2023년 2월

석사학위 논문

Signal Anomaly Detection Algorithm Using Deep Learning in NPPs

조선대학교 대학원

원 자 력 공 학 과

최 윤 희

Signal Anomaly Detection Algorithm Using Deep Learning in NPPs

딥러닝 기법을 사용한 원자력발전소에서의 이상 신호 탐지 알고리즘

2023년 2월 24일

조선대학교 대학원
원 자 력 공 학 과
최 윤 희

Signal Anomaly Detection Algorithm Using Deep Learning in NPPs

지도교수 김 종 현

이 논문을 공학 석사학위신청 논문으로 제출함

2022년 10월

조선대학교 대학원

원 자 력 공 학 과

최 윤 희

최윤희의 석사학위논문을 인준함

위원장	조선대학교 교수	나 만 균	(인)
위 원	조선대학교 교수	공 태 영	(인)
위 원	조선대학교 교수	김 종 현	(인)

2022년 12월

조선대학교 대학원

Table of Contents

List of Tables

List of Figures

초 록

딥러닝 기법을 사용한 원자력 발전소에서의 이상 신호 탐지 알고리즘

최 윤 희
지도 교수 : 김 종 현
원자력공학과
조선대학교 대학원

원자력 발전소의 센서는 현재 발전소의 상태 및 상황을 운전원과 제어 시스템에게 전달하는 역할을 하고 있기에 원자력 발전소의 안전한 운영을 위해서는 신호가 매우 중요하다. TMI 사고와 후쿠시마 사고 때 알 수 있듯이 비상상황에서의 잘못된 신호는 제어시스템과 운전원에게 혼란을 초래하고 이는 사고로 이어질 수 있다. 또한, 자율 및 자동 제어에 대한 관심이 높아지면서 신뢰할 수 있는 신호의 중요성이 높아졌다. 본 논문은 원자력 발전소에서의 신호가 급격히 변화하는 비상 상황에서의 이상 신호 탐지를 위한 알고리즘을 제안한다. 알고리즘은 딥러닝 기법의 한 종류인 Variational Autoencoder (VAE)와 Long Short-Term Memory (LSTM)을 기반으로 한다. 또한, 알고리즘은 3개의 최적화 단계를 통해 최적화된다. 최적화 단계는 1) 최적의 입력값 선택, 2) 매개변수 선택, 3) 문턱값 선정 등으로 구성된다. 마지막으로, 제안된 알고리즘은 Compact Nuclear Simulator (CNS)를 통해 검증된다. 검증 결과 제안된 알고리즘은 비상상황에서의 신호 고장을 97% 이상 검출한다.

Abstract

Signal Anomaly Detection Algorithm Using Deep Learning in NPPs

Younhee Choi
Advisor : Prof. Jonghyun Kim , Ph.D.
Department of Nuclear Engineering
Graduate School of Chosun University

The validity and correctness of signals are critical to the safe operation of nuclear power plants (NPPs). Faulty signals as well as sensors may degrade the performances of both control systems and operators under the emergency situations, as learned from the past accidents in NPPs. Moreover, increasing interest in autonomous and automatic controls also highlights the importance of reliable signals because successful controls largely rely on the integrity of input signals. This paper aims to propose an algorithm for the signal anomaly detection in emergency situations in which signals are dramatically changing over time in NPPs. The algorithm is based on a combination of Variational Auto-Encoder (VAE) and Long Short-Term Memory (LSTM) that employs unsupervised learning. The optimization of algorithm is also conducted for selecting inputs, determining hyper-parameters of the network, and determining thresholds to identify signal failures. Lastly, the proposed algorithm is validated by using the Compact Nuclear Simulator (CNS). The result presents that the suggested algorithm could detect more than 97% of the status of signals successfully in the emergency situations.

I. Introduction

To operate nuclear power plants (NPPs) safely and efficiently, signals from sensors to operators must be valid and correct. Faulty signals as well as sensors may impair the performances of control systems as well as of operators. This may consequently lead to icky situations that compromise the safety of NPPs [1]. In particular, operator's misjudgment under the emergency situation resulting from faulty signals could be a main contributor to a severe accident, as learned from past experiences such as Three Mile Island and Fukushima Daiichi NPP accidents [2-6]. Moreover, recent interest in autonomous or automatic controls is improving and then the reliability of signal becomes more important for the successful operation because signals are inputs to those control systems.

For this reason, many researches have been performed for developing techniques for the signal anomaly detection [7-30]. Those approaches can be classified into model-based and data-driven approaches. Model-based approaches [7-14] are based on the understanding of physical mechanisms of the system and the accurate models. Data-driven approaches [15-30] are using historical operational data without accurate model presentations. Data-driven approaches seem more suitable for complex systems like NPPs, because it is virtually difficult to develop accurate physical mechanisms or models of them.

Lately, the increasing availability of enormous datasets of signal measurement has been favoring the data-driven approach over an analytical model-based approach for signal reconstruction. [31]. Typical examples of data-driven approaches include Artificial Neural Networks (ANNs) [15-21], Principal Component Analysis (PCA) [22-25], Independent

Component Analysis (ICA) [26], Auto-Associative Kernel Regression (AAKR) [27], Multivariate State Estimation Technique (MSET) [28], Support Vector Machines (SVMs) [28], and Fuzzy Similarity (FS) [29]. A comparison study pointed out that the auto-associative method, which is a kind of unsupervised learnings, is suitable because of its quickness and robustness [30].

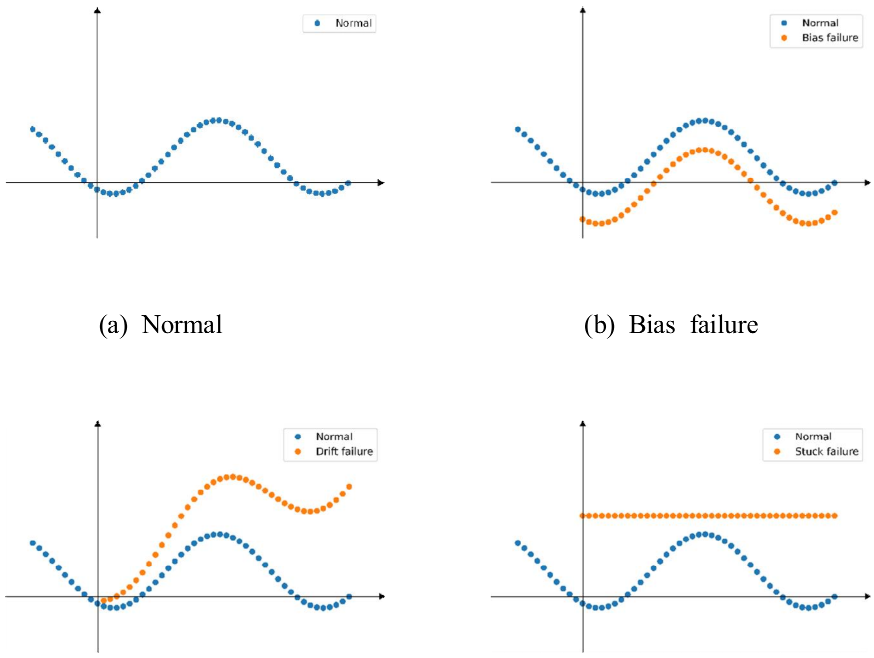
This study aims to propose an signal anomaly detection algorithm under the emergency situations in which signals are dramatically changing over time in NPPs. The proposed algorithm is based on Variational Auto-Encoder (VAE) and Long Short-Term Memory (LSTM), i.e., one of ANNs. First, this study discusses the signal behaviors under the emergency situation in NPPs and methods (i.e., VAE and LSTM). Then, an algorithm for signal anomaly detecting is proposed by applying LSTM and VAE. The optimization of algorithm is also conducted for selecting input sets, determining hyper-parameters of the network, and determining thresholds to identify signal anomaly. Finally, the suggested algorithm is validated by using the Compact Nuclear Simulator (CNS).

