




Article

A Wavelet Transform-Assisted Convolutional Neural Network Multi-Model Framework for Monitoring Large-Scale Fluorochemical Engineering Processes

Xintong Li ^{1,†}, Kun Zhou ^{1,†} , Feng Xue ¹, Zhibing Chen ², Zhiqiang Ge ³, Xu Chen ¹ 
and Kai Song ^{1,*} 

¹ School of Chemical Engineering and Technology, Tianjin University, Tianjin 300350, China; 3014207218@tju.edu.cn (X.L.); kzhou@tju.edu.cn (K.Z.); 3014207238@tju.edu.cn (F.X.); xchen@tju.edu.cn (X.C.)

² Zhejiang Juhua Co., Ltd. Fluor-Polymeric Plant, Quzhou 324004, Zhejiang, China; czb20200430@sina.com

³ State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, Zhejiang, China; gezhiqiang@zju.edu.cn

* Correspondence: ksong@tju.edu.cn

† These authors contributed equally to this work.

Received: 8 October 2020; Accepted: 16 November 2020; Published: 17 November 2020



Abstract: The barely satisfactory monitoring situation of the hypertoxic fluorochemical engineering processes requires the application of advanced strategies. In order to deal with the non-linear mechanism of the processes and the highly complicated correlation among variables, a wavelet transform-assisted convolutional neural network (CNN) based multi-model dynamic monitoring method was proposed. A preliminary CNN model was first trained to detect faults and to diagnose part of them with minimum computational burden and time delay. Then, a wavelet assisted secondary CNN model was trained to diagnose the remaining faults with the highest possible accuracy. In this step, benefitting from the scale decomposition capabilities of the wavelet transform function, the inherent noise and redundant information could be filtered out and the useful signal was transformed into a higher compact space. In this space, a well-designed secondary CNN model was trained to further improve the fault diagnosis performance. The application on a refrigerant-producing process located in East China showed that not only regular faults but also hard to diagnose faults were successfully detected and diagnosed. More importantly, the unique online queue assembly updating strategy proposed remarkably reduced the inherent time delay of the deep-learning methods. Additionally, the application of it on the widely used Tennessee Eastman process benchmark strongly proved the superiority of it in fault detection and diagnosis over other deep-learning methods.

Keywords: fluorochemical engineering processes; convolutional neural network (CNN); wavelet transform; fault detection and diagnosis (FDD); deep learning

1. Introduction

As high performance and value-added products, fluorides and fluorocarbons are extensively applied in various fields of industry and our daily lives, such as medicine, chemical engineering, nuclear industry and so on. Due to the hypertoxic characteristics of octafluoroisobutylene, hydrofluoric acid and other raw materials and intermediate products of fluorochemical engineering processes, it is critical to secure their safe use for public safety and environmental protection. Therefore, it is very important to improve the monitoring performance not only to detect faults but also to diagnosis them correctly. However, the corrosion, aging, fouling and other changes of important parts or

equipment in a fluorochemical engineering process make it a strong time-varying process. Such strong time-varying characteristics, as well as complicated non-linear mechanisms make the performance of regular fault detection and diagnosis (FDD) methods barely satisfactory. In practice, the major safety control strategy is still over maintenance based on the basic control system. Consequently, advanced monitoring methods are urgently needed.

Although, multivariate monitoring methods (including principal component analysis (PCA) [1], partial least squares [2], independent component analysis [3] and so on) and shallow learning methods (including support vector machine (SVM) [4], artificial immune system [5] and artificial neural network (ANN) [6]) are still the primary monitoring methods for large-scale industrial processes, their limited learning ability makes their drawbacks more and more obvious with the ever expanding engineering processes. Additionally, the fast-developing computer techniques (i.e., computing speed, storage capacity) make it increasingly feasible to apply deep-learning methods in monitoring practices. Under such circumstances, over the past few years, deep-learning methods have attracted more research attention and have shown better performance than other data-driven methods in many fields.

Deep-learning methods with multilayer nonlinear structure emphasize the importance of extracted features layer by layer. Hinton and Salakhutdinov [7] proved that the features learned by an ANN with multiple hidden layers had a more essential characterization of the data, which contributed to better visualization or classification performance. At the same time, the “layer-wise pre-training” was suggested to effectively overcome the difficulty of training deep neural networks. These proposals inspired a new wave of deep learning in both academic and industrial fields, and gradually developed the deep artificial neural network [7], deep belief network (DBN) [8], deep convolutional neural network (CNN) and so on.

