

Article

A Fault Detection and Isolation Method via Shared Nearest Neighbor for Circulating Fluidized Bed Boiler

Minseok Kim , Seunghwan Jung , Eunbyeong Kim , Baekcheon Kim , Jinyong Kim  and Sungshin Kim * 

Department of Electrical and Electronics Engineering, Pusan National University, Busan 46241, Republic of Korea; minskey@pusan.ac.kr (M.K.); shjung0091@pusan.ac.kr (S.J.); kimeunbyeong@pusan.ac.kr (E.K.); 1001000@pusan.ac.kr (B.K.); skes1234@pusan.ac.kr (J.K.)

* Correspondence: sskim@pusan.ac.kr; Tel.: +82-51-510-2374

Abstract: Accurate and timely fault detection and isolation (FDI) improve the availability, safety, and reliability of target systems and enable cost-effective operations. In this study, a shared nearest neighbor (SNN)-based method is proposed to identify the fault variables of a circulating fluidized bed boiler. SNN is a derivative method of the k-nearest neighbor (kNN), which utilizes shared neighbor information. The distance information between these neighbors can be applied to FDI. In particular, the proposed method can effectively detect faults by weighing the distance values based on the number of neighbors they share, thereby readjusting the distance values based on the shared neighbors. Moreover, the data distribution is not constrained; therefore, it can be applied to various processes. Unlike principal component analysis and independent component analysis, which are widely used to identify fault variables, the main advantage of SNN is that it does not suffer from smearing effects, because it calculates the contributions from the original input space. The proposed method is applied to two case studies and to the failure case of a real circulating fluidized bed boiler to confirm its effectiveness. The results show that the proposed method can detect faults earlier (1 h 39 min 46 s) and identify fault variables more effectively than conventional methods.

Keywords: fluidized bed boiler; fault isolation; shared nearest neighbor; k-nearest neighbor



Citation: Kim, M.; Jung, S.; Kim, E.; Kim, B.; Kim, J.; Kim, S. A Fault Detection and Isolation Method via Shared Nearest Neighbor for Circulating Fluidized Bed Boiler. *Processes* **2023**, *11*, 3433. <https://doi.org/10.3390/pr11123433>

Academic Editors: Stefania Tronci, Massimiliano Errico, Riccardo Bacci Di Capaci, Ingmar Nopens, Elena Torfs and Michael Short

Received: 29 November 2023
Revised: 12 December 2023
Accepted: 13 December 2023
Published: 15 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Modern industrial processes continue to grow owing to efforts to expand facility sizes and establish efficient operational strategies. To support the growth rate of these processes, monitoring technologies such as fault detection and diagnosis (FDD) that are capable of ensuring facility safety are required [1]. Faults can occur for various reasons, including equipment defects and malfunctions. For example, in a circulating fluidized bed boiler, critical equipment, such as turbines, compressors, and generators, operates in hazardous environments (e.g., high pressure and temperature). Such operating conditions cause potential failures; if the fault is not diagnosed early, the smearing effect can ultimately lead to unplanned downtime. Accordingly, monitoring technology is crucial for preventing failures. Since early FDD can avoid unexpected downtimes and high maintenance costs, many scholars have recently proposed methods for fault diagnosis technology in recent years [2–4].

Process monitoring is performed in four steps: fault detection, isolation, diagnosis, and recovery [5]. Fault detection and isolation (FDI) determines whether a fault has occurred in the target system and then identifies the variables related to the failure. Fault diagnosis determines the type of failure in which the variables determined in the previous step are based on the failure logging history. Finally, the recovery stage repairs any failures that occur in the target system and restores them to a normal state.

As explained above, it is essential to identify the variables that cause the failure of an accurate fault diagnosis. Commonly employed methods for detecting and identifying fault

variables include multivariate statistical process monitoring (MSPM) methods, such as principal component analysis (PCA) [6,7], independent component analysis (ICA) [8,9], and partial least squares (PLS) [10,11]. In the past decades, MSPM has been successfully applied to various processes [1]. However, in multivariate processes, accurate fault diagnosis is difficult because of the smearing effect of fault variables. The smearing effect is a problem in which a fault variable affects a normal variable such that the normal variable is regarded a fault. If the fault smearing is large, the isolation results through contribution analysis may confuse technicians and engineers, and accurate diagnosis results cannot be guaranteed. As industrial systems become larger and more interconnected, additional process knowledge is required to determine the root cause. Thus, conventional methods have clear limitations. Therefore, a method for identifying the fault variables that considers the reflection effect is essential.

Fault isolation is primarily divided into contribution analyses [12] and reconstruction-based contribution (RBC) methods [13]. The contribution analysis calculates the contribution of each variable in the failure situation and selects the variable with the largest contribution as the cause of failure [14]. The calculated contribution of each variable can be compared to the normal operating condition (NOC), which is the upper limit of the contribution of normal data, to identify the faulty variable easily. Methods derived from contribution analysis include complete decomposition contribution (CDC), partial decomposition contribution (PDC), and diagonal contribution (DC), which have been proposed [15]. Kourti [16] conducted a contribution analysis for a high-pressure, low-density polyethylene reactor. Liu [17] proposed a modified contribution plot-based approach to reduce the back-burying effect of non-fault variables. However, traditional contribution analysis methods may increase or decrease the contributions of normal variables owing to the smearing effect, leading to misdiagnosis [18]. Westerhuis [12] showed that traditional attribution analysis methods are affected by the ask-back effect, which can lead to a misdiagnosis. In contrast, reconstruction-based methods reconstruct fault detection indices to calculate the contributions of the variables. Alcalá [15] demonstrated that RBC methods can accurately identify fault variables in processes with smearing effects. Owing to these advantages, research on RBC methods continues to be conducted. Xu [19] proposed a weighted RBC fault isolation method based on the Tennessee Eastman (TE) process. Wang [20] used a reconstruction method based on the distance calculated by kNN instead of the test statistic derived from PCA to identify fault variables. Zhou [18] proposed a new isolation index based on kNN. Compared with conventional methods, these methods reconstruct the detection index based on the variables of the original measurement space, thereby preventing the smearing effect. However, the kNN must select the number of close neighbors for FDI. These conditions have the potential to degrade the model performance. For example, if the neighbors selected from healthy and faulty data are located within the decision boundaries of normal, FDI performance will degrade. Therefore, FDI performance can be improved by removing unnecessary neighbors for fault detection.

In this study, we propose a FDI method based on a shared nearest neighbor (SNN). As depicted in Figure 1, the FDI procedure of the proposed method is divided into off-line monitoring, which includes the selection of SNN neighbors and weighting of SNN neighbors, and on-line monitoring, which identifies the process variables that cause the fault from the period after the fault occurs until the target system is stopped. First, the normal operation data stored in the database are normalized to equalize the mean and variance of the process variables. Data normalization was first performed, because it ensures that important variables for FDI are not hidden by other variables. Upon performing data normalization, neighbors between the training data and the query vector are searched using the SNN. Unlike kNN, the neighbor selection of SNN determines a shared neighbor among adjacent neighbors as the final neighbor. Since the training data consist of normal data, neighbors that share it with each other are likely to be close to the normal data. In other words, it has a more robust performance than kNN. In particular, the proposed method assigns weights based on the number of shared neighbors among the neighbors to

adjust the distance value to be lower when they are close to the normal data. If a sample has no shared neighbors, the farthest neighbor is selected and assigned a higher weight, resulting in a greater distance. The distance calculated from each neighbor was used as a detection index to identify the fault variable.

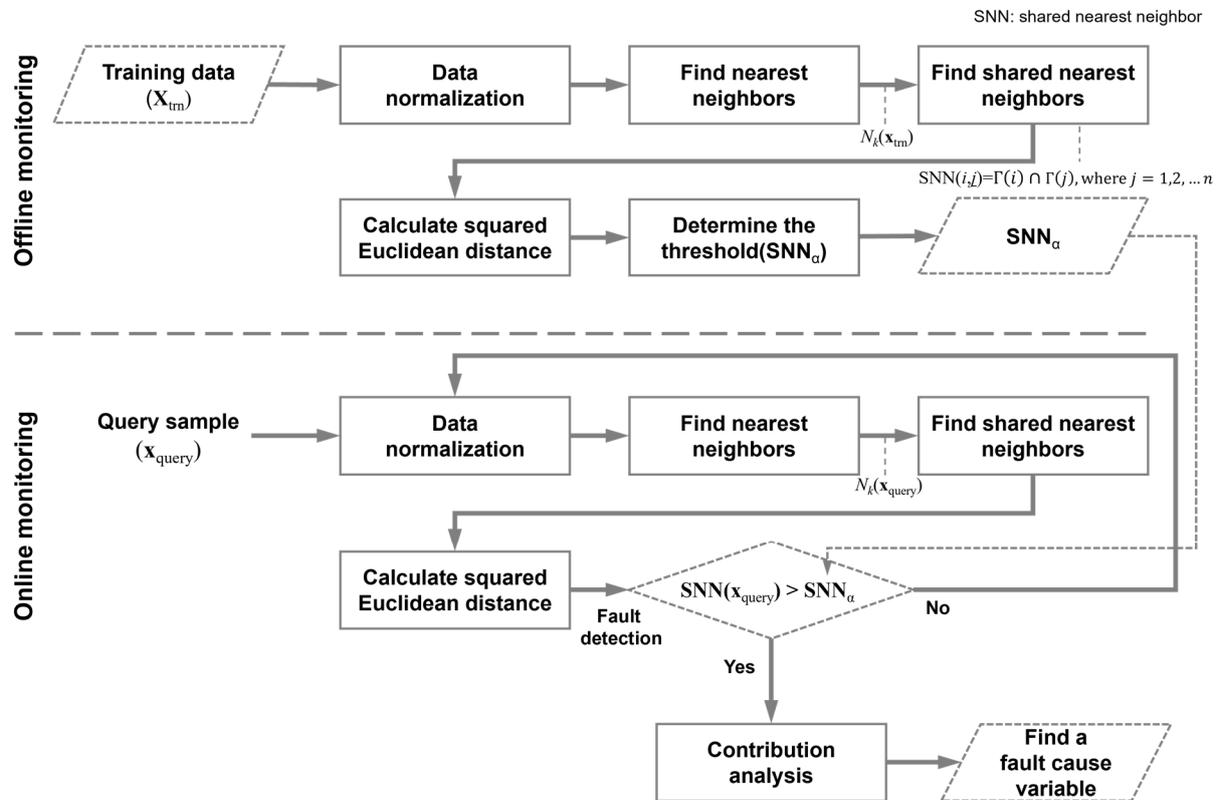


Figure 1. Flowchart of the fault detection procedure using the proposed method.