II. Signal Behavior and VAE & LSTM

This section reviews the signal behavior under the emergency situation in NPPs. Then, VAE and LSTM will be briefly introduced.

A. Signal Behavior in the Emergency Situation

Signal failures may occur due to many reasons such as abnormalities of sensor, transmitter, and/or cable, which can be caused by internal, external or environmental problems, e.g., pollution, vibration, extreme temperature, and aging [32]. Typical failure modes of sensors in NPPs can be divided into bias, drift, and stuck failures, as illustrated in Fig. 1. In the bias, a constant value is added to the normal, intact signal, while the drift is a time-correlated permanent offset failure. The stuck failure is one in which the signal is wrongly indicating a constant value. Typical stuck failures in NPPs are 'stuck at the highest value' (called, stuck-high), 'stuck at the lowest value' (called, stuck-low), and 'stuck at the current value at the time of failure' (called, stuck-as-is).



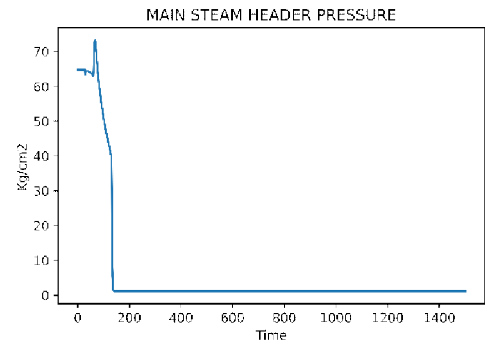
(a) Normal (b) Bias failure

(c) Drift failure (d) Stuck failure

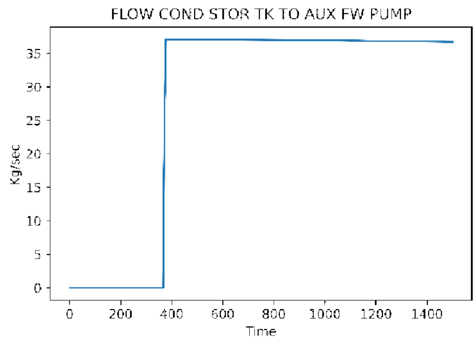
Fig. 1. Types of signal failure

Emergency situations are a situation in which the detection of signal failures is difficult by operators' visual inspection as well as even a computerized technique. In the normal situation, since plant parameters usually have a steady value, a parameter indicated by a faulty signal could be distinctive from the normal state of signal. However, in the emergency situation, many parameters are changing dramatically and it is difficult to distinguish whether a change of parameter are caused by the emergency situation or signal failure. Especially, the stuck failures may cause operator's misunderstanding about the situation if they regard the faulty signal as normal wrongly. For instance, Fig. 2 presents different behaviors of parameters in the emergency situation (e.g., LOCA scenario).

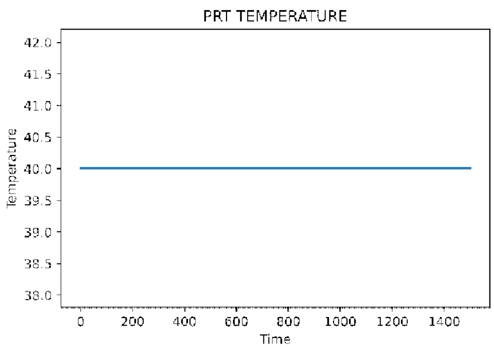
As shown in Fig. 2 (a), some parameters may show the minimum value of measurement, which is a similar pattern to the stuck-low failure. In addition, some parameters are expected to show the maximum value like Fig. 2 (b), which is similar to the stuck-high failure, while some parameters scarcely change like Fig. 2 (c). Therefore, if the stuck failures of signals are added to this situation, operators may have incorrect situation awareness and this may influence negatively their mitigations in the emergency situation.



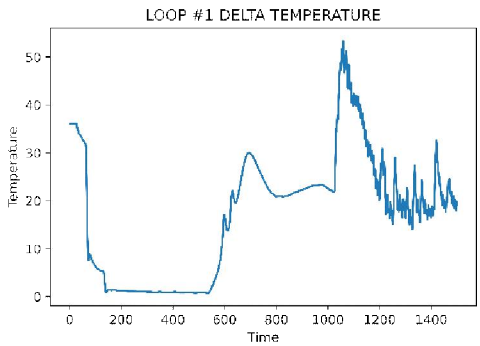
(a) Main steam head pressure, indicating the minimum value



(b) Aux feedwater flow from condensate storage tank, indicating the maximum value



(c) Pressurizer relief tank temperature, indicating a certain value



(d) Loop #1 delta temperature, changing over time

Fig. 2. Signal behaviors of NPPs in the emergency situation

B. Methods

This section introduced VAE, which is a kind of unsupervised learning, used to detect the signal failure, and LSTM used to process time-series data.

1. VAE

VAE is a variant of an Auto-encoder (AE) rooted in Bayesian inference [33]. VAE is a method of unsupervised learning 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 [33]. 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. 3.

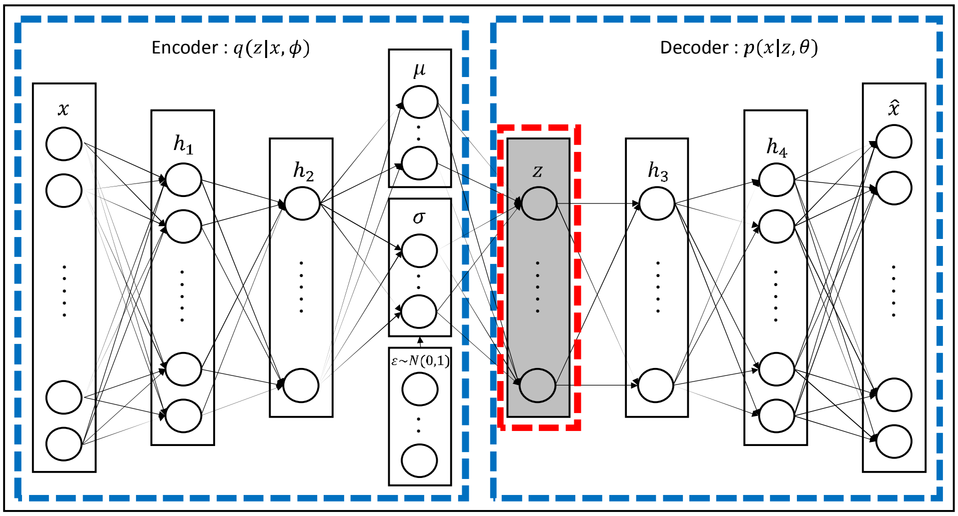


Fig. 3. The VAE structure

Goal of VAE is to model the distribution of observations $p(x)$ and generate new data by introducing latent random variables z. With the VAE, the posterior distribution is defined as $p(x) = \int p_\theta(z) p_\theta(x \mid z) dz$. Latent variable z is generated from a prior distribution $p(x)$, Θ 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 l $\log p(x) = D_{KL}(q_\Theta(z \mid x) \mid \mid p_\theta(z)) + L_{vae}(\Theta, \theta; x)$ notation from [39], where D_{KL} is Kullback-Leibler (KL) divergence from a prior $p_\theta(z)$ to the variational approximation $q_\Theta(z \mid x)$ of $p(z \mid x)$ and L_{VAE} is the variational lower bound of the data x by Jensen's inequality [33].

