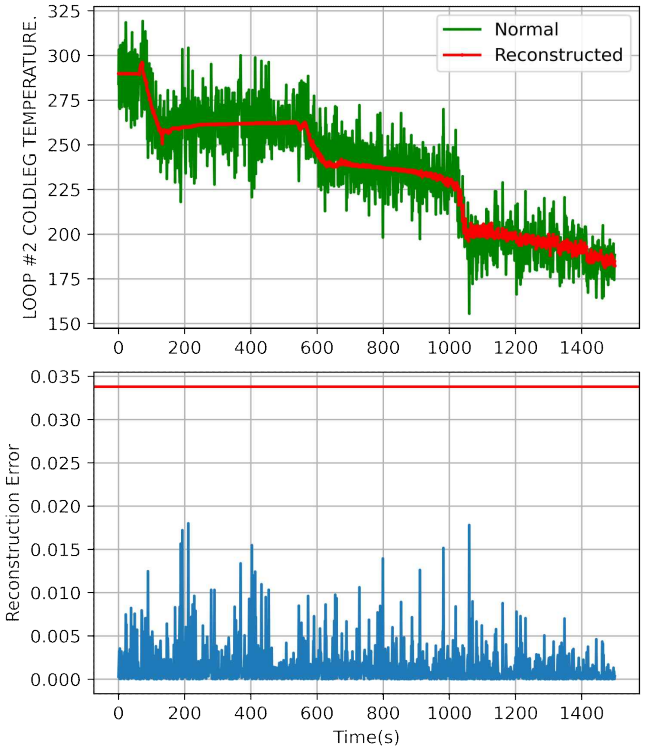

Fig. 10. Reconstruction result of Loop 2 cold leg temperature

The type of failure applied to the verification is a stuck fault, and according to the target signal, it is divided into stuck at a high value, stuck at a low value, and stuck as is current.

In the verification sequence, 1) injects respectively target signals fault for 54 scenarios. 2) enter the failure-injected data into the optimized VAE-LSTM model to calculate the RE using the difference between the reconstructed signal and the input signal. 3) compare the calculated RE with the optimal threshold determined for the target signal. At this time, if the RE exceeds the threshold, the algorithm determines that the signal is faulty.

5.1 Validation result

After preparing for the verification of the proposed algorithm, an experiment was conducted to demonstrate the extent to which the proposed algorithm detects anomaly signals. Each failure of 26 target signals used for verification was 80,697 failures for 54 scenarios in LOCA accidents.

Fig. 11 and Fig. 12 are types of high and low value faults at PRZ temperature, 300 seconds later the fault signal was injected. At this time, we can confirm that RE exceeds the threshold, and the proposed algorithm determines the signal for PRZ temperature as a failure signal. Other signals that have not been injected with the defect are on the right side of Fig. 11 and Fig. 12.