The highlighted advantages of the proposed method are as follows: (1) It does not require knowledge of the process required for fault isolation. (2) Unlike statistical monitoring methods, this method can be applied without assuming a specific data distribution. (3) The number of neighbors does not need to be assigned. (4) The influence of faulty data can be reduced by selecting shared neighbors as the final neighbors. (5) The FDI performance can be improved by assigning weights to each shared neighbor. (6) Smearing is prevented by reconstructing the detection index in the original space. (7) This can also be applied to the process by applying multimode normalization.

The remainder of this paper is organized as follows: Section 2 explains the FDI method using the SNN. Section 3 describes the experimental data used to verify the model's performance and the SNN's neighbor selection. In Section 4, we discuss case studies of two types of failures (single and multiple) and the results of fault isolation in an actual circulating fluidized bed boiler. Finally, we discuss our conclusions and future work in Section 5.

2. Proposed Method

This section introduces this concept and provides a detailed description of SNNs for FDI.

2.1. Shared Nearest Neighbor

The SNN is a method derived from kNN that calculates the similarity between two points using common neighbors [21]. This method can be applied to nonlinear systems and multimodal and time-varying processes, because it can determine normality or failure

based on the similarity of common neighbors between the query vector and training data [22]. The detailed concept of SNNs can be found in [21]. In this study, the distance was adjusted based on the number of neighbors by assigning weights to the selected neighbors in the conventional SNN to represent the difference in the location of the normal and fault data. The fault isolation procedure of the proposed method is performed as follows: (1) search for the nearest neighbor using a distance function, (2) search for shared neighbors among neighbors, and (3) weight allocation based on the number of shared neighbors and calculation of a new distance value (detection index) using the final number of neighbors and weights.

(1) Collect normal operation data ($\mathbf{x}_{\text{trn}} \in \mathbb{R}^{n \times m}$ with n samples m variables) from the target system. Normal data are used to assign weights to the query vector. The similarity between the query vector ($\mathbf{x}_{\text{query}} \in \mathbb{R}^{1 \times m}$ with m variables) and the training data (\mathbf{x}_{trn}) stored in the database is calculated using the Euclidean distance, as shown in Equation (1):

$$d(\mathbf{x}_{\text{trn}}, \mathbf{x}_{\text{query}}) = \sqrt{(x_{\text{trn}}^1 - x_{\text{query}}^1)^2 + (x_{\text{trn}}^2 - x_{\text{query}}^2)^2 + \dots + (x_{\text{trn}}^m - x_{\text{query}}^m)^2}, \quad (1)$$

where \mathbf{x}_{trn} and $\mathbf{x}_{\text{query}}$ denote the training data and query vectors, respectively. Select k training data samples adjacent to $\mathbf{x}_{\text{query}}$ based on the distance calculated using Equation (1). In the case of kNN, faulty data adjacent to normal data can be considered normal because of the need to select neighbors for the query vector. In contrast, the SNN calculates similarity by selecting common neighbors; thus, it can effectively detect the fault even when there is fault data adjacent to the query vector.

(2) The common neighbors of \mathbf{x}_{trn} and $\mathbf{x}_{\text{query}}$ are selected as follows based on the distance values calculated in Equation (1):

$$\text{SNN}_i(\mathbf{x}_{\text{trn}}^i, \mathbf{x}_{\text{query}}) = \Gamma(\mathbf{x}_{\text{trn}}^i) \cap \Gamma(\mathbf{x}_{\text{query}}), \quad i = 1, \dots, n, \quad (2)$$

where Γ indicates the cardinality of the corresponding neighbor. Unlike kNN, SNN selects neighbors only when common neighbors exist. If common neighbors do not exist, the neighbor with the furthest distance is selected from among the k neighbors adjacent to $\mathbf{x}_{\text{query}}$. For example, even if three neighbors are selected in kNN, if only one shared neighbor exists, only one final neighbor is selected.

(3) To calculate the detection index, the distance values of the selected neighbors are calculated as shown in Equation (3):

$$D_i^2 = \sum_{j=1}^s d_{i,j}^2, \quad s = 1, \dots, k, \quad (3)$$

where D_i^2 is the squared distance and number of common neighbors, respectively. The calculated distribution of D_i^2 approximately follows a non-central chi-square distribution, and the distance value of D_i^2 is adjusted such that the weight is multiplied by the number of neighbors. For example, if more than 50% of the neighbors are shared, the distance value is adjusted by multiplying it by a weight of 0.05. Consequently, if the query vector is located within the normal data distribution, the distance decreases.

2.2. Nearest Neighbor Difference Normalization

Modern industrial processes are operated in various modes. Accordingly, the process data may follow different distributions, such as Gaussian and non-Gaussian distributions [23]. Data normalization is necessary to detect and isolate faults effectively. If data normalization is not performed, there exists a risk that mode data with a large difference in value compared to other normal mode data may be considered a fault. Therefore, in this study, we performed nearest neighbor difference (NND)-based normalization to normalize multimodal data to unimodal data. The NND was proposed by [1] to improve the fault detection performance of processes using multimodal data. A detailed description of NND

can be found in [1], and the procedure for normalization is as follows: (1) first-order NND calculation and (2) second-order NND calculation.

- (1) To compute the NND, kNN is used to explore the nearest neighbors. The neighbors of a query vector are subtracted from the corresponding query vector to compute the first-order NND, as shown in Equation (4):

$$\tilde{\mathbf{x}} = \mathbf{x}_{\text{query}} - \mathbf{x}_{\text{trn}}^{(k)} \quad (4)$$

where $\mathbf{x}_{\text{query}}$ and $\mathbf{x}_{\text{trn}}^{(k)}$ denote the nearest neighbors of the training data and query vectors, respectively. The first-order NND, computed using Equation (4), removes the multicenter structure, while preserving the position information between the current sample and its nearest neighbors.

- (2) Subsequently, the second-order NND was calculated to convert the multimode data into single-mode data, as shown in Equation (5):

$$\tilde{\mathbf{x}}' = \omega [\tilde{\mathbf{x}}_{\text{query}} - \tilde{\mathbf{x}}_{\text{trn}}^{(k)}], \quad (5)$$

where $\omega = 1 / \|\tilde{\mathbf{x}}_{\text{trn}}^{(k)}\|$ denotes a weight parameter used to map to a single mode. $\|\tilde{\mathbf{x}}_{\text{trn}}^{(k)}\|$ is the Euclidean distance between the query vector and the kth neighbor training data. The original data are converted into a single mode with multiple modes using a second-order NND. In addition, after the second-order NND, each variable follows a multivariate Gaussian distribution. A detailed explanation of how multimode characteristics are removed using NND can be found in [1]. In this study, a SNN was applied to data from which multimode characteristics were removed to calculate the detection index for identifying failure variables.

3. Numerical Simulation Study

In this section, two failure processes (single and multiple) were studied to verify the fault isolation performance of the proposed method. Section 3.1 describes the simulation data used to identify the fault variables and compares the fault isolation performance of conventional methods and the proposed method. We used CPU Intel Core i9-9900K (Intel, Santa Clara, CA, USA) and GPU Nvidia RTX 2070 (Nvidia, Santa Clara, CA, USA) hardware to run these tests. The algorithm was implemented in MATLAB 2021b.

3.1. Multimode Numerical Example

Carlos [15] proposed a simple numerical example to compare the diagnostic results when single or multiple faults of moderate sizes occur. As shown in Equation (6), this example is calculated using the simulation model. Zhou [18] used simulation data to compare the performances of the proposed model. This example consists of six variables that can be generated using the following system:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} -0.3441 & 0.4815 & 0.6637 \\ -0.2312 & -0.5936 & 0.3545 \\ -0.5060 & 0.2495 & 0.0739 \\ -0.5552 & -0.2405 & -0.1123 \\ -0.3371 & 0.3822 & -0.6115 \\ -0.3877 & -0.3868 & -0.2045 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} + \text{noise}, \quad (6)$$

where t_1, t_2, t_3 are random variables with uniform distributions in $[0, 1]$, $[0, 1.6]$, and $[0, 1.2]$, respectively. *noise* refers to white noise with zero mean and a standard deviation of 0.01.

The training data and query vectors generated in Equation (6) are 3000 and 1500, respectively. The query vectors operate normally from 1 to 500, and in samples 501 to 1500,

a fault of an appropriate size occurs in single or multiple variables. Single or multiple variable failures were added, as shown in Equation (7):

$$\mathbf{x}_{\text{single fault}} = \mathbf{x}^* + \xi_p f_p, \quad (7)$$

$$\mathbf{x}_{\text{multiple fault}} = \mathbf{x}^* + \xi_p f_p + \xi_q f_q, \quad (p \neq q), \quad (8)$$

where the fault-free measurements (\mathbf{x}^*) are normal variables generated using Equation (8). The defect size (f_p) is uniformly distributed between 0 and 5. The fault variables (p and q) were chosen uniformly at random from the six variables (p and q were not equal).