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

$$L_{vae}(\Theta,\theta;x) = -D_{KL}(q_{\Theta}(z \mid x) \mid \mid p_{\theta}(z)) + E_{q_{\Theta}(z \mid x)}[\log p_{\theta}(x \mid z)] \quad (1)$$

The first term of Eq. (1) 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. (1) is the reconstruction of x by maximizing the log likelihood $\log p_{\theta}(x \mid z)$ with sampling from $\log q_{\Theta}(z \mid x)$.

Signal detection through VAE is based on the probability of successful reconstruction. The VAE is trained for the reconstruction of the normal signal. If the VAE reconstructs the input signal successfully, it means that the characteristics of input signal are similar to those of the normal, trained signal. If the difference between the generated and input signals is large, it is likely that the input is not trained and so can be a faulty signal.

2. LSTM

LSTM is a kind of recurrent neural network (RNN), capable of learning long-short term dependency in sequence data [34-36]. The LSTM is designed to avoid the long-term dependency problem of RNN [37]. 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 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. 4.

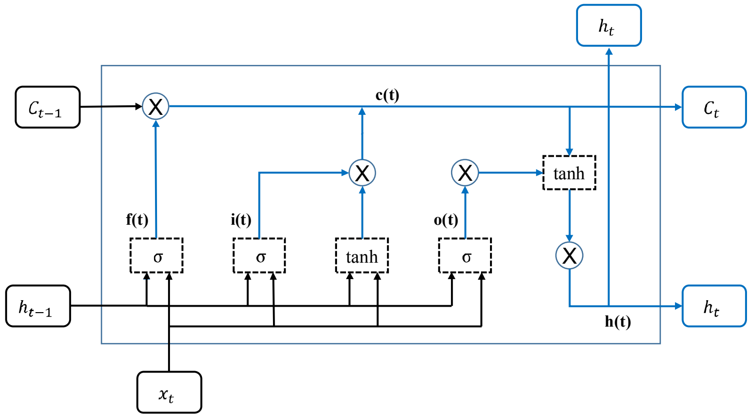


Fig. 4. The LSTM structure

Eq. (2-5) describe the output from each gate unit in a LSTM cell:

$$i_t = \sigma(x_t W_{xi} + h_{t-1} W_{xi} + b_i) \tag{2}$$

$$f_t = \sigma(x_t W_{xf} + h_{t-1} W_{xf} + b_f) \tag{3}$$

$$o_t = \sigma(x_t W_{xo} + h_{t-1} W_{xo} + b_o) \tag{4}$$

$$c_t = f_t c_{t-1} + i_t \tanh(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \tag{5}$$

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-1}). 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} 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 [38].

III. Development of Signal Anomaly Detection Algorithm in the Emergency Situation

An signal anomaly detection algorithm is proposed that can detect stuck failures of signals in the emergency situation by using a combined method of VAE and LSTM. Fig. 5 (i.e., the left part) presents the overview of the suggested algorithm. The details will be introduced in Section III.1. This study also carried out optimization activities to improve the performance of algorithm for the selection of input parameters, determination of hyper-parameters of network, and determination of thresholds as shown in the right part of Fig. 5. The LOCA was considered as an emergency situation using the compact nuclear simulator (CNS). The optimization activity will be presented in Section III.2.

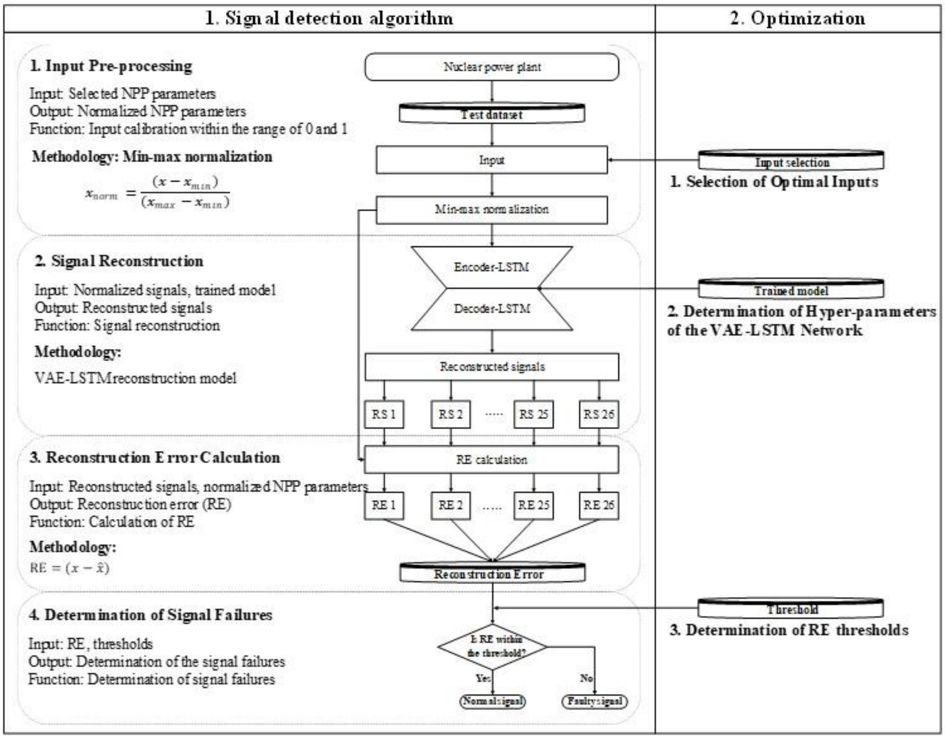


Fig. 5. The overview of the signal anomaly detection algorithm

A. Algorithm for Signal Anomaly Detection

The algorithm for the signal anomaly detection comprises four main steps: input preprocessing (step 1), signal reconstruction (step 2), reconstruction error calculation (step 3), and determining the signal failures (step 4). The optimization conducted concurrently with these steps is covered in the next section.

1. Step 1 (Input preprocessing)

Step 1 is to normalize selected input signals to be suitable for the input of VAE-LSTM network in the next step. Plant signals have different ranges of values or states (e.g., feedwater temperature: 220℃, steam generator

(SG) level: 50%, and valve state: open or closed). Generally, variables with higher values will have a larger impact on the network result. However, higher values are not necessarily more important for prediction. This problem causes local minima. To reduce this problem, the input pre-processing obtains the regular plant parameters as input and then outputs the normalized plant parameters that will be utilized by the network of the next step.