The convolutional neural network (CNN) was first proposed in the late 1980s, to process data that comes in the form of multiple arrays [9]. Since then, it has had great success in object detection and recognition in the computer-vision domain [10]. The architecture of a CNN is analogous to that of the connectivity pattern of neurons in the human brain and was inspired by the organization of the visual cortex [11]. Individual neurons respond to stimuli only in a restricted region of the visual field known as the receptive field. A collection of such fields is overlapped to cover the entire visual area. Correspondingly, CNN consists of a data input layer, convolutional layers, pooling layers, a fully connected layer and an output layer [12]. The convolution layers locally sample the input feature maps while the role of pooling layers is to compress parameters. Then, the fully connected layer learns a possibly non-linear function for classification. The unique weight sharing and sparsity of connections of CNN greatly improve the model generalization ability and reduce the number of parameters need to be optimized [10]. With the improvement of computer performance and the widespread acceptance of deep learning theory, the investigation on CNN has been intensified by degrees. A series of representative CNN architectures excel in computer vision and object recognition: AlexNet [13], ZFNet [14], GoogLeNet [15] and ResNet [16] and so on.

According to the characteristics of the input data on different time axes, the parameters of CNN are adaptively adjusted to realize the dynamic iteration of the model training. In other words, the data process procedure of CNN is a way to realize dynamic modelling. Therefore, CNN is very suitable for dynamic issues like fault detection and diagnosis (FDD) of industrial processes [17,18]. Due to the aforementioned characteristics of a fluorochemical engineering process, unfortunately, the performance of such a regular CNN could not meet the requirements on both fault diagnosis rates and the computational cost (especially the time delay caused by model training and updating).

On the other hand, wavelet transform (WT) and other data preprocessing methods hold high potential for improving the FDD performance by filtering noise and transforming data [19]. Compared with Fourier analysis (FA) [20] and empirical mode decomposition (EMD), WT is a much faster and more effective method in industrial applications [21]. In recent studies, attempts have been made to imbed wavelet transform into CNN framework [22]. Although the purpose is on improving the efficiency of image processing, it also provides a reference for industrial applications.

Therefore, a wavelet transform-assisted convolutional neural network-based multi-model (WCNN) dynamic monitoring method was proposed to improve the FDD performance as well as to balance the conflicting requests on fault diagnosis rates and the computational cost. The application of it in a R-22, a refrigerant-producing process of a fluorochemical engineering factory located in East China, proved the theoretical considerations.

The rest of the article is structured as follows: Section 2, a brief introduction of the background and related existing methods; Section 3, a description of the proposed WCNN method in details; Section 4, the results and comparisons on the FDD performance of R-22 producing process and the Tennessee Eastman (TE) process benchmark obtained by the proposed method and related deep-learning methods; Section 5, conclusion.

2. Background and Methods

2.1. Brief Introduction to R-22 Refrigerant Producing Process

R-22, also known as HCFC-22, is one of the major fluorides. Although the application of R-22 as a refrigerant or a propellant is controversial, the output of R-22 has shown a steady-state growth because it has been proved to be an indispensable material for tetrafluoroethylene (TFE, primarily used for polytetrafluoroethylene resins, copolymers and food product aerosols) and other fluoropolymer products.

R-22 is produced by the reaction of AHF (anhydrous hydro fluoric acid) and chloroform and is purified by water and alkali to remove residual HCL and HF. Correspondingly, the main operating units of R-22 producing procedure include a feed, a reactor, two rectifying columns, a water scrubber and a separator as shown in Figure 1.