- Case 1. The system was initially running normally, and then an arbitrary bias-type single fault was added from the 501st to the 1500th sample.
- Case 2. The system was initially running normally, and then arbitrary bias-type multiple faults were added from the 501st to 1500th samples.

To identify the fault variables, the threshold of the conventional and proposed methods was set to 0.01. The number of principal components for PCA, DPCA, ICA, and DICA was set to five under the same conditions, and the lag variables for DPCA and DICA were set to two. The LOF sets the number of neighbors to 60. The NND and k values of the SNN were set to 15 and 45, respectively. In SNN, the weight based on the number of common neighbors was set to 0.1 for over 90%, 0.5 for over 70%, and k when in the absence of shared neighbors.

First, fault detection is performed to identify the faults. The false alarm rate (FAR) indicates the result of detecting samples in the section after the faulty section of the query vector as faults and is used to compare the fault detection performance of each model. The fault detection indices of the PCA, DPCA, ICA, and DICA are the squared prediction error (SPE), and LOF and SNN are the LOF and D^2 , respectively. LOF is a method of detecting failures using the density of neighbors selected in kNN [24]. It was used as a comparison model, because it can detect faults more effectively than the conventional kNN.

Table 1 compares the FAR of the proposed and conventional methods for single and multiple faults. Among conventional methods, DPCA and DICA can effectively detect single faults. The performance of the dynamic-based method was improved compared to PCA and ICA by increasing the number of original variables through lagged variables. LOF resulted in a higher FAR than the other methods, because the normal data were distributed around the faulty data. In the case of multiple faults, the detection index value of the models increases when two variables have faults. Therefore, faults can be detected more easily than a single fault. DICA had a FAR of 0.4 compared to a single fault, which was lower than other conventional methods. LOF has a lower FAR compared to a single fault but still exhibits a higher error compared to the other models. In contrast, the proposed method has a higher performance than conventional models, with an FAR of zero for both single and multiple fault cases. Compared with LOF, faults can be effectively detected even though they are located adjacent to normal data. The proposed method readjusts the distance according to the neighbors shared by each other in the calculated distance value to capture the failure data adjacent to the normal data quickly. Consequently, the results show that the proposed method can effectively detect faults even when the fault size is small compared to conventional methods.

Table 1. Performance indices of the proposed method and comparison methods (lower is better).

Method	PCA (SPE)	DPCA (SPE)	ICA (SPE)	DICA (SPE)	LOF (LOF)	SNN (D^2)
	FAR	FAR	FAR	FAR	FAR	FAR
Single fault	10.2	6.93	25	2.2	26.1	0
Multiple fault	6.93	2.53	48.7	0.4	15.1	0

As mentioned above, fault isolation during process monitoring is performed after fault detection. The occurrence of a single fault ranged from 501 to 1500 samples, and the randomly selected failure variable (p) is 3. Figure 2 shows the fault isolation results for each model for a single fault. The greater the contribution to failure, the darker the color of the graph. Generally, statistical-based methods (PCA, ICA, DPCA, and DICA) suffer from the smearing effect, which causes the contribution color to be attributed to the fault variable even though the other variables are normal. As a result, PCA, DPCA, ICA, and DICA show that the normal variables appear as red bars, similar to the fault variables after failure occurs. The LOF was not affected by smearing; therefore, the color of the normal variable did not change to that of the fault variable. However, it could not accurately identify the fault variables. On the other hand, the proposed method was not affected by the smearing effect and only calculated the contribution of the faulty variables.

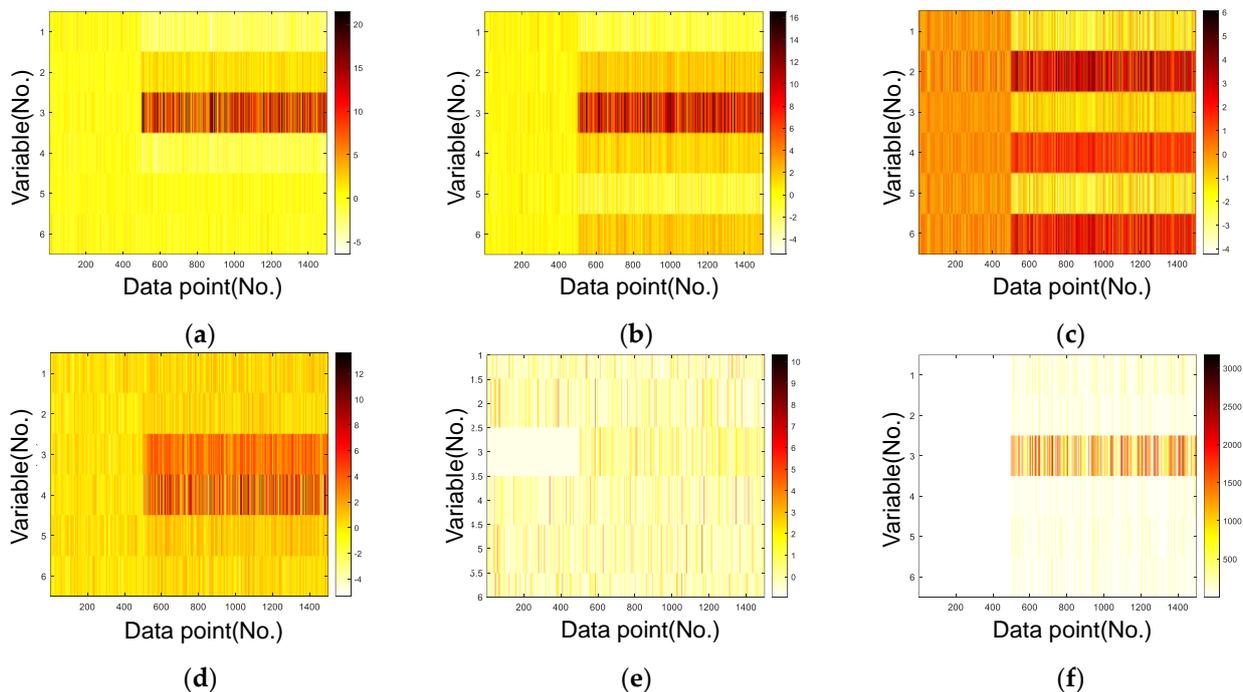


Figure 2. Single fault case. Comparison of fault isolation results: (a) PCA, (b) DPCA, (c) ICA, (d) DICA, (e) LOF, (f) SNN.

The fault isolation results for multiple faults are shown in Figure 3. Multiple faults also occur from the 501st sample to 1500, and the randomly selected fault variables (p and q) are three and four. PCA and DPCA resulted in large contribution values for fault variables 3 and 4. However, because the fault variables affected other variables, other normal variables were identified as failures, as in the single fault cases. ICA failed to identify fault variables under the same conditions as PCA and DPCA (principal components = 5); therefore, all variables were considered faults. DICA calculated the high contribution of fault variables 3 and 4, as well as other variables (5 and 6). Unlike conventional statistical methods, LOF did not show the smearing effect of other variables and did not accurately identify the fault variable. In contrast, the proposed method maintains the contribution values in yellow until the 500th sample, when the contributions of variables 3 and 4 increase, and the graph changes. In particular, we show that only the fault variables can be found without being affected by the smearing effect, compared to other methods. This is because the detection index is calculated in the original space; therefore, it does not affect the other variables. Moreover, it rescales the distance values by assigning weights based on common neighbors. The results demonstrated that the proposed method can effectively identify fault variables while avoiding smearing effects.

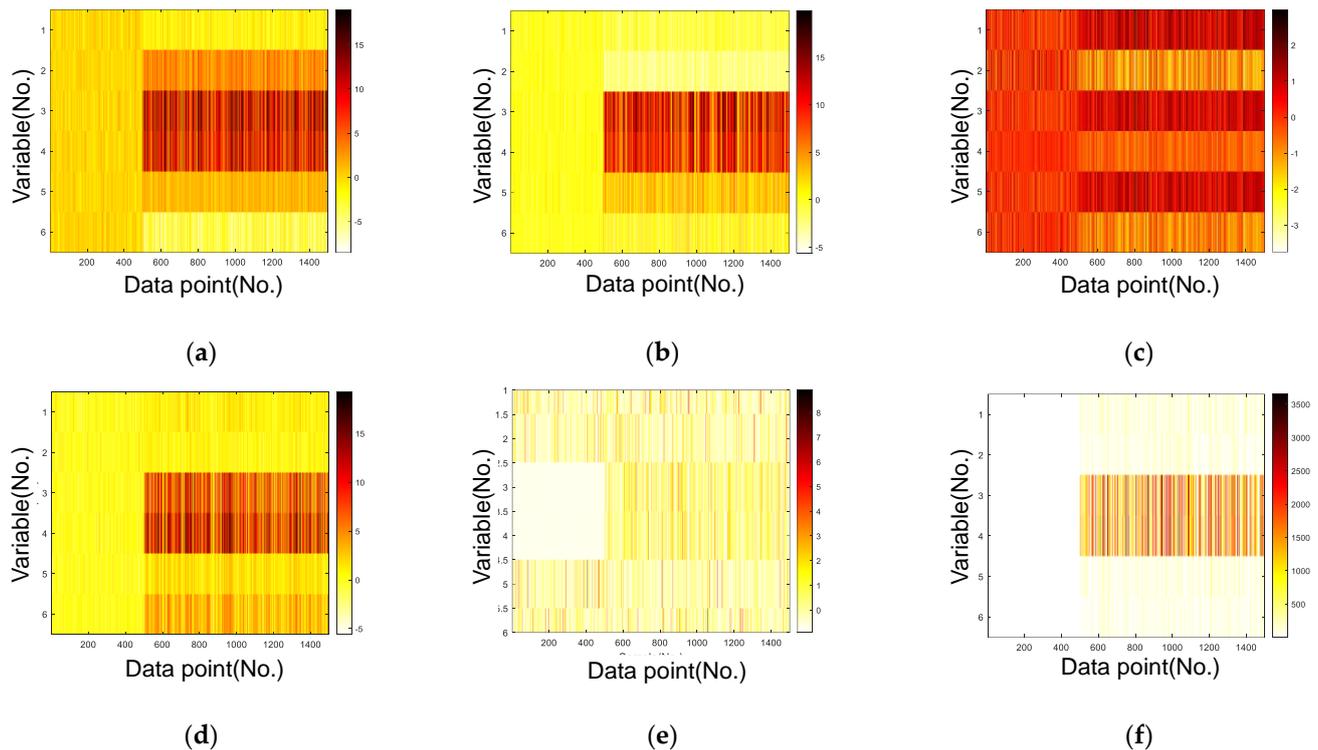


Figure 3. Multiple fault case. Comparison of fault isolation results: (a) PCA, (b) DPCA, (c) ICA, (d) DICA, (e) LOF, (f) SNN.