Min-max normalization is used to prevent local minima and increase the learning speed. A signal from the NPP is transformed to a value between 0 and 1 by using Eq. 6. x_t is the current value of the signal, while $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of collected data for that signal, respectively.

$$X_{norm} = \frac{(x_t - x_{\min})}{(x_{\max} - x_{\min})} \tag{6}$$

This step receives the selected signals as inputs. The signals that are highly related to the signal are selected by the Pearson correlation analysis to achieve the high performance in the detection of signal failure. The process to determine the inputs will be discussed in Section III.A.2.

2. Step 2 (Signal reconstruction using VAE-LSTM)

Step 2 of signal reconstruction attempts to produce the same value as each pre-processed input resulting from the previous step. This step is implemented by using a VAE-LSTM network, as shown in Fig. 6. The encoder receives the normalized signals from the previous step as inputs. It is trained to extract their features with the mean and standard deviation of normal distribution. Then, the decoder is expected to reconstruct the output as much as same as the input. The LSTM in both the encoder and decoder is used to handle time-series data because the plant signal is dynamic and the prior information is important in the prediction. The repeat vector was also used to increase the dimension of input values to the LSTM. The encoder and decoder consist of several layers and nodes, and the structure of network, i.e., the number of layers and nodes, was optimized to achieve the best reconstruction performance, which will be introduced in Section III.A.3.

Well trained VAE-LSTM network would produce the same output for each normalized input. The network is trained using the normal data that contain no signal failure, which is an unsupervised learning. Then, after the network is well trained, if a faulty signal that is not trained comes in, the deviation between the input and output, called the reconstruction error (RE), becomes large.

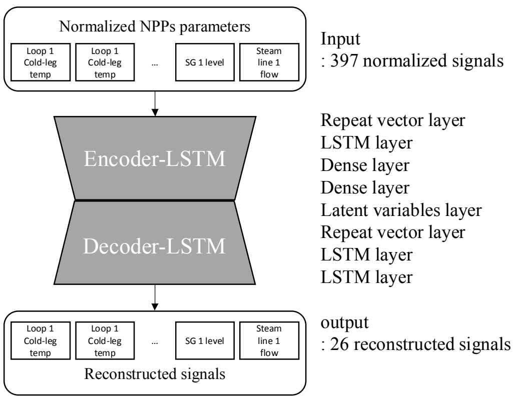


Fig. 6. Architecture of the Step 2

3. Step 3 (RE calculation)

Step 3 is to calculate the difference between the reconstructed signal generated from the previous step and the normalized input signal, i.e., RE.

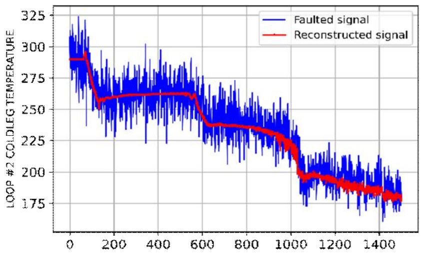
The RE is calculated using Eq. 7 as below;

$$RE = (x_t - \hat{x})^2 \quad (7)$$

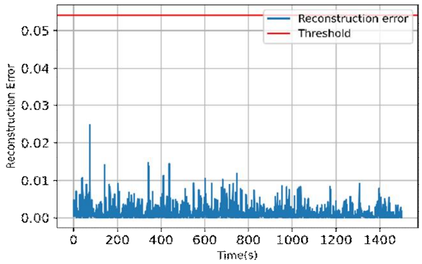
where x_t is the normalized value of original signal that is the input to the signal reconstruction step in Step 1, and $\hat{x}$ is the reconstructed value from Step 2.

Fig. 7 presents an example of how the reconstructed value is generated for the normal signal with no failure and the RE is calculated. Fig. 7 (a) depicts the original signal (blue line) for Loop #2 coldleg temperature in the LOCA that is obtained from the CNS and the reconstructed value (red line) from the VAE-LSTM network. Since the network is well trained for the normal signal, the RE would be very small as shown in Fig. 7 (b). Fig. 8 presents an example for handling faulty signals. As shown in Fig. 8 (a), the temperature signal fails at 300 sec. Since the network was not trained for this faulty signal, the difference between the reconstructed value and the faulty input signal becomes large as shown in Fig. 8 (b).

If we choose an appropriate criterion of RE that can discriminate normal and faulty signals, it is possible to detect signal failures. The process for the determination of the criterion, named a threshold, will be discussed in Section III.A.4.

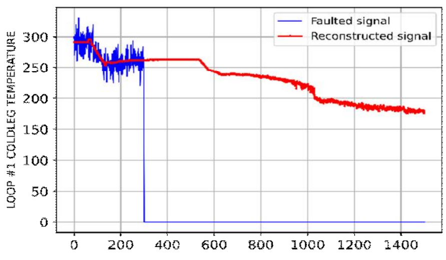


(a) Reconstruction result of loop2 coldleg temperature

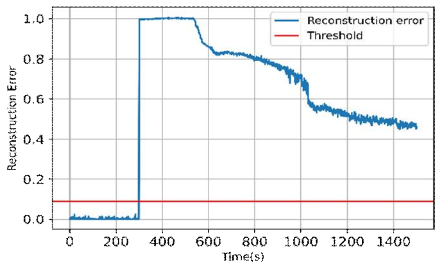


(b) RE calculation result of loop2 coldleg temperature

Fig. 7. Reconstruction result of loop2 coldleg temperature



(a) Reconstruction result of loop1 coldleg temperature



(b) RE calculation result of loop1 coldleg temperature

Fig. 8. Reconstruction result of loop1 coldleg temperature

4. Step 4 (Determination of signal failures)

Step 4 determines whether the signal is normal or faulty by comparing the RE calculated from the previous step and the pre-defined threshold. As discussed in the previous section, if the RE for a signal is smaller than the threshold, the signal is finally labeled as normal. On the other hand, if the RE exceeds the threshold, the signal is determined to be a faulty signal, as shown in Fig. 9.

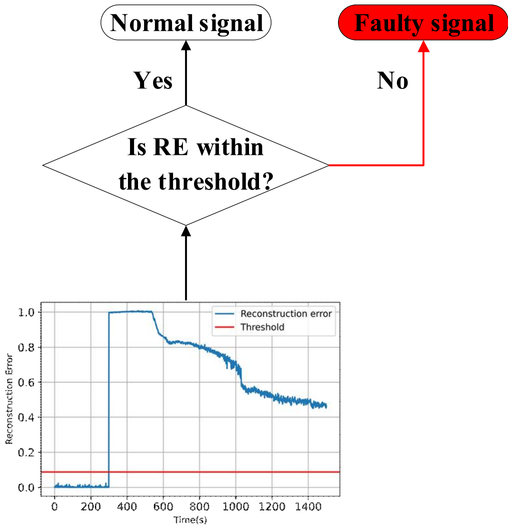


Fig. 9. Example for the Step 4

B. Optimization

1. Testbed

To improve the performance of algorithm, this study carried out the optimizations for 1) selecting inputs to the VAE-LSTM network, 2) determining the hyper-parameters of the VAE-LSTM network, and 3) defining the thresholds of RE for determining faulty signals. For the optimization, the CNS was used to simulate emergency situations. The CNS was developed by Korea Atomic Energy Research Institute (KAERI) with the reference to a Westinghouse 3 loop 900MW Pressurized Water Reactor (PWR) [39]. Fig. 10 shows the display of the CNS as an overview.