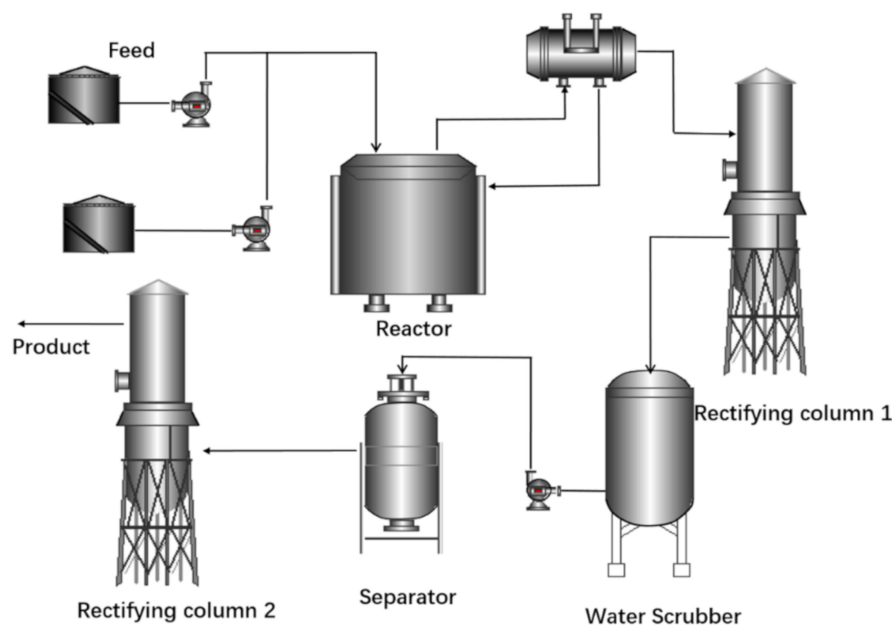


Figure 1. The producing process of R-22.

All materials and byproducts like AHF, HCL and HF become intensely corrosive when they dissolve in the water in the air. Even tiny amounts leaked into the environment can cause terrible damage to equipment and workers. Therefore, it is vital to public safety and environmental protection to improve the performance of the monitoring system to secure the safety of the R-22 producing process. However, the complicated non-linear relationship among variables makes the performance of traditional FDD methods far from satisfied. In addition, the corrosion, aging, fouling and other changes of important parts or equipment in a R-22 producing process make it a strong time-varying process. Such

complicated non-linear and time-varying characteristics of it require sophisticated monitoring methods. Therefore, we proposed a wavelet transform-assisted convolutional neural network multi-model based dynamic monitoring method to improve the fault detection and diagnosis performance for a R-22 producing process and other fluorochemical or chemical engineering processes.

2.2. Brief Introduction of Wavelet Transform Algorithm

The wavelet transform (WT) algorithm, originated from the idea of dilation and translation, analyzes signals using varying scales in time domain and frequency domain [23]. Due to the advantage of multi-resolution analysis, WT has already shown its tremendous superiority in data preprocessing in many industrial applications [23,24]. Traditionally, the wavelet transform can be categorized as continuous wavelet transform, discrete wavelet transform (DWT) and wavelet packet transform [23,25]. The DWT method simplifies the transformation process while still providing a very effective and precise analysis. Therefore, it was adopted as a data preprocessing tool in our method.

According to the DWT method, a time series $x(n)$ can be decomposed into a finite summation of shifted wavelets in different scales using Equation (1) [26,27]:

$$x(n) = \sum_j \sum_k c_{jk} \Psi(a_0^j nT - \tau) \quad (1)$$

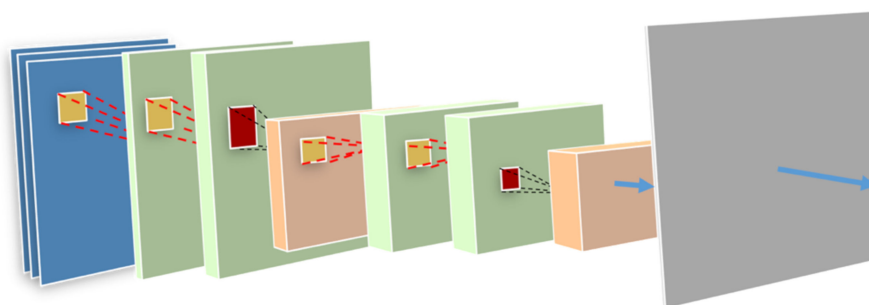
where n represents the n th point of the discrete signal, c_{jk} is a set of wavelet coefficients, $\Psi(a_0^j nT - \tau)$ denotes the wavelet on j th scale shifted by the proper time shifts τ , T represents discrete time, $a_0 = 2$ for the dyadic wavelet.