4. Actual Failure Case Study

In this section, two cases are studied to verify the performance of fault isolation between the proposed and conventional methods in an unplanned shutdown case that occurs in an actual circulating fluidized bed boiler. Section 4.1 describes the structure and data of the circulating fluidized bed boiler and compares the early fault detection time and fault isolation performance before an actual failure occurs. We used CPU Intel i9-9900K and GPU Nvidia RTX 2070 hardware to run these tests. The algorithm was implemented in MATLAB 2021b. The target system used in the experiment was a circulation boiler operating in the Saha-gu industrial complex in Busan, South Korea.

4.1. Circulating Fluidized Bed Boiler Structure

A circulating fluidized bed combustion boiler (CFBC) is a power generation system that produces electrical energy using biomass and domestic waste fuel. They are widely utilized in small- and medium-sized power generation owing to their low combustion temperatures, low NO_x emissions, and ease of SO₂ reduction [25]. As depicted in Figure 4, a CFBC is largely composed of a combustion furnace, superheater, reheater, cyclone, and an economizer. Unlike other power generation systems, CFBCs are gaining popularity, because they can increase fuel flexibility and combustion efficiency using a fluid medium [26–29]. However, bed materials such as industrial waste, sand, and alkali salt contained in the exhaust gas adhere to the bed material and heat pipe in the boiler, causing erosion, corrosion, and agglomeration [30]. The bed material can be quickly deposited as it passes through the rough areas on the inner surface of the pipe. For example, the KCl(s) in domestic waste fuels react with Cr and Cr₂O₃ oxide films in boiler pipes to produce K₂CrO₄ and Cl(g) films, respectively. This erosive activity in the circulating bed results in problems such as flow disorders, pipe overheating, and waterside and fireside corrosion. Tube-related failures were observed several hours before the boiler was shut down, because the response varied based on the location of the steam leak [31]. Consequently, FDI technologies based on sensors measured in the CFBC are required. In this study, we compare the performance of FDI between the proposed method and conventional methods for cases where unplanned

shutdowns occurred owing to tube holes in the superheater, economizer, and sediment in the furnace.

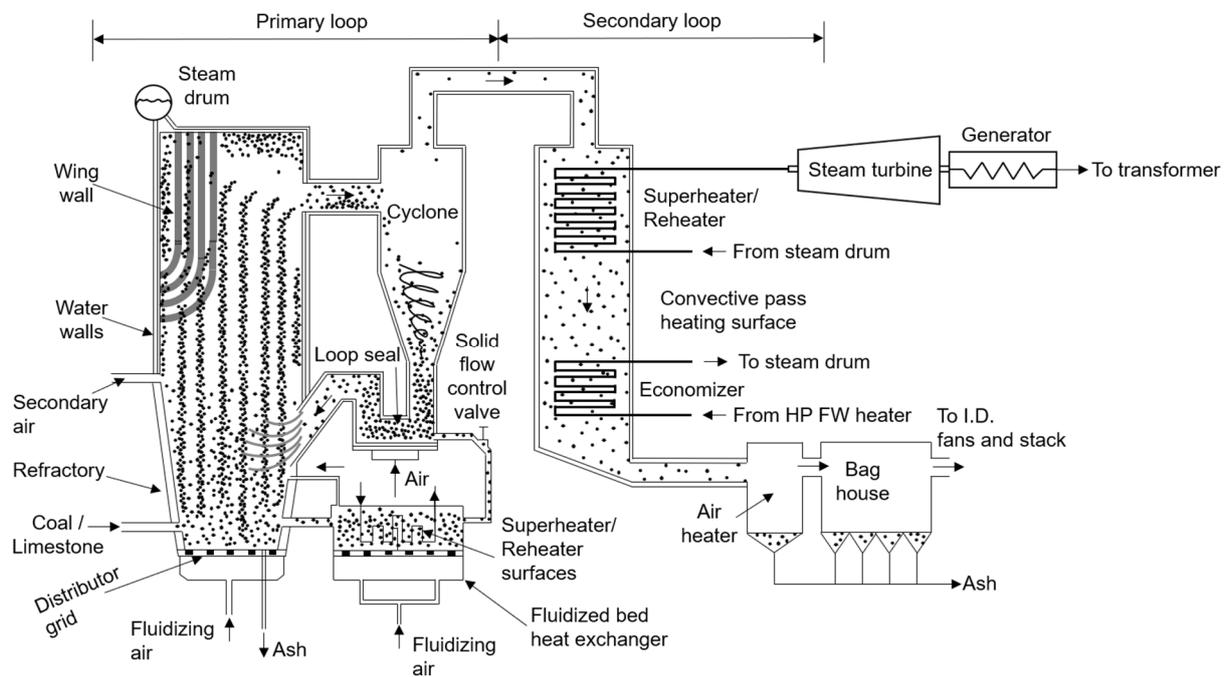


Figure 4. Distribution diagram of CFBC.

The CFBC failed when the boiler was opened and inspected at 2:35 p.m. on 9 September 2020, owing to the increased moisture in the combustion furnace. The main causes of failure include leakage in the tubes of the superheater and economizer and sediment formation in the furnace and cyclone. The training and validation datasets for identifying the cause of failure were 80,000 and 4471, respectively. As presented in Table 2, the 110 boiler and steam-related variables measured at each facility of the circulating fluidized bed boiler were used. Based on the knowledge of the CFBC operators, critical variables were set among the values measured at various facilities. Unlike conventional power generation, the operating modes of the target system change over time based on the steam output. Therefore, the multimode data were normalized to single-mode data using NND before applying the SNN. Figure 5 shows the main causes of unplanned shutdowns, such as leakage in superheaters and economizers and deposits in furnaces and cyclones. As depicted in Figure 5a,b, the tube leaks owing to pipe overheating and roadside corrosion. It was confirmed that the size of the holes gradually increased, affecting other facilities and combustion furnaces. Figure 5c,d show the dislodged deposits in the furnace and cyclone, respectively.

Table 2. Summary of monitored variables for CFBC.

No.	Description	Unit	No.	Description	Unit	No.	Description	Unit
x1	Amount of H ₂ O	%	x38	Steam press. of SCR	mmH ₂ O	x75	Inlet temp. inlet of upper place furnace	°C
x2	Amount of O ₂ in eco.	%	x39	Press. of steam supplied of upper place furnace	MPa	x76	Inlet temp. inlet of furnace	°C

Table 2. Cont.

No.	Description	Unit	No.	Description	Unit	No.	Description	Unit
x3	Diff. of press. furnace and top of cy.	mmH ₂ O	x40	Combustor bed press. of lower furnace feedwater (sensor A)	mmH ₂ O	x77	Inlet temp. inlet of cyclone and boiler front-end	°C
x4	Diff. of press. 2nd and 1st S/H.	mmH ₂ O	x41	Combustor bed press. of lower furnace feedwater (sensor B)	mmH ₂ O	x78	Inlet temp. inlet of cyclone and boiler terminal	°C
x5	Diff. of press. 1st S/H and 2nd eco.	mmH ₂ O	x42	Press. of lower place furnace	mmH ₂ O	x79	Diff. of temp. 2nd and 1st S/H	°C
x6	Diff. of press. 2nd and 1st eco.	mmH ₂ O	x43	Press. of middle place furnace	mmH ₂ O	x80	Inlet temp. inlet of 1st S/H	°C
x7	Diff. of press. of 1st and new eco.	mmH ₂ O	x44	Press. of upper place furnace	mmH ₂ O	x81	Diff. of temp. 1st S/H and 2nd eco.	°C
x8	Diff. of press. of new eco.	mmH ₂ O	x45	Press. between cyclone and boiler	mmH ₂ O	x82	Inlet temp. inlet of 2nd S/H	°C
x9	Output of steam ratio (sensor A)	%	x46	Press. of 1st S/H	mmH ₂ O	x83	Diff. of temp. 2nd and 1st eco.	°C
x10	Output of steam ratio (sensor B)	%	x47	Press. of 2nd economizer	mmH ₂ O	x84	Inlet temp. inlet of 1st eco.	°C
x11	Output of steam ratio (sensor C)	%	x48	Press. of Air pre-heater	mmH ₂ O	x85	Diff. of temp. 1st S/H and new eco.	°C
x12	Steam output of feedwater pipe 1(sensor A)	t/h	x49	Press. of lower supply cyclone (sensor A)	mmH ₂ O	x86	Inlet temp. inlet of 2nd eco.	°C
x13	Steam output of feedwater pipe 1(sensor B)	t/h	x50	Press. of middle place cyclone	mmH ₂ O	x87	Diff. of temp. new eco. and bag filter	°C
x14	Steam output of feedwater pipe 2 (sensor C)	t/h	x51	Press. of middle place furnace	mmH ₂ O	x88	Outlet temp. of air pre-heater terminal	°C
x15	Steam output of fluidized bed material supply	t/h	x52	Press. of lower place furnace	mmH ₂ O	x89	Outlet temp. of dry reactor front-end	°C
x16	Aux. steam output of lower feedwater pipe	t/h	x53	Press. of dry reactor and bag filter	mmH ₂ O	x90	Diff. of temp. cyclone and boiler	°C
x17	Inlet output of feedwater pipe 1	%	x54	Steam flow of air pre-heater and dry reactor	mmH ₂ O	x91	Inlet temp. of cyclone fluidized bed material supply	°C
x18	Outlet output of feedwater pipe 1	%	x55	Output of feedwater ratio (sensor A)	%	x92	Steam output of steam drum	t/h
x19	Outlet output of feedwater pipe 2	%	x56	Output of feedwater ratio (sensor B)	%	x93	Amount of outlet steam flow 2nd S/H	t/h
x20	Steam flow of fluidized bed material supply	t/h	x57	Inlet temp. of dry reactor and bag filter	°C	x94	Amount of inlet steam flow 2nd S/H	t/h

Table 2. Cont.