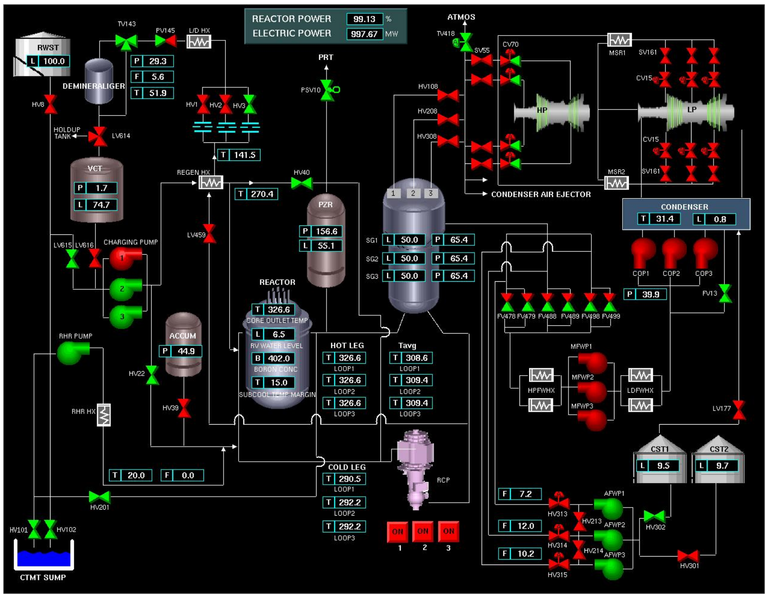


Fig. 10. Overview of the CNS

A total of 26 signals are selected for the optimization of the signal validation algorithm, as listed in Table 1. In other words, optimization is conducted to detect the stuck failures of these 26 signals.

Table 1. Selected signals for the optimization

NPP parameter
Feedwater pump outlet pressure
Feedwater line 1 flow
Feedwater line 2 flow
Feedwater line 3 flow
Feedwater temperature
Main steam flow
Steam line 1 flow
Steam line 2 flow
Steam line 3 flow
Main steam header pressure
Charging line outlet temperature
Loop 1 cold−leg temperature
Loop 2 cold−leg temperature
Loop 3 cold−leg temperature
Pressurized temperature
Core outlet temperature
Net letdown flow
Pressurized level
Pressurized pressure
Loop 1 flow
Loop 2 flow
Loop 3 flow
Steam generator 1 level
Steam generator 2 level
Steam generator 1 pressure
Steam generator 1 pressure

Table 2 presents the detailed list of collected data for this study. Data #1 includes normal signals from 49 LOCA scenarios and is used for the VAE-LSTM network training. Data #2 also includes the data of normal signals from five scenarios, is used for Optimizations 1 and 2. The data for faulty signals are divided for the purposes of optimization and validation. Data #3 includes the stuck failures of 26 selected variables. Note that the stuck-low dataset includes only the failures of 12 variables because the other 14 signals indicate the lowest values in the scenarios without any faults and are not distinguishable from the stuck-low failures. Data #4 is used for validation.

Table 2. The detailed list of collected data

Situation	Data #N	Failure Types		Number of Datasets
LOCA	Data #1	Normal		49 scenarios × 1,500 s = 72,627 datasets
	Data #2	Normal		5 scenarios × 1,500 s = 8,070 datasets
	Data #3	Faulty	Stuck−high	54 scenarios × 1,500 s × 26 signals = 2,098,122 datasets
			Stuck−low	54 scenarios × 1,500 s × 12 signals = 968,364 datasets
			Stuck−as−is	54 scenarios × 1,500 s × 12 signals = 2,098,122 datasets
	Data #4	Faulty	Stuck−high	18 scenarios × 1,500 s × 26 signals = 702,468 datasets
			Stuck−low	18 scenarios × 1,500 s × 12 signals = 270,180 datasets
			Stuck−as−is	18 scenarios × 1,500 s × 26 signals = 702,468 datasets

2. Optimization 1 (Selected of optimal input sets)

The objective of this optimization is to find out the set of optimal inputs to the VAE-LSTM network to reconstruct the normal signals of 26 plant variables. Different sets of inputs in the VAE-LSTM network would show different reconstruction performances. A correlation analysis was performed to choose the optimal set of inputs among 2,200 variables that are available in the CNS. The Pearson Correlation Analysis [40] has been used by applying the correlation coefficient given in Eq. 8.

$$r = \frac{\sum\left(\left(\frac{X_i - \overline{X}}{s_X}\right)\left(\frac{Y_i - \overline{Y}}{s_Y}\right)\right)}{N-1} \tag{8}$$

Here, N is the number of observations, X_i and Y_i are the values for the i-th observation where X indicates the 26 target variables for signal validation through stuck failure detection and Y indicates all the available variables in the CNS (i.e., 2,200 plant variables), and s is the standard deviation. Pearson's coefficient r has a value between -1 and 1, where the larger the absolute value of r, the higher the correlation. An r value approaching 1 means that there is positive linearity, while that approaching -1 means that there is negative linearity. A coefficient of 0 indicates that there is no linear correlation between the two variables.

As shown in Eq. 8, r is calculated among the 26 target variables and the CNS-available variables. Fig. 11 shows a portion of the calculation. Plant variables with correlation coefficients higher than a specific threshold are selected as the optimal input; this threshold is determined here through an experimental approach.

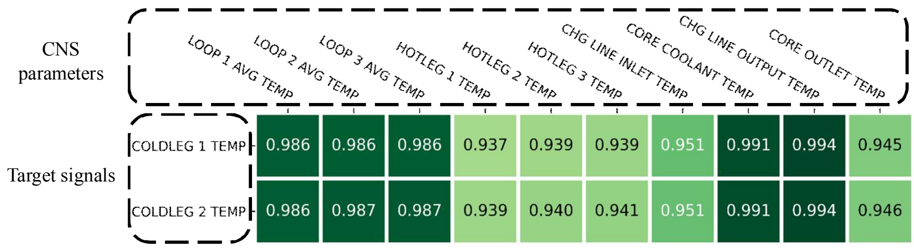


Fig. 11. Pearson correlation analysis result

The criterion was determined through an experimental approach. The accuracy in reconstructing target signals was investigated by varying the correlation coefficient. Table 3 presents the optimization result for the selection of inputs. The result indicates that the reconstruction of target signals presents the highest accuracy with the correlation coefficient r=0.985. Finally, total 397 variables that have the correlation coefficients higher than r=0.985 were selected as the inputs to the VAE-LSTM network.

Table 3. Number of inputs and reconstruction ratio according to the r

r value	Reconstruction accuracy of the target signal (%)	# of inputs
0.995	94.2%	157
0.985	99.8%	397
0.975	97.5%	604

3. Optimization 2 (Determination of VAE-LSTM hyperparameters)

This optimization determines hyper-parameters of the VAE-LSTM network, i.e., the number of batches, layers, and nodes. In general, these hyper-parameters affect reconstruction performances of the network.