A signal will be decomposed by DWT into two ancillary signal shape components: the approximation component A and the detail component D. Component A comprises the large-scale and low-frequency component of the signal, while component D represents the small-scale but high-frequency component.

2.3. Brief Introduction to Convolutional Neural Network (CNN)

A convolutional neural network (CNN), a typical deep-learning algorithm, was inspired by biological processes in that the connectivity pattern between neurons resembles the organization of the animal visual cortex [11]. The superiority of it in non-linear supervised learning applications have been proved [28]. Recently, because of the fast development of computer techniques, CNN has also been applied in process monitoring [17,18,29]. Unlike PCA and other traditional methods, which need auxiliary methods for further fault diagnosis, CNN can finish fault detection and fault diagnosis in one step. More importantly, the convolutional functions (also called filters) make CNN good at dealing with dynamic problems [30,31]. It is consequently suitable for continuous process FDD.

The common CNN architecture, shown in Figure 2, consists of an input and an output layer as well as multiple hidden layers. These hidden layers typically consist of a series of convolutional layers, pooling layers, fully connected layers and normalization layers.



Input-Convolutional-Convolutional-Pooling-Convolutional-Convolutional-Pooling- fully connected layer

Figure 2. An example of the structure of the convolutional neural network (CNN).

The essential function of a convolutional layer is the feature extraction, which is the most important function of CNN. The common convolution kernels are squares (i.e., 3×3 , 5×5 and so on). They extract the variable features from rows and columns uniformly. The pooling layer after the convolution layer simplifies the calculation process to increase the speed of network learning and to avoid overfitting. For pattern recognition applications, the fully connected layer is performed to take the feature map as input, which are extracted from the convolutional layer and the pooling layer, and the sample category as output. The dropout function is used to further prevent overfitting [13]. The Adam algorithm [32], the advanced stochastic gradient descent (SGD) algorithm, and other techniques [33,34] are also used to train CNN.

3. The Proposed Wavelet Transform-Assisted Convolutional Neural Network (WCNN)-Based Multi-Model Framework for Dynamic Process Monitoring

As mentioned above, unfortunately, the complicated characteristics of fluorochemical engineering processes make some faults hard to diagnose even by such a high-performance deep-learning algorithm like CNN in one step. A natural idea is to deepen the depth of the network, but the problem is that the computing cost is also greatly increased correspondingly, which in turn increases the challenges of applying CNN in industrial practices. For fault detection and diagnosis issues, detecting all faults with the minimum time-delay is most important. Therefore, a multiple CNN model was proposed by us: (1) the first CNN model is comparatively simple to make sure all faults can be detected as soon as possible and to diagnose faults with simple characters; (2) the second CNN model is more complicated than the first to make sure all faults even those with difficult characters can be diagnosed correctly. Using this two-stage strategy, the requirement on minimizing fault detection and on maximizing the diagnosis accuracy can be fully met and the computational time for CNN models can be well optimized.

On the other hand, considering that the time-varying characteristics of some faults in the fluorine chemical process mainly focus on low frequency, using the wavelet transform to emphasize the feature information under larger scale may improve the FDD performance. Consequently, we proposed a WCNN-based multi-model dynamic monitoring method, whose procedure is listed below and shown in Figure 3. According to background knowledge or use of clustering methods, process faults are classified into three classes: no faults (normal), easy to diagnose (ETD) and hard to diagnose (HTD), then.

1. A preliminary diagnosis model is trained with CNN algorithm to detect all faults with the minimized time-delay. Additionally, for ETD faults, the corresponding diagnosis information is also provided by this preliminary model at the same time for a further response.
2. For HTD faults, a wavelet transform algorithm is introduced to preprocess sampled data by filtering out the inherent noise and transforming it into a compacter space, then a secondary CNN model is trained to diagnose them.
3. For online monitoring, a queue assembly updating method is proposed to reduce the time delay in FDD, whose details will be described in Section 3.3.

As a result, the following distinguishing advantages are brought up by diagnosing faults with the proposed multi-model strategy:

1. The priceless background knowledge can be utilized by labelling faults into ETD and HTD classes.
2. Different kinds of conventional function and structure of CNNs can be used in the preliminary and secondary models, which remarkably reduces the training burden for both models.
3. The conventional functions and structures of CNN in the secondary model can be more specifically designed to further improve the diagnosis accuracies for all HTD faults without causing any time-delay in fault detection.
4. The performance of the secondary CNN model can be improved by introducing a wavelet transform function for data preprocessing.