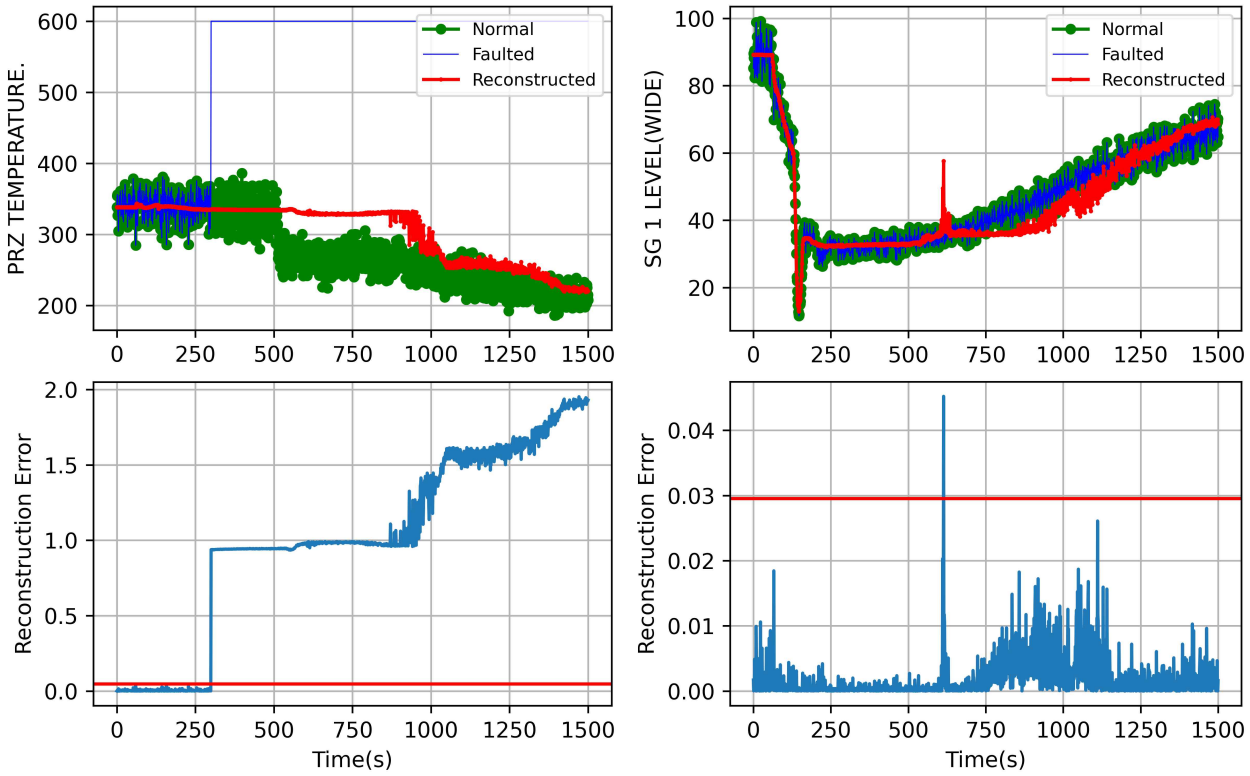

Fig. 11. PRZ temperature stuck at a high value

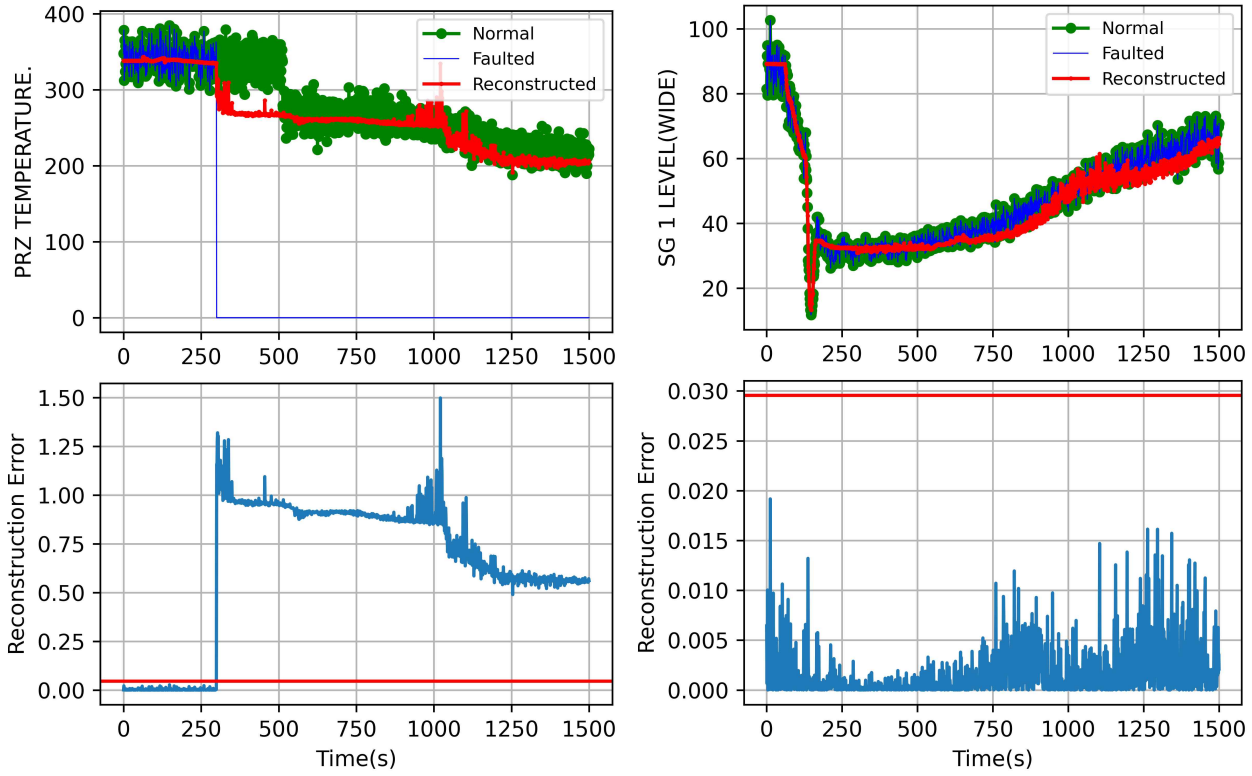

Fig. 12. PRZ temperature stuck at a low value

And, Fig. 13 is the type for failure, such as the fail as is the current value for PRZ temperature, and after 300 seconds the fault signal was injected. At this time, we can confirm that RE exceeds the threshold, and the proposed algorithm determines the signal for PRZ temperature as a failure signal. Other signals that did not inject defects are on the right side of Figure 13.