Table 4 presents the comparison of RE and loss for eight different configurations by changing the number of batches, LSTM layers, LSTM nodes, and latent nodes. The REs and losses in Table 4 were calculated at 300 epochs. The loss is a number indicating how bad the network's prediction was. If the network's prediction is perfect, the loss is zero; otherwise, the loss is greater. One epoch means one iteration about an entire training data. The epoch is comprised of one or more batches that are number of sampling data. As mentioned in the previous section, ninety percent (90%) of training data were used for the model training and the other 10 percent of data was used for the test in the optimization.

Fig. 12 presents the trend of losses for eight configurations over epochs. This Figure indicates that these models have saturated losses around 300 epochs. As an example, Fig. 13 compares the signal reconstruction of Configurations 1 and 4 for SG #1 pressure in the LOCA scenario that has a rupture at the coldleg 1. It presents that Configuration 4 reconstructs the original signal more accurately and stably than Configuration 1. Finally, Configuration 4 that presents the smallest RE and loss was selected for the hyper-parameters of VAE-LSTM network. Consequently, the VAE-LSTM network in the signal reconstruction step has 3 LSTM layers, 4 LSTM nodes, and 8 latent nodes with 32 batches, as shown in Fig. 6.

Table 4. Performance comparison for different hyper-parameters

Configuration No.	Batch	LSTM layer	LSTM node	Latent node	RE	Loss
1	32	2	2	4	3.961E−2	1.129E−3
2	32	2	4	8	2.748E−2	8.721E−4
3	32	3	2	4	2.251E−3	9.017E−4
4	**32**	**3**	**4**	**8**	**1.074E−3**	**5.816E−4**
5	64	3	4	8	1.259E−3	8.753E−4
6	64	3	8	16	3.392E−3	7.139E−4
7	32	4	4	8	2.319E−3	1.010E−3
8	64	4	4	8	2.310E−3	1.090E−3

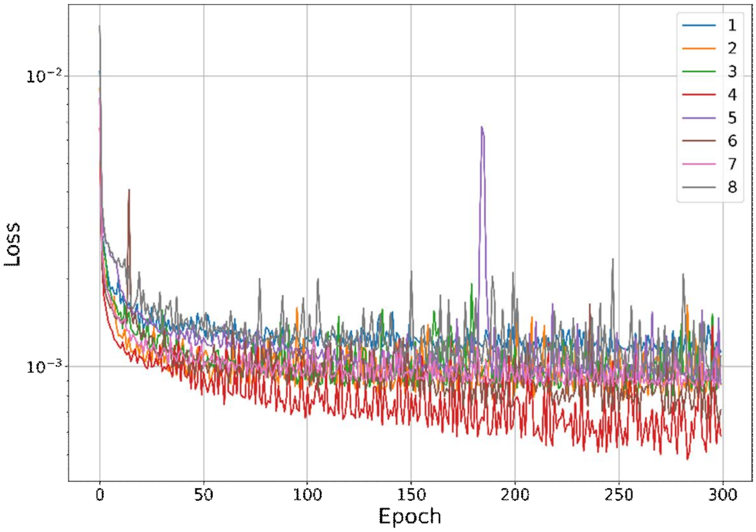


Fig. 12. Losses of different configuration over epochs

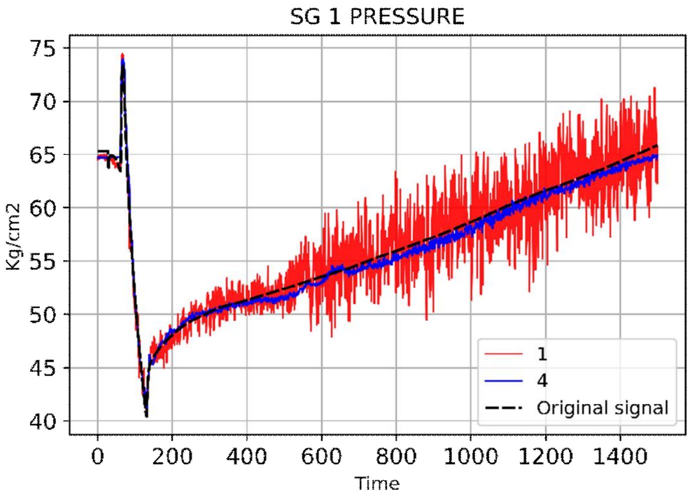


Fig. 13. Comparison of signal reconstruction for 1 and 4

4. Optimization 3 (Determination of RE thresholds)

This optimization determines the threshold of RE to judge whether an input signal is normal or faulty. Fig. 14 illustrates how the threshold is determined. As mentioned earlier, in the suggested algorithm, the RE is large for faulty signals because those signals are not trained, whereas the RE has small values for normal and trained signals. The threshold is a cutoff value of RE that divides normal or faulty signals. If the RE of a signal is higher than the threshold, the signal is regarded as a faulty one. If the RE of the signal is lower, it is normal. If the threshold is chosen too high, i.e., Case 1 in Fig. 14, the algorithm generates the result that both normal and faulty signals are normal. Therefore, the faulty signal is regarded as normal, which is Type 1 error. If the threshold is chosen too low, i.e., Case 3, the algorithm judges that both normal and faulty signals are faulty. In this case, the normal signal is detected as faulty, which is Type 2 error. If the threshold is chosen properly like Case 2, the algorithm becomes capable of distinguishing the normal and faulty signals correctly.

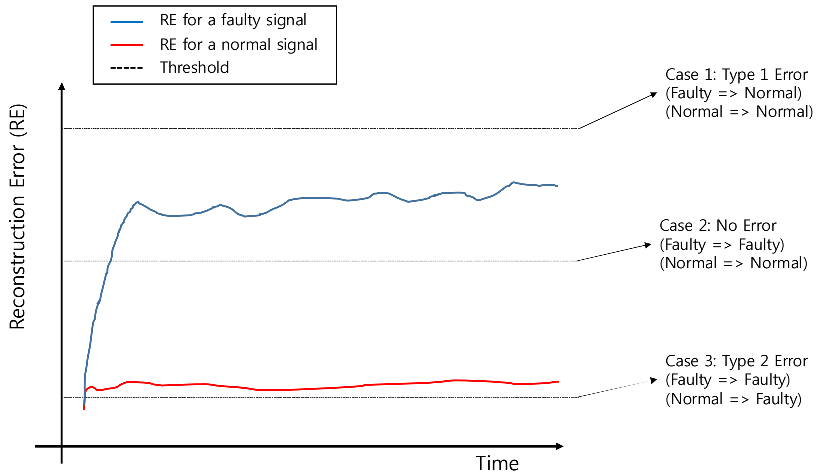


Fig. 14. Cases of thresholds

The RE threshold is determined based on the statistical method proposed by Shewhart's [41]. Shewhart's control charts are widely used to calculate changes in process features from the in-control state using Eq.9.

$$RE\ Thresholds = \mu + k\sigma \tag{9}$$

Here, μ and σ represent the mean and standard deviation, respectively, of the RE for each variable in the training data (i.e., Data #1), and k is a constant. This optimization step calculates the results of distinguishing normal and faulty signals for the 26 target variables by entering different k values (i.e., k = 0.5, 1, 2, or 3). By testing for Type 1 and Type 2 errors according to the k value, the optimal k can be determined.