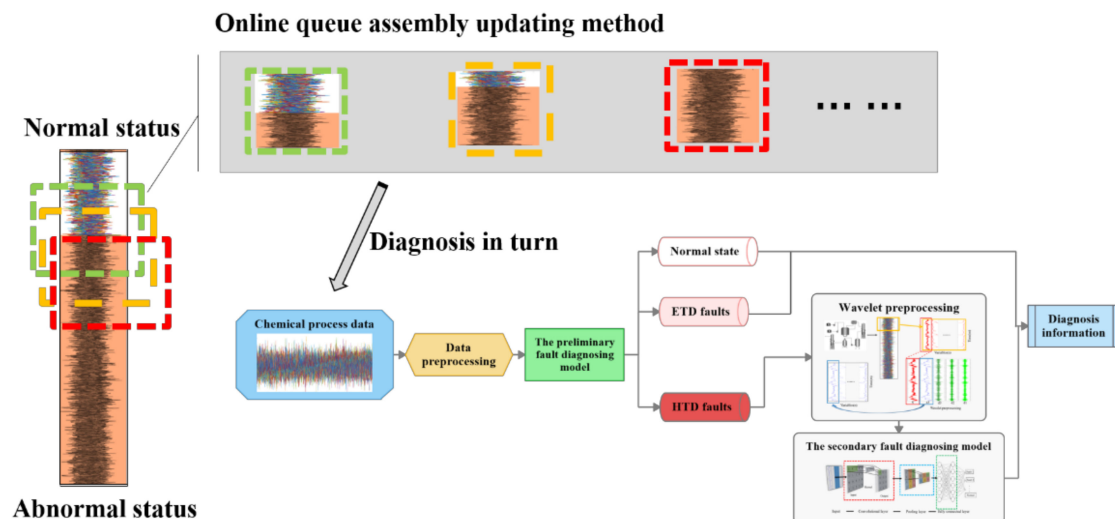


Figure 3. The pipeline of the wavelet transform-assisted convolutional neural network (WCNN) multi-model monitoring method. Three different color dashed boxes represent the data matrices, input into the WCNN model, obtained through the online queue assembly updating method.

3.1. The Preliminary CNN Fault Detecting and Diagnosing Model

The aim of the preliminary fault diagnosing model is to diagnose normal status and ETD faults with a possibly minimum computing burden. In this step, diagnosis information of ETD faults are simultaneously given out. By contrast, all HTD faults are temporally classified in one class.

As mentioned above, for image recognition cases, square convolution kernels are very effective. In our study, however, the input data matrix $X_{m \times n} = [x_1(t), x_2(t), \dots, x_n(t)]$ consists of n process variables sampled at m time points. Obviously, the rows and columns of $X_{m \times n}$ contain completely different kinds of information: the rows contain values of different variables sampled at the same time while the columns contain time serials of values of the same variables sampled at different times. Correspondingly, the correlation among rows (variables) and among columns (time serials of variables) are completely different from each other. Therefore, the square convolution kernels are not suitable anymore. In order to extract information among variables, rectangular kernels along the variable direction (for example 1×3 , 1×5 , and so on) are adopted in a CNN-based monitoring system for the first time [35,36]. The rectangular kernel convolution operation is shown in Figure 4I. The corresponding rectangular kernels used in our model are defined in Equation (2):

$$Y_{mm} = \sum_{v=1}^2 k_{1v} \times X_{m,n+v-1} \quad (2)$$

where k is a rectangular kernel constructed along the variable direction, X and Y are input and output matrix, respectively. v is the number of columns of k . To extract high-dimensional features as comprehensive as possible, multiple convolution kernels are used in one convolutional layer. Correspondingly, the output of a convolutional layer can be calculated as Equation (3):

$$x_j^l = f\left(\sum_{i=1 \dots M} x_i^{l-1} * K_{ij}^{l-1} + b_j^l\right) \quad (3)$$

where f represents the activation function, M is the number of convolution kernels of the j th convolution layer, x_j^l is the j th output feature map of layer l , x_i^{l-1} is the i th input feature map of layer $l-1$, K is the convolution kernel and b_j^l represents the bias of the j th filter.