No.	Description	Unit	No.	Description	Unit	No.	Description	Unit
x21	Steam flow between feedwater pipe 1 and 2	t/h	x58	Inlet temp. of SCR and SGR	°C	x95	Steam drum level of feedwater tank	mm
x22	Steam flow between feedwater pipe 1 and 2	t/h	x59	Inlet temp. of SGR and combustor	°C	x96	Steam drum level of eco.	t/h
x23	Metering bin A outlet conveyor	rpm	x60	Outlet temp. of feedwater pipe 1	°C	x97	Outlet press. of 2nd S/H 1-1	MPa
x24	Diff. press. between feedwater pipe 1	mmH ₂ O	x61	Outlet temp. of feedwater pipe 2	°C	x98	Outlet press. 2nd S/H	MPa
x25	Diff. press. of feedwater pipe 1 (sensor A and B)	mmH ₂ O	x62	Outlet temp. of upper place furnace	°C	x99	Inlet press. 2nd S/H	MPa
x26	Diff. press. between dry reactor and bag filter	mmH ₂ O	x63	Inlet temp. inlet of cyclone and boiler middle point	°C	x100	Outlet press. steam supplied of 2nd S/H	MPa
x27	Sum of steam output of feedwater pipe 1 and 2	mmH ₂ O	x64	Outlet temp. of upper place boiler	°C	x101	Outlet press. of 2nd S/H 1-1	MPa
x28	Furnace press. of feedwater pipe 2	mmH ₂ O	x65	Inlet temp. of feedwater pipe 1 (sensor B)	°C	x102	Temp. of steam supplied of boiler silencer	°C
x29	Furnace press. of feedwater pipe 2 (sensor A)	mmH ₂ O	x66	Inlet temp. of feedwater pipe 2 (sensor B)	°C	x103	Outlet temp. of 1st S/H	°C
x30	Furnace press. of feedwater pipe (sensor B)	mmH ₂ O	x67	Inlet temp. of feedwater pipe 1	°C	x104	Temp. of steam supplied of boiler silencer	°C
x31	Press. of fluidized bed material supply	mmH ₂ O	x68	Inlet temp. of feedwater pipe 2	°C	x105	Output of steam drum	%
x32	Press. of 2nd S/H	mmH ₂ O	x69	Inlet temp. of feedwater pipe 1 (sensor A)	°C	x106	Outlet temp. of 1st S/H	°C
x33	Press. of lower supply cyclone (sensor B)	mmH ₂ O	x70	Inlet temp. of feedwater pipe 1 (sensor A)	°C	x107	Inlet temp. of 2nd S/H (sensor A)	°C
x34	Inlet press. of feedwater pipe 2	mmH ₂ O	x71	Inlet temp. of feedwater pipe 2 (sensor A)	°C	x108	Inlet temp. of 2nd S/H (sensor B)	°C
x35	Press. of air pre-heater and dry reactor	mmH ₂ O	x72	Inlet temp. inlet of fluidized bed material supply	°C	x109	Inlet temp. of 1st S/H (sensor A)	°C
x36	Press of upper place combustor	mmH ₂ O	x73	Inlet temp. inlet of lower place furnace (sensor A)	°C	x110	Inlet temp. of 1st S/H (sensor B)	°C
x37	Press. of SCR terminal	mmH ₂ O	x74	Inlet temp. inlet of lower place furnace (sensor B)	°C			

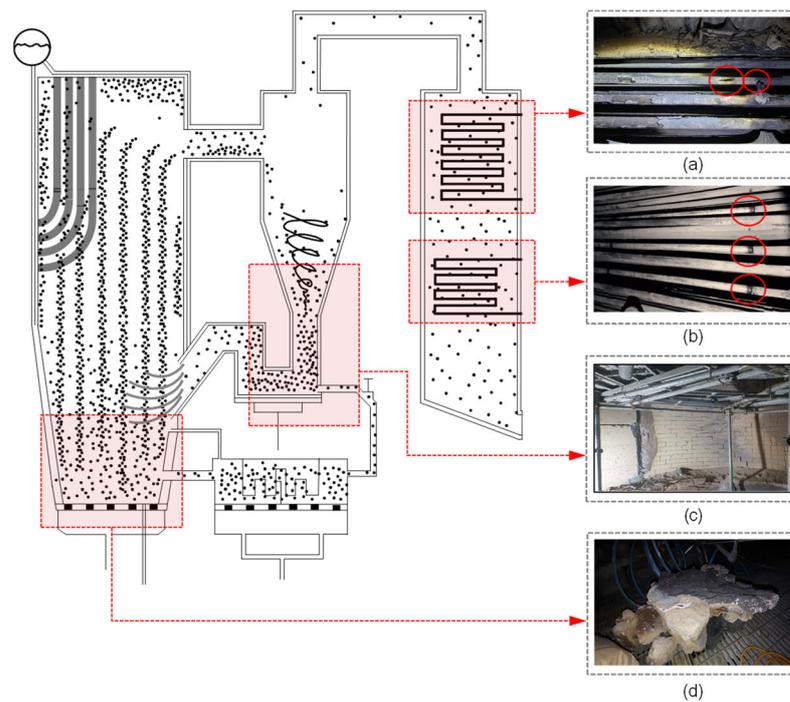


Figure 5. Tube leakages and deposits of CFBC: (a) superheater tube leakage, (b) economizer tube leakage, (c) sediment in cyclone, (d) sediment in furnace.

Figure 6 illustrates the pressure in the furnace and temperature in the feedwater pipe of the CFBC. As described previously, the target system changes its operating modes based on the steam output; therefore, the mean and variance of the data are different for each mode. Each operation mode should not be considered a failure because it contains normal operational data. Therefore, in this study, NND was used to convert and normalize the multiple modes to a single mode. After normalization, the multimodal structure is removed, and each process variable follows a multivariate Gaussian distribution. Figure 6c,d present the results of normalization with NND. The modes are converted into a single mode compared to existing operational modes. In particular, it can be observed that the phenomenon of bouncing values is maintained while changing to a single mode of operation.

Figure 7 shows the fault detection results based on the number of neighbors of the NND and SNN, which are the parameters used in the proposed method. The parameter settings of the proposed method involve fault detection by increasing k . The parameter with the smallest error is selected as the appropriate value for the model. As depicted in Figure 7, when $NND = 12$ and $k = 16$, the fault detection is faster than before the boiler shutdown, and the FAR is lower. However, when k exceeds 20, the calculated D^2 continuously exceeds the threshold. In other words, values greater than 25 were excluded from the model setup, because this proved to be challenging to perform early fault detection. As a result, the numbers of neighbors for NND and SNN were set to 12 and 16, respectively. The threshold value for fault detection was set using the kernel density estimation (KDE). KDE is a widely used method in fault diagnosis for estimating the distribution of normal operation data. The probability distribution of D^2 calculated from the SNN was estimated using the 'ksdensity' function built into MATLAB, and the significant level (α) was set to 0.01.