Table 5 presents the comparison of Types 1 and 2 errors for the different k values. When k=0.5, which is the lowest value, Type 1 error was the smallest. However, this lowest threshold resulted in the largest Type 2 error that misjudges normal signals into faulty ones. As the k value increases, Type 1 error increases, but Type 2 error decreases. Based on the comparison in Table 5, this study selected k=1 that presents the best performance considering both Type 1 and 2 errors. Therefore, the following equation is applied to the threshold RE for the 'determination of signal failure' step in Fig. 9,

$$RE\ Thresholds = \mu + \sigma \tag{10}$$

The μ and σ refer to the mean and standard deviation of reconstruction error in each variable for the training data, respectively.

Table 5. Types 1 and 2 errors with different k values

Parameter	k = 0.5		k = 1		k = 2		k = 3	
	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2	Type 1	Type 2
FEEDWATER PUMP OUTLET PRESS	0	0.07	0	0	0	0.06	0	0
FEEDWATER LINE 1 FLOW	0	0.02	0	0	0	0	0	0
FEEDWATER LINE 2 FLOW	0	0	0	0	0	0	0	0
FEEDWATER LINE 3 FLOW	0	0.001	0	0	0	0	0	0
FEEDWATER TEMP	0	0.11	0	0	0.90	0.09	0.49	0
MAIN STEAM FLOW	0	0.06	0	0	0.93	0.06	0.06	0
STEAM LINE 1 FLOW	0	0.011	0	0	0	0.10	0	0
STEAM LINE 2 FLOW	0	0.011	0	0	0	0.10	0	0
STEAM LINE 3 FLOW	0	0.011	0	0	0	0.11	0	0
MAIN STEAM HEADER PRESSURE	0	0.011	0	0	0	0.10	0	0
CHARGING LINE OUTLET TEMP	0	3.5	0.06	0.06	0.67	0.21	1.28	0
LOOP 1 COLDLEG TEMP	0	3.5	0.06	0.02	0.14	0.38	0.58	0
LOOP 2 COLDLEG TEMP	0	3.5	0.12	0.02	0.03	2.97	0.95	0.005
LOOP 3 COLDLEG TEMP	0	3.5	0.06	0.01	0.23	0.23	0.64	0
PZR TEMP	0	3.5	0	0.02	0.14	0.14	0.43	0
CORE OUTLET TEMPE	0	3.5	0.35	0.01	0.55	0.12	0.98	0
NET LETDOWN FLOW	0	0.002	0	0	0	0.002	0	0
PZR LEVEL	0.75	0.05	0.41	0	1.07	0.04	0.87	0
PZR PRESSURE	0	3.5	0.49	0.02	2.05	0.15	3.04	0
LOOP 1 FLOW	0	3.56	0	0	0	0.13	0	0
LOOP 2 FLOW	0	3.56	0	0	0	0.15	0	0
LOOP 3 FLOW	0	3.56	0	0	0	0.11	0	0
SG 1 LEVEL (WIDE)	0	3.5	0	0.01	0	1.24	0.29	0
SG 2 LEVEL (WIDE)	0	3.5	0	0.01	0	1.90	0.20	0
SG 1 PRESSURE	0	3.5	0.52	0.002	0.20	0.02	1.53	0
SG 2 PRESSURE	0	0.99	0.35	0.001	0.61	0.02	1.56	0
Sum	0.75	47.41	2.40	0.19	7.52	8.44	12.91	0.005

IV. Results of the Optimization

Table 6 presents the accuracy of the algorithm obtained as a result of the three aforementioned optimizations using Data #3 from Fig. 12. The algorithm determined 99.81% of the normal signals as "normal." It also could detect 97.6% of signal failures (i.e., 100% of the stuck-high, 98.92% of the stuck-low, and 93.88% of the stuck-as-is failures).

Table 6. Optimization result using Data #3

Failure mode		Classification result (%)	
		Faulty	Normal
Failed	Stuck−high	100	0
	Stuck−low	98.92	1.08
	Stuck−as−is	93.88	6.12
	Total	97.6	2.4
Normal		0.19	99.81

V. Validation

Fig. 15 shows an example of the process by which the signal anomaly detection algorithm process detects stuck signal failures. In this Fig 15, the algorithm receives two signals as inputs from the LOCA scenario. The loop 1 coldleg temperature signal is faulty signal, namely a stuck-high, while the other signal, PZR pressure, is normal. Step 1 of the algorithm normalizes these signal inputs to a range of 0 to 1. Step 2 attempts to reconstruct the normalized signals similarly to the in-put signals. Then step 3 calculates the RE from the difference between the normalized and reconstructed signals. Step 4 compares the calculated RE to the threshold defined in the third optimization.

As shown in Fig 15, the RE of the loop 1 cold-leg temperature is larger than the threshold, and based on the comparison, step 5 determines that the input signal is faulty.

The proposed algorithm was also validated using the data that were not used in either the training or optimization, i.e., Data #4. Table 7 presents the accuracy of the signal anomaly detection algorithm using validation data. It is indicated that the algorithm can detect 96.70% of all stuck failures and 98.29% of all normal signals in the scenarios from this dataset. These validation results are similar to those achieved via the optimization shown in Table 7.

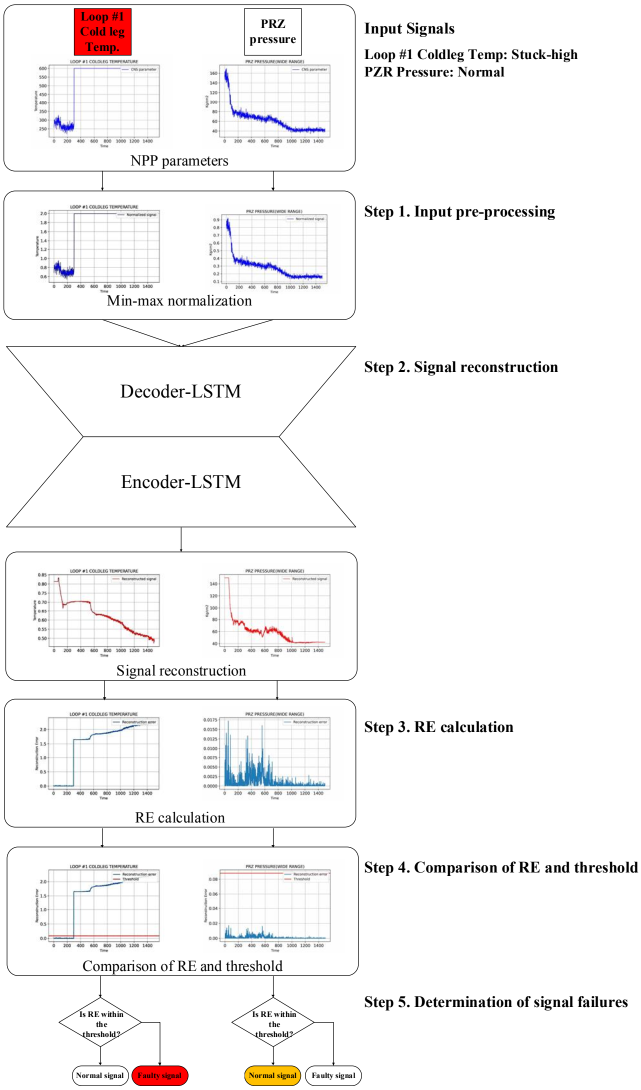


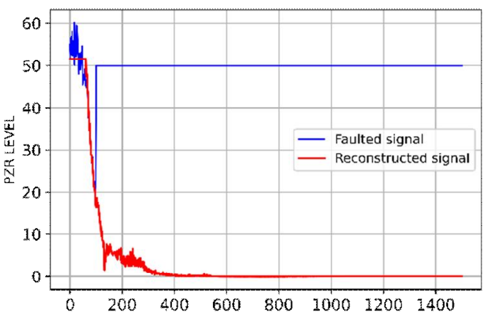
Fig. 15. The validation process of algorithm