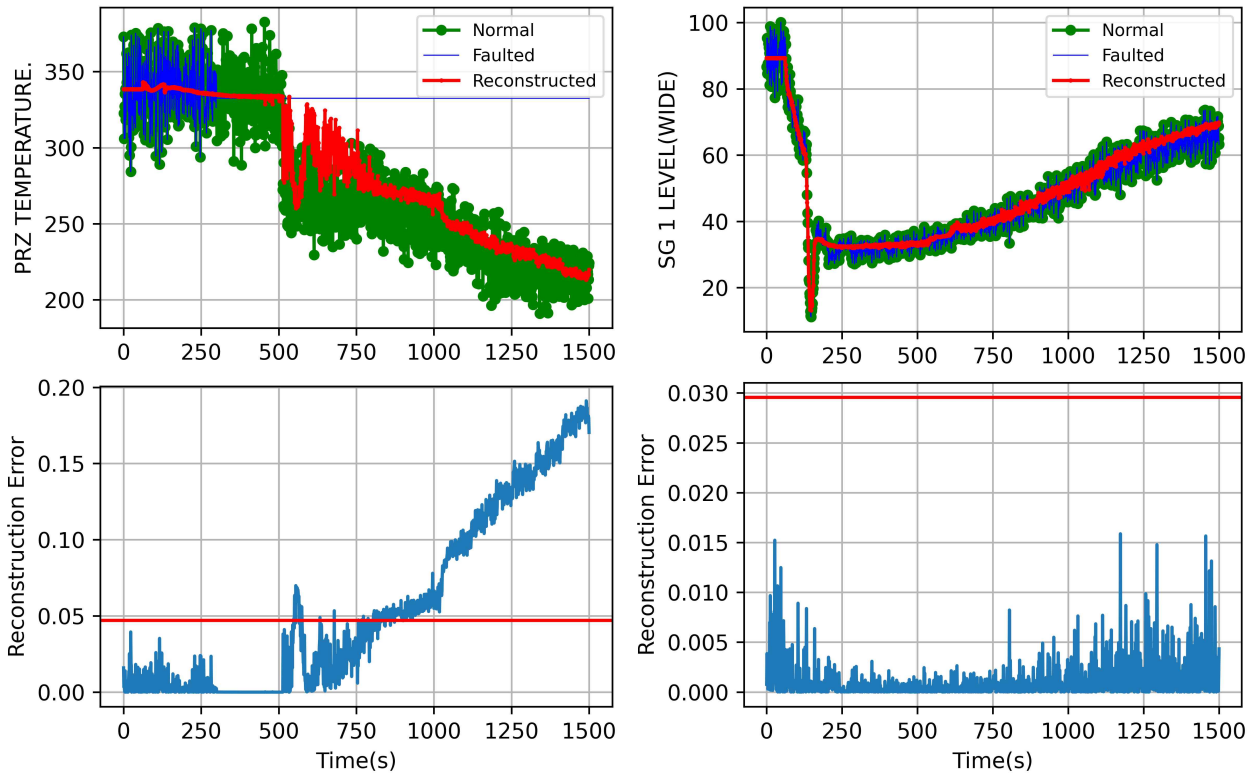

Fig. 13. PRZ temperature stuck as is current.

The performance of the proposed anomaly signal detection algorithm is shown in Table 7.

Table 7. Anomaly signals detection ratio

Unit: %		Output	
		Failed	Normal
Input	Failed	97.39 (3366/3456)	2.61 (90/3456)
	Normal	0.19 (158/84240)	99.81 (84082/84240)

The anomaly detection ratio is calculated as:

$$Failed-Failed=\frac{Number\ of\ failures\ detected\ by\ the\ algorithm}{Number\ of\ posiible\ failure\ cases} \tag{14}$$

$$Normal-Normal=\frac{Number\ of\ normals\ detected\ by\ the\ algorithm}{Number\ of\ posiible\ normal\ cases} \tag{15}$$

Through experiments, it was confirmed that the probability of detecting a fault was 97.39% and that of detecting a normal was 99.81%. Although the proposed model deals with one emergency scenario (i.e., LOCA), it has shown its applicability to a real NPPs considering its high accuracy.

6. Conclusion

In this study, an algorithm for the detection of the anomaly signals is proposed. Algorithms are divided into two steps: 1) Algorithms for determining signal failures and 2) optimizing processes. The algorithm for the determination of signal failure is the determination of signal failure and, the optimization of processes is to improve the performance of algorithms. To compensate for the cons of unsupervised learning during the optimization of the process, the input of the model was determined using Pearson correlation analysis, and the model was trained by obtaining optimal hyper-parameters from the determined inputs. Besides, each threshold for the target signal was determined experimentally to improve the algorithm's failure detection performance. The algorithm is proposed in an emergency situation, so it needs to consider thousands of signals and various types of failures in NPPs. For this reason, the proposed algorithms used the VAE model, which is a representative production model of unsupervised learning, and the LSTM model, which processes time-series data. In the validation, the single fault was injected for the target signal value to verify the anomaly signals detection ability. As a result, the proposed algorithm showed a 97.39% failure detection ability.

Reference

[1] H. Basher, J. S. Neal, and L. L. C. UT-Battelle, Autonomous Control of Nuclear Power Plants, United States, Department of Energy, 2003.

[2] M. Weightman, The Great East Japan Earthquake Expert Mission. IAEA International Fact Finding Expert Mission of the Fukushima Dai-ichi NPP Accident Following the Great East Japan Earthquake and Tsunami, Mission Report IAEA, 2011.

[3] F. Daiichi, Ans committee report. A Report by The American Nuclear Society Special Committee on Fukushima, 2012.

[4] J. E. Yang, Fukushima Dai-Ichi accident: lessons learned and future actions from the risk perspectives. Nuclear Engineering and Technology, 46(1), pp.27-38, 2014.

[5] Younhee Choi, Gyeongmin Yoon, Subong Lee , Jonghyun Kim, Algorithm for the Detection of Sensor Failure in the Emergency Situation Using VAE-LSTM, kns, 2020.

[6] Foo, G. H. B., Zhang, X., Vilathgamuwa, D. M., A sensor fault detection and isolation method in interior permanent-magnet synchronous motor drives based on an extended Kalman filter, IEEE Transactions on Industrial Electronics, pp.3485-3495, 2013.