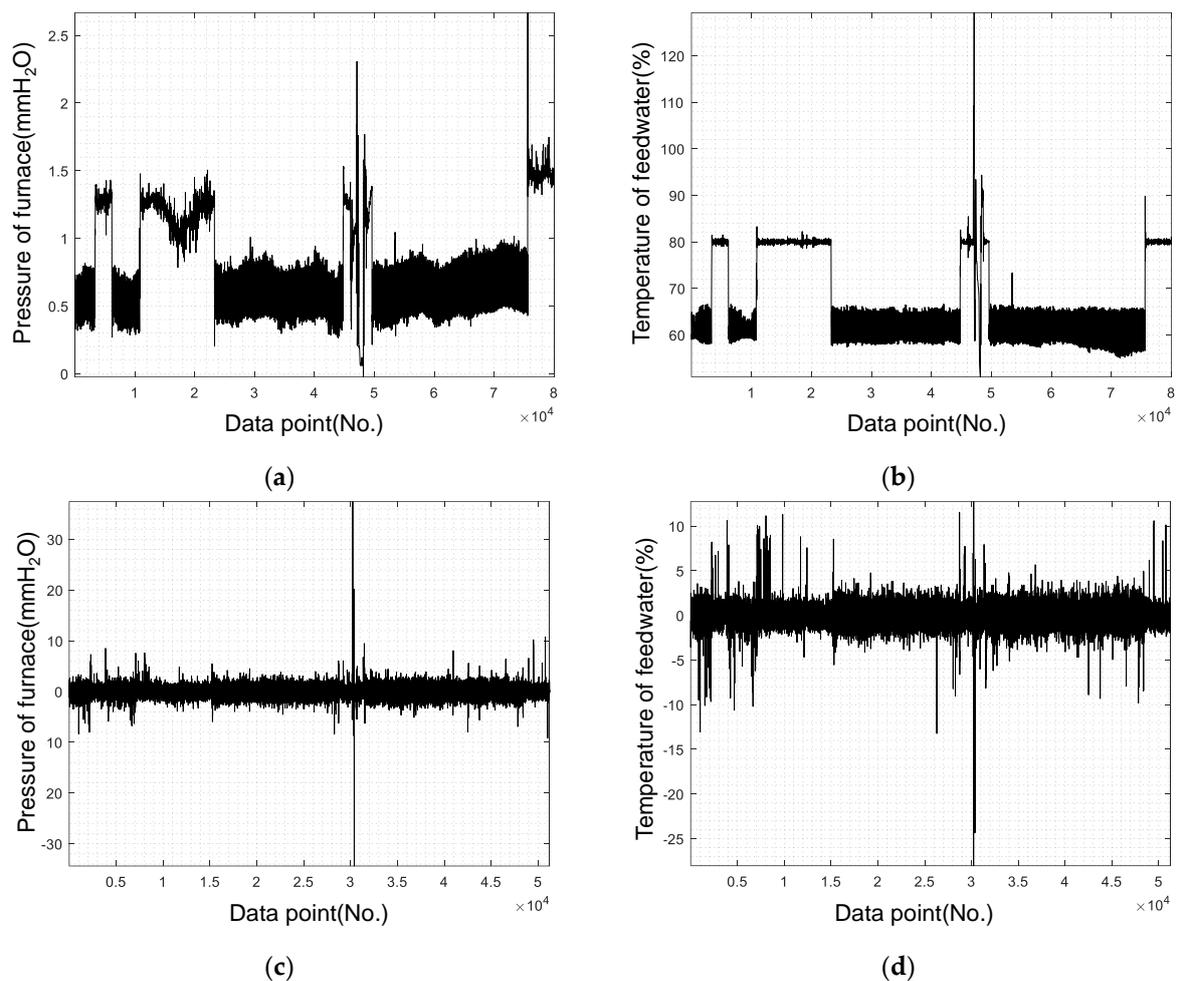


Figure 6. Distribution diagram of CFBC and clinker and clogged nozzle of furnace: (a) pressure in furnace before normalization, (b) temperature in feedwater before normalization, (c) pressure in furnace after normalization, (d) temperature in feedwater after normalization.

Figure 8 shows the CFBC fault detection results of the conventional and proposed methods. The fault detection times were compared based on the unplanned shutdown on 9 September 2020, at 2:35:06 p.m. (data point: 1080). The conventional method uses PCA, DPCA, ICA, and DICA, which are widely used statistical methods in fault detection, and LOF, which is a non-parametric method for detecting faults based on nearest neighbors. Cumulative percentage variance (CPV) was used to determine the number of data reductions for each statistical model, and the threshold value of CPV was set to 90%. The number of neighbors in the LOF was 10. In the proposed method, NND and k are set to 12 and 16, respectively. Fault detection was considered successful when the value derived from the model consistently exceeded a threshold value. Table 3 presents a comparison of the early failure detection times for each model. PCA exceeds the threshold from the 1074th sample until an unplanned shutdown occurs. The fault detection time was 2:34:05 p.m., which detected a failure one minute before the target system stopped. DPCA began to deviate from the threshold for the 998th sample, which was earlier than that of PCA. The fault detection time was 2:06:26 p.m., which was approximately 28 min and 40 s early. For DPCA, the performance improvement over PCA appears to be owing to the increased number of variables that can be utilized. The ICA and DICA failed to detect the fault, because they deviated from the threshold value after an unplanned shutdown. For both methods, the poor performance can be attributed to the fact that the decision is based on the number of principal components selected in PCA and DPCA for fault detection comparison. In general,

the number of principal components in the ICA and DICA is set to the number of principal components determined by the PCA; therefore, an improved method is required. The LOF shows that the query vector is outside the threshold starting from the 994th sample (2:20:56 p.m.). In other words, the LOF detected the fault 14 m 10 s before boiler shutdown. The fault detection result of the proposed method was 12:26:06 p.m., which was approximately 2 h 9 m. In particular, it was confirmed that abnormal signs were detected 1 h 39 m 46 s faster than DPCA, which detected failure first among the conventional methods. Therefore, the proposed method can also be applied to fault detection.

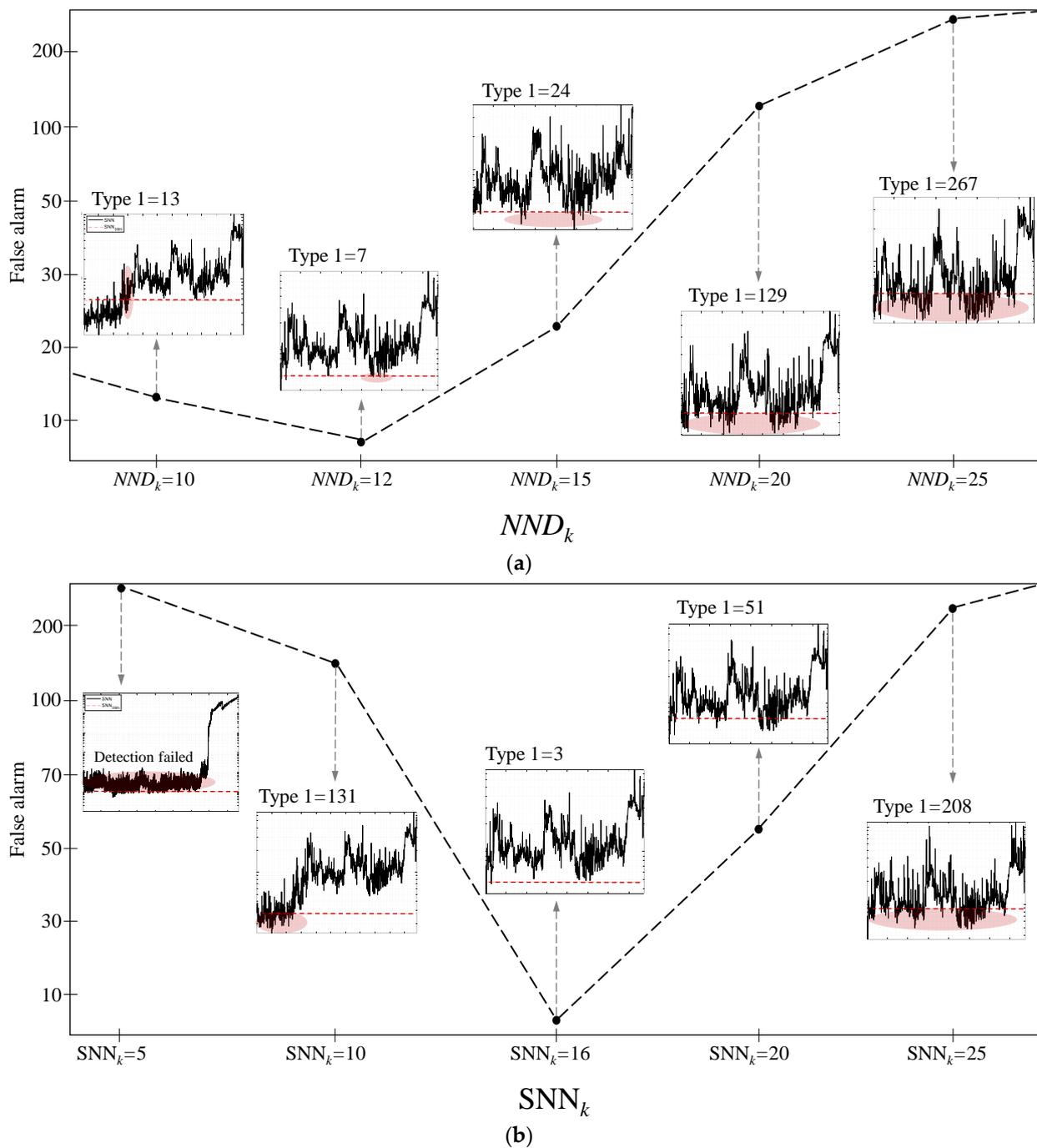


Figure 7. Distribution diagram of CFBC and clinker and clogged nozzle of furnace: (a) FAR according to the neighbors of NND, (b) FAR according to neighbors of SNN.