Table 7. Accuracy of the signal anomaly detection algorithm using the Data #4

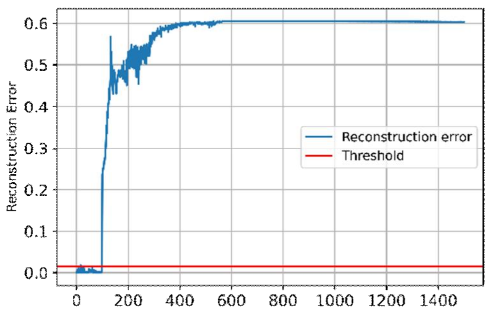
Failure modes		Classification results (%)	
		Failed	Normal
Failed	Stuck−high	100	0
	Stuck−low	97.92	2.08
	Stuck−as−is	92.18	7.82
	Total	96.70	3.30
Normal		1.71	98.29

VI. Discussion

The signal validation algorithm using unsupervised learning can detect the entire range of stuck failures that are not close to normal signal values. For instance, Fig. 16 (a) shows a 50% stuck failure of the PZR signal at 150 s. The stuck-high (100%), stuck-low (0%), and stuck-as-is (20%) failures were tested and shown to be detectable, as described in the previous section. Fig. 16 (b) demonstrates that the "stuck at 50%" failure is also detectable by the algorithm because the RE is larger than the pre-defined threshold for this parameter.



(a) PZR level signal stuck at 50%



(b) Detection of signal

Fig. 16. An example of detecting stuck failure at other values

VII. Conclusion

An algorithm has been proposed and optimized for the signal anomaly detection under emergency situations by using deep learning methods. The algorithm comprises four main steps: input preprocessing (step 1), signal reconstruction (step 2), reconstruction error calculation (step 3), and determining the signal failures (step 4). It also applied the VAE-LSTM which is based on the deep learning method. The algorithm has been also optimized to detects three types of signal failures, i.e., stuck-high, stuck-low, and stuck-as-is, using the CNS. The validation result presented that the algorithm can detect the signal failures successfully under an emergency situation that is the LOCA.

Since the algorithm was developed based on the unsupervised learning, it has the capability of detecting a wide range of stuck failures, which is practically impossible for the supervised-based learning. This study is going to be extended to different types of emergency situations such as main steam line break (MSLB) and steam generator tube rupture (SGTR), and to multiple signal failures.

REFERENCES

[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] P. Le Bot, Human reliability data, human error and accident models-illustration through the Three Mile Island accident analysis. Reliability Engineering & System Safety, 83(2), pp. 153-167, 2004.

[3] B. R. Upadhyaya, T. W. Kerlin, and Jr. P. J. Gaudio, Development and testing of an integrated signal validation system for nuclear power plants. Combustion Engineering, Inc., Stamford, CT (USA); Tennessee Univ., Knoxville, TN (USA). Dept. of Nuclear Engineering, Vol. 1, 1989.

[4] 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.

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

[6] 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.

[7] 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.

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

[9] 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.

[10] 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.

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

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

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

[14] 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.

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

[16] 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.

[17] 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.

[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] 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.

[20] S. Seker, E. Ayaz,and E. T?rkcan, Elman's recurrent neural network applications to condition monitoring in nuclear power plant and rotating

machinery. Engineering Applications of Artificial Intelligence, 16(7-8), pp. 647-656, 2003.

[21] 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.

[22] C. K. Yoo, K. Villez, I. B. Lee, S. Van Hulle, & P. A. Vanrolleghem, Sensor validation and reconciliation for a partial nitrification process, Water science and technology, 53(4-5), 513-521 2006.

[23] 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.

[24] 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.

[25] P. Baraldi, A. Cammi, F. Mangili, & E. Zio, An ensemble approach to sensor fault detection and signal reconstruction for nuclear system control, Annals of Nuclear Energy, 37(6), 778-790, 2010.

[26] H. Al-Bazzaz and X. Z. Wang, Statistical process control chart for batch operations based on independent component analysis, Ind. Eng. Chem. Res., vol. 43, pp. 6731-6741, 2004.

[27] 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.

[28] N. Zavaljevski, K. C. Gross, Sensor fault detection in nuclear power plants using multivariate state estimation technique and support vector machines. No. ANL/RA/CP-103000. Argonne National Lab., Argonne, IL (US), 2000.

[29] E. Zio, F. Di Maio, A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. Reliability Engineering & System Safety, 95(1), pp. 49-57, 2010.

[30] P. Baraldi, F. Di Maio, D. Genini,and E. Zio, Comparison of data-driven reconstruction methods for fault detection. IEEE Transactions on Reliability, 64(3), pp. 852-860, 2015.
[31] J. D. Boskovic, 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] Zúñiga, Andrés A., et al. Classical Failure Modes and Effects Analysis in the Context of Smart Grid Cyber-Physical Systems. Energies, 13.5: 1215, 2020.
[33] J. An, S. Cho, Variational autoencoder based anomaly detection using reconstruction probability, Special Lecture on IE, 2(1), pp.1-18, 2015.
[34] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation, 9(8), pp.1735-1780, 1997.
[35] J. Yang, J. Kim, Accident diagnosis algorithm with untrained accident identification during power-increasing operation, Reliability Engineering & System Safety, vol.202, 2020.
[36] J. Yang, J. Kim, An accident diagnosis algorithm using long short-term memory, Nuclear Engineering and Technology, 50(4), pp.582-588, 2018.
[37] F. A. Gers, J. Schmidhuber, F. Cummins, Learning to forget: Continual prediction with LSTM, 1999.
[38] 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.
[39] KAERI, 'Advanced compact nuclear simulator textbook, KAERI, Nucl. Training Center Korea At. Energy Res. Inst., Daejeon, South Korea, Tech. Rep., 1990.
[40] H. Xu, Y. Deng, Dependent evidence combination based on shearman coefficient and pearson coefficient, IEEE Access, 6, pp. 11634-11640, 2017.

[41] H. Z. Nazir, M. Schoonhoven, M. Riaz, and R. J. Does, Quality quandaries: a stepwise approach for setting up a robust Shewhart location control chart. Quality Engineering, 26(2), pp. 246-252, 2014.