[7] P. Sundvall, P. Jensfelt,and B. Wahlberg, Fault detection using redundant navigation modules. IFAC Proceedings Volumes, 39(13), pp.522-527, 2006.

[8] Chow, E. Y. E. Y., and Alan Willsky, Analytical redundancy and the design of robust failure detection systems, IEEE Transactions on automatic control , pp.603-614, 1984.

[9] Gertler, J.J., Singer, D., 1990. A new structural framework for parity equation-based failure detection and isolation. Automatica 26, pp.381-388.

[10] Gertler, J.J., Fault detection and isolation using parity relations. Control Engineering Practice , pp.653-661, 1997.

[11] R. Isermann, Fault diagnosis of machines via parameter estimation and knowledge processing-tutorial paper, Automatica, 29(4), 815-835, 1993.

[12] T. Jiang, K. Khorasani, S. Tafazoli, Parameter estimation-based fault detection, isolation and recovery for nonlinear satellite models, IEEE Transactions on control systems technology, 16(4), 799-808, 2008.

[13] Z. Li, Y. Wang, A. Yang and H. Yang, Drift detection and calibration of sensor networks. In 2015 International Conference on Wireless Communications & Signal Processing (WCSP), pp.1-6, 2015.

[14] G. Yun-hong, Z. Ding, L. Yi-bo, Small UAV sensor fault detection and signal reconstruction, International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC), IEEE, pp. 3055-3058, 2013.

[15] J. Yu, J. Jang, J. Yoo, J. H. Park, S. Kim, Bagged auto-associative kernel regression-based fault detection and identification approach for steam boilers in thermal power plants, Journal of Electrical Engineering & Technology, 12(4), 1406-1416, 2017.

[16] F. Di Maio, P. Baraldi, E. Zio, R. Seraoui, Fault detection in nuclear power plants components by a combination of statistical methods, IEEE Transactions on Reliability, pp.833-845, 2013.

[17] M. G. Na, Y. R. Sim, K. H. Park, B. R. Upadhyaya, B. Lu, and K. Zhao, Failure Detection using a Fuzzy Neural Network with an Automatic Input Selection Algorithm, in Power Plant Surveillance and Diagnostics, D. Ruan and P. F. Fantoni, Eds. Vienna, Austria: Springer- Verlag, vol. 14, pp.221-242, May 2002.

[18] Xu, Xiao, J. Wesley Hines, and Robert E. Uhrig, Sensor validation and fault detection using neural networks, Proc. Maintenance and Reliability Conference (MARCON 99), pp.10-12, 1999.

[19] W. Li, M. Peng, Y. Liu, N. Jiang, H. Wang, Z. Duan, Fault detection, identification and reconstruction of sensors in nuclear power plant with optimized PCA method, Annals of Nuclear Energy, 113, 105-117, 2018.

[20] A. S. Heger, K. H. Holbert, and A. M. Ishaque , Fuzzy Associative Memories for Instrument Fault Detection, Annals of Nuclear Energy, pp.739-756 , 1996.

[21] N. Kaistha, B. R. Upadhyaya], Incipient fault detection and isolation of field devices in nuclear power systems using principal component analysis, Nuclear technology, 136(2), 221-230, 2001

[22] E. Eryurek, B. R. Upadhyaya, Sensor Validation for Power Plants using Adaptive Backpropagation Neural Network, IEEE Transactions on Nuclear. Science, 37(2),

pp.1040-1047, 1990.

[23] J. W. Hines, R. E. Uhrig, D. J. Wrest, Use of autoassociative neural networks for signal validation, Journal of Intelligent and Robotic Systems, 21(2), 143-154, 1998.

[24] P. F. Fantoni, A. Mazzola, Multiple-failure signal validation in nuclear power plants using artificial neural networks. Nuclear technology, 113(3), 368-374, 1996.