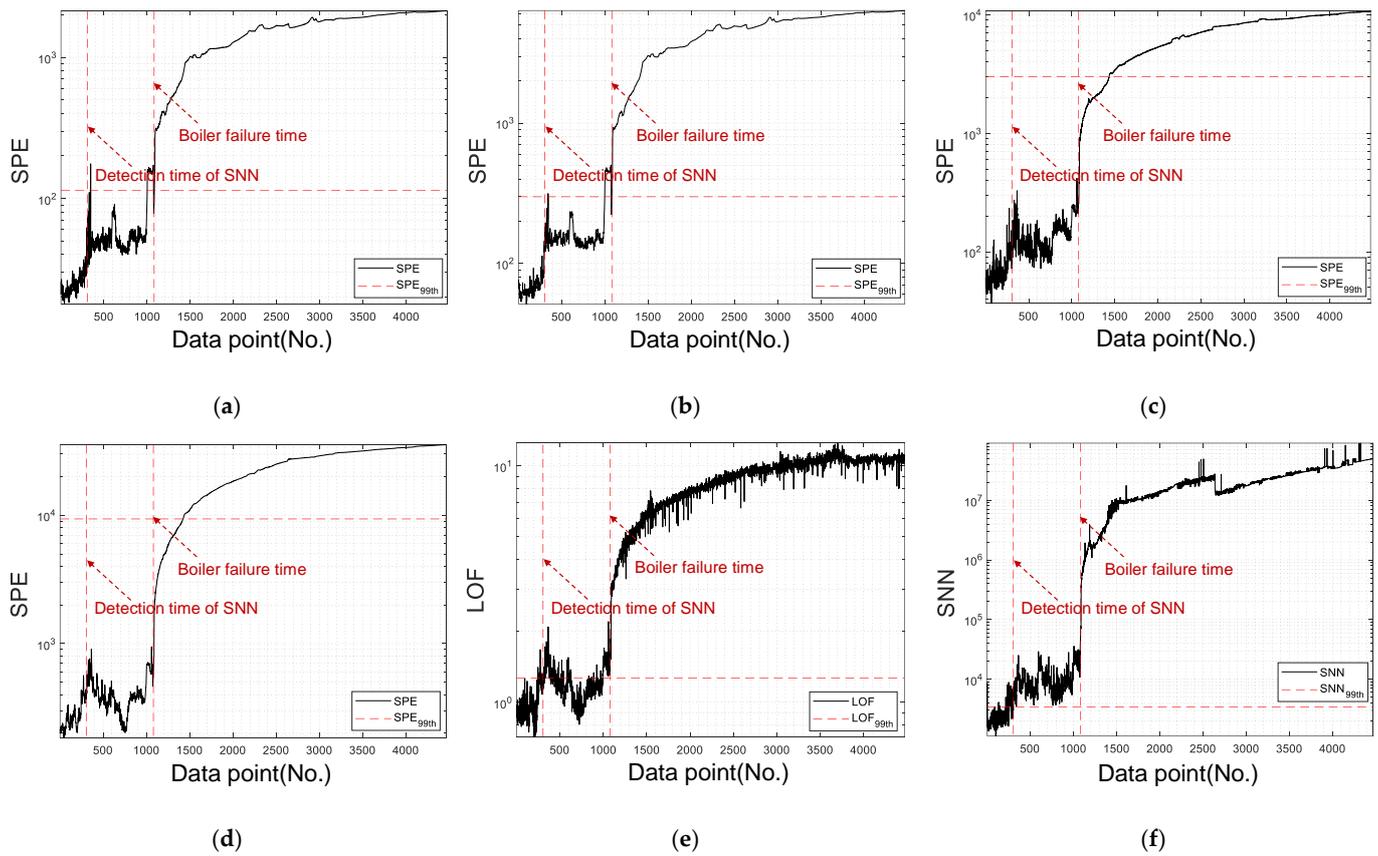


Figure 8. Comparison of the fault detection results: (a) PCA, (b) DPCA, (c) ICA, (d) DICA, (e) LOF, (f) SNN.

Table 3. Fault detection time of the proposed method and comparison methods.

	PCA	DPCA	ICA	DICA	LOF	SNN
Failure time			14 h 35 m 6 s			
Detection time	14 h 34 m	14 h 27 m	-	-	14 h 20 m	12 h 26 m
Early detection time	1 m ago	28 m 40 s ago	-	-	14 m 16 s ago	2 h 9 m ago

Figure 9 shows the contributions of the conventional and proposed methods. The contribution of each variable was compared from the time the SNN detected the earliest fault (12:26:06 p.m.) to the time the unplanned shutdown occurred (2:35:06 p.m.). In general, statistical-based methods (PCA, DPCA, ICA, and DICA) can confirm that the contributions of all variables are high, because the fault variable affects other variables owing to the smearing effect. Consequently, the contribution colors of the normal variables in PCA, DPCA, ICA, and DICA appear in red. Owing to tube leakage and sediment generation, the output volume of the feedwater pipe, temperature of the SGR and combustion passage, and internal temperature of the S/H in the PCA, DPCA, ICA, and DICA contributions were calculated to be high. However, the contributions of the other normal variables also increased, failing to accurately identify the fault variable. Although ICA and DICA failed to detect faults, the contribution of the variables related to tube leakage was calculated to be higher than that of PCA and DPCA. Compared with statistical-based methods, LOF can identify fault variables without suffering from the smearing effect. LOF yielded the highest contribution value for the output of the feedwater pipe as a failure-causing variable owing to tube leakage. In addition, the S/H variable also made a high contribution. LOF can identify variables related to the cause of failure compared with statistical-based methods.

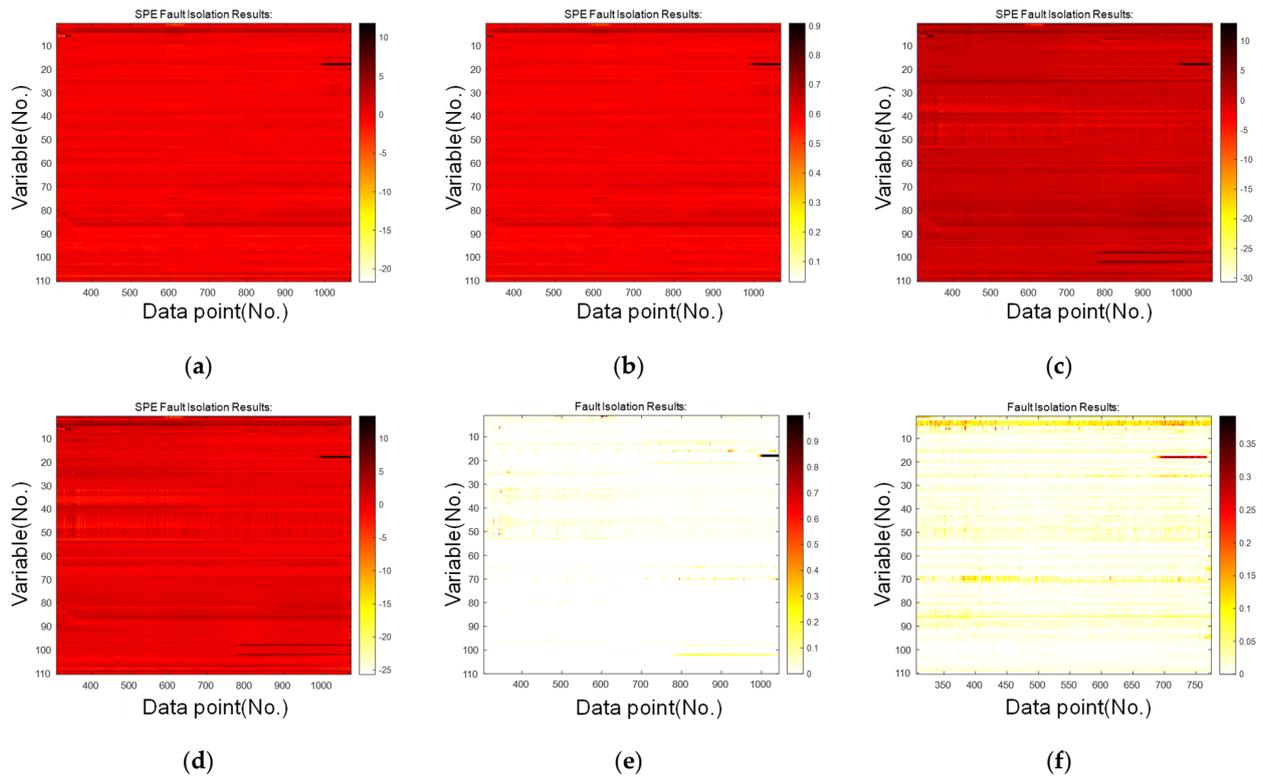


Figure 9. Comparison of the fault isolation results: (a) PCA, (b) DPCA, (c) ICA, (d) DICA, (e) LOF, (f) SNN.

Finally, the proposed method can prevent the smearing effect, which is the same as LOF. LOF had a low contribution of variables other than the output amount of the feedwater pipe, but as observed in Figure 10, the proposed method’s difference in the pressure furnace and top of the cyclone and the difference in the pressure show that the pressure difference between second and first S/H, the temperature inside the economizer, and the variables of the cyclone’s fluidized bed material input can be considered failures. In particular, the proposed method shows the contribution of variables related to failure changes by detecting faults earlier than the other models. In other words, the proposed method can effectively identify the variables related to failures caused by unplanned shutdowns of circulating fluidized bed boilers compared with conventional methods. Furthermore, by applying multimodal normalization, it was shown that the method can be applied to time-varying processes owing to variations in the boiler output.

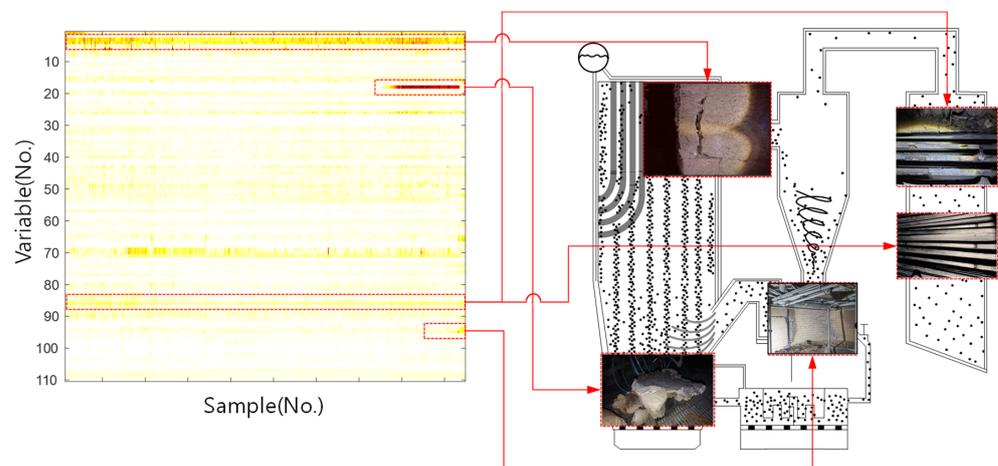


Figure 10. Fault isolation results of the proposed method.