[25] J. H. Choi, S. J. Lee, Consistency Index-Based Sensor Fault Detection System for Nuclear Power Plant Emergency Situations Using an LSTM Network, Sensors, Vol.20, no.6, 2020.

[26] P. F. Fantoni and A. Mazzola, A Pattern Recognition-Artificial Neural Networks Based Model for Signal Validation in Nuclear Power Plants, Annals of in Nuclear Power Plants, Annals of Nuclear Energy, pp.1069-1076, 1996.

[27] S. G. Kim, Y. H. Chae, P. H. Seong, Signal fault identification in nuclear power plants based on deep neural networks. Annals of DAAAM & Proceedings, 2019.

[28] P. Cunningham, M. Cord, S. J. Delany, Supervised learning, In Machine learning techniques for multimedia (pp. 21-49), Springer, Berlin, Heidelberg, 2008.

[29] P. Dayan, M. Sahani, G. Deback, Unsupervised learning, The MIT encyclopedia of the cognitive sciences, 857-859, 1999.

[30] N. Mehranbod, M. Soroush, and C. Panjapornpon, A method of sensor fault detection and identification. Journal of Process Control, p.321-339, 2005.

[31] J. D. Bošković, and R. K. Mehra, Failure detection, identification and reconfiguration in flight control. In Fault Diagnosis and Fault Tolerance for Mechatronic Systems: Recent Advances, Springer, Berlin, Heidelberg, pp.129-167, 2003.

[32] B. Pan, Bias error reduction of digital image correlation using Gaussian pre-filtering. Optics and Lasers in Engineering, 51(10), pp.1161-1167, 2013.

[33] B. Ergöçmen, Fault tolerant flight control applications for a fixed wing UAV using linear and nonlinear approaches, MIDDLE EAST TECHNICAL UNIVERSITY, 2019.

[34] R. Dorr, F. Kratz, J. Ragot, F. Loisy, J. L. Germain, Detection, isolation, and identification of sensor faults in nuclear power plants. IEEE Transactions on Control Systems Technology, 5(1), pp.42-60, 1997.

[35] D. Roverso, J. E. Farbrot, F. S. Gran, and M. Hoffmann, Condition monitoring and diagnostic systems integration-guiding principles and practical examples. No. HWR--744. Institutt for energiteknikk, 2005.

[36] S. Kiss and S. Lipcseia, Combined validation of noise signals measured in NPPs. Annals of Nuclear Energy, 26(4), pp.327-346, 1999.

[37] D. Park, Y. Hoshi, C. C. Kemp, A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder, IEEE Robotics and Automation Letters, 3(3), pp.1544-1551, 2018.

[38] J. An, S. Cho, Variational autoencoder based anomaly detection using reconstruction probability, Special Lecture on IE, 2(1), pp.1-18, 2015.

[39] D. P. Kingma, M. Welling, Auto-encoding variational Bayes, arXiv preprint arXiv:1312.6114v10, 2014.

[40] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation, 9(8), pp.1735-1780, 1997.

[41] J. Yang, J. Kim, Accident diagnosis algorithm with untrained accident identification during power-increasing operation, Reliability Engineering & System Safety, vol.202, 2020.

[42] J. Yang, J. Kim, An accident diagnosis algorithm using long short-term memory, Nuclear Engineering and Technology, 50(4), pp.582-588, 2018.

[43] F. A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: Continual prediction with LSTM, 1999.

[44] D. Lee, A. M. Arigi, J. Kim, Algorithm for Autonomous Power-increase Operation Using Deep Reinforcement Learning and a Rule-Based System. IEEE Access, 2020.

[45] KAERI, "Advanced compact nuclear simulator textbook," KAERI, Nucl. Training Center Korea At. Energy Res. Inst., Daejeon, South Korea, Tech. Rep., 1990.

[46] H. Xu, Y. Deng, Dependent evidence combination based on shearman coefficient and pearson coefficient, IEEE Access, 6, 11634-11640, 2017.

[47] W.A. Shewhart, Economic Control of Quality of Manufactured Product, New York: Van Nostrand, 1931.