5. Conclusions

In this study, a SNN-based FDI method was proposed to identify the cause of failure in a circulating fluidized bed boiler. First, data normalization to a single mode was performed for FDI of the target system operating in multiple modes. Normalization was performed, and the SNN was used to detect the faults and identify the variables causing the faults. The performance of the proposed method was evaluated by applying the conventional and proposed methods to simulated data with single and multiple faults and unplanned shutdowns in a real circulating fluidized bed boiler. The experimental results demonstrated that the proposed method can effectively detect faults and identify fault variables while avoiding the smearing effect. In the case of circulating fluidized bed failure, the proposed method was able to detect a fault earlier than conventional methods before an unplanned shutdown occurs. It detected the fault 1 h 39 m earlier than DPCA. In fault detection, the earlier a fault is detected, the more likely it is that proper maintenance can be performed to reduce costly downtimes [2,18]. In addition, early fault detection is essential for fault diagnosis, because fault isolation can only proceed when fault detection is performed [32–34]. The proposed method provides early fault detection to identify failure signs in advance. As a result of the fault isolation, the conventional statistical method could not detect the fault properly, because the contribution of other normal variables was calculated to be high owing to the smearing effect. LOF captured the variables related to the feedwater supply owing to tube leakage. However, it does not capture the problems caused by leakage in the economizer tube or deposits in the furnace and cyclone. By contrast, the proposed method captured the relevant variables owing to tube leakage and changes related to sediment formation and shedding. Therefore, the proposed method can be applied to processes operating in multiple modes.

In future studies, we will consider the following two topics: First, we look at the number of neighbors of NND and SNN, which are parameters used in the proposed method, set through repeated experiments; therefore, it takes time to find the appropriate model parameters. To solve this problem, we plan on setting the parameters using the optimization method required for the number setting. Second, weighting according to shared neighbors in a SNN has the disadvantage of having to be set for each target system. Therefore, we will investigate a weight allocation method based on the distribution of shared neighbors.

Author Contributions: S.J. conceived and designed the simulations; J.K. and E.K. analyzed the data; B.K. advised on the whole process of manuscript preparation; M.K. analyzed the data and wrote the paper. The analysis results and the paper were supervised by S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C2009667).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, C.; Gao, X.; Xu, T.; Li, Y. Nearest neighbor difference rule-based kernel principal component analysis for fault detection in semiconductor manufacturing processes. *J. Chemom.* **2017**, *31*, e2888. [[CrossRef](#)]
2. Guo, J.; Tangming, Y.; Yuan, L. Fault detection of multimode process based on local neighbor normalized matrix. *Chemom. Intell. Lab. Syst.* **2016**, *154*, 162–175. [[CrossRef](#)]
3. Ge, Z.; Gao, F.; Song, Z. Two-dimensional Bayesian monitoring method for nonlinear multimode processes. *Chem. Eng. Sci.* **2011**, *66*, 5173–5183. [[CrossRef](#)]
4. Jie, Y. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chem. Eng. Sci.* **2012**, *68*, 506–519.
5. Yu, J.; Jang, J.; Yoo, J.; Park, J.H.; Kim, S. Leakage detection of steam boiler tube in thermal power plant using principal component analysis. In Proceedings of the Annual Conference of the PHM Society 2016, Denver, CO, USA, 3–6 October 2016.
6. Nomikos, P.; MacGregor, J.F. Multivariate SPC charts for monitoring batch processes. *Technometrics* **1995**, *37*, 41–59. [[CrossRef](#)]

7. Wise, B.M.; Gallagher, N.B. The process chemometrics approach to process monitoring and fault detection. *J. Process Control* **1996**, *6*, 329–348. [[CrossRef](#)]
8. Ali, A.; Daneshvar, M. Data driven approach for fault detection and diagnosis of turbine in thermal power plant using Independent Component Analysis (ICA). *Int. J. Electr. Power Energy Syst.* **2011**, *43*, 728–735.
9. Chun-Chin, H.; Su, C.-T. An adaptive forecast-based chart for non-Gaussian processes monitoring: With application to equipment malfunctions detection in a thermal power plant. *IEEE Trans. Control Syst. Technol.* **2010**, *19*, 1245–1250.
10. Wang, G.; Liu, J.; Zhang, Y.; Li, Y. A novel multi-mode data processing method and its application in industrial process monitoring. *J. Chemom.* **2015**, *29*, 126–138. [[CrossRef](#)]
11. Yu, J.; Jang, J.; Yoo, J.; Park, J.H.; Kim, S. A fault isolation method via classification and regression tree-based variable ranking for drum-type steam boiler in thermal power plant. *Energies* **2018**, *11*, 1142. [[CrossRef](#)]
12. Westerhuis, J.A.; Gurden, S.P.; Smilde, A.K. Generalized contribution plots in multivariate statistical process monitoring. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 95–114. [[CrossRef](#)]
13. Conlin, A.K.; Martin, E.B.; Morris, A.J. Confidence limits for contribution plots. *J. Chemom.* **2000**, *14*, 725–736. [[CrossRef](#)]
14. Alcalá, C.F.; Qin, S.J. Reconstruction-based contribution for process monitoring. *Automatica* **2009**, *45*, 1593–1600. [[CrossRef](#)]
15. Alcalá, C.F.; Qin, S.J. Analysis and generalization of fault diagnosis methods for process monitoring. *J. Process Control* **2011**, *21*, 322–330. [[CrossRef](#)]
16. Kourti, T.; MacGregor, J.F. Multivariate SPC methods for process and product monitoring. *J. Qual. Technol.* **1996**, *28*, 409–428. [[CrossRef](#)]
17. Liu, J. Fault diagnosis using contribution plots without smearing effect on non-faulty variables. *J. Process Control* **2012**, *22*, 1609–1623.
18. Zhou, Z.; Wen, C.; Yang, C. Fault isolation based on k-nearest neighbor rule for industrial processes. *IEEE Trans. Ind. Electron.* **2016**, *63*, 2578–2586. [[CrossRef](#)]
19. Xu, H.; Yang, F.; Ye, H.; Li, W.; Xu, P.; Usadi, A.K. Weighted reconstruction-based contribution for improved fault diagnosis. *Ind. Eng. Chem. Res.* **2013**, *52*, 9858–9870. [[CrossRef](#)]
20. Wang, G.; Liu, J.; Li, Y. Fault diagnosis using kNN reconstruction on MRI variables. *J. Chemom.* **2015**, *29*, 399–410. [[CrossRef](#)]
21. Zhang, L.; Lin, J.; Karim, R. Sliding window-based fault detection from high-dimensional data streams. *IEEE Trans. Syst. Man Cybern. Syst.* **2016**, *47*, 289–303. [[CrossRef](#)]
22. Guo, J.; Wang, X.; Li, Y. kNN based on probability density for fault detection in multimodal processes. *J. Chemom.* **2018**, *32*, e3021. [[CrossRef](#)]
23. Basu, S.; Debnath, A.K. *Power Plant Instrumentation and Control Handbook: A Guide to Thermal Power Plants*; Academic Press: Cambridge, MA, USA, 2014.
24. Öngelen, G.; İnkaya, T. LOF weighted KNN regression ensemble and its application to a die manufacturing company. *Sādhanā* **2023**, *48*, 246. [[CrossRef](#)]
25. Rostek, K.; Morytko, Ł.; Jankowska, A. Early detection and prediction of leaks in fluidized-bed boilers using artificial neural networks. *Energy* **2015**, *89*, 914–923. [[CrossRef](#)]
26. Van Caneghem, J.; Brems, A.; Lievens, P.; Block, C.; Billen, P.; Vermeulen, I.; Dewil, R.; Baeyens, J.; Vandecastelle, C. Fluidized bed waste incinerators: Design, operational and environmental issues. *Prog. Energy Combust. Sci.* **2012**, *38*, 551–582. [[CrossRef](#)]
27. Obernberger, I. Decentralized biomass combustion: State of the art and future development. *Biomass Bioenergy* **1998**, *14*, 33–56. [[CrossRef](#)]
28. Khan, A.A.; de Jong, W.; Jansens, P.J.; Spliethoff, H. Biomass combustion in fluidized bed boilers: Potential problems and remedies. *Fuel Process. Technol.* **2009**, *90*, 21–50. [[CrossRef](#)]
29. Kim, M.; Jung, S.; Kim, B.; Kim, J.; Kim, E.; Kim, J.; Kim, S. Fault Detection Method via k-Nearest Neighbor Normalization and Weight Local Outlier Factor for Circulating Fluidized Bed Boiler with Multimode Process. *Energies* **2022**, *15*, 6146. [[CrossRef](#)]
30. Li, Y.S.; Sanchez-Pasten, M.; Spiegel, M. High temperature interaction of pure Cr with KCl. *Mater. Sci. Forum* **2004**, *461*, 1047–1054. [[CrossRef](#)]
31. Oka, S. *Fluidized Bed Combustion*; CRC Press: Boca Raton, FL, USA, 2003.
32. Huang, C.-C.; Chen, T.; Yao, Y. Mixture Discriminant Monitoring: A Hybrid Method for Statistical Process Monitoring and Fault Diagnosis/Isolation. *Ind. Chem. Res.* **2013**, *52*, 10720–10731. [[CrossRef](#)]
33. Dai, X.; Gao, Z. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Trans. Ind. Inform.* **2013**, *9*, 2226–2238. [[CrossRef](#)]
34. Zhong, S.; Langseth, H.; Nielsen, T.D. A classification-based approach to monitoring the safety of dynamic systems. *Reliab. Eng. Syst. Saf.* **2014**, *121*, 61–71. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.