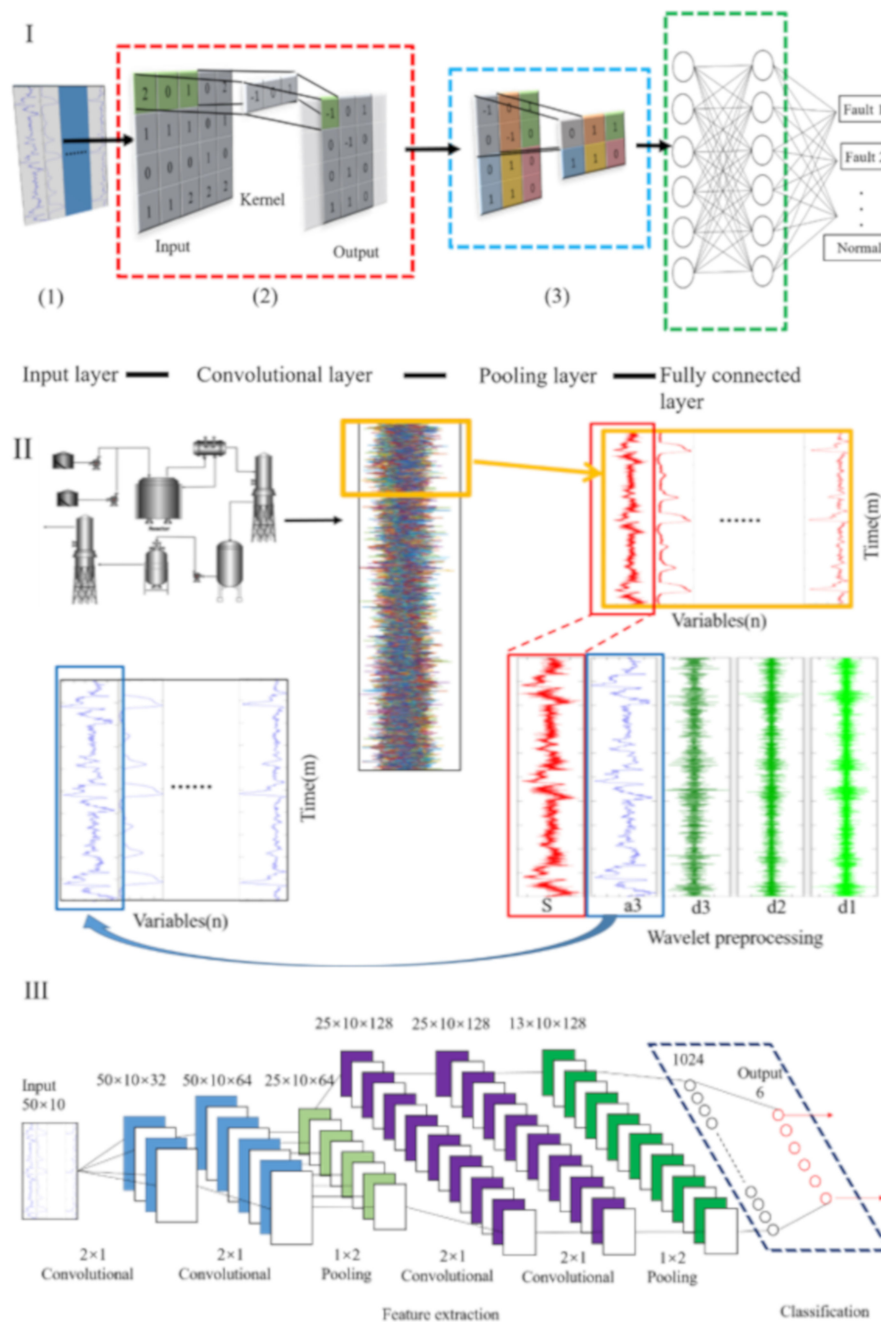


Figure 4. Implementation details of WCNN multi-model monitoring method. (I) details of the preliminary fault detecting and diagnosing model; (II) process of data acquisition and wavelet preprocessing; (III) details of the secondary fault diagnosing model.

Similarly, in order to preserve more time-varying information contained in a time series of variables for better dynamic fault diagnosis performance, the rectangular pooling layer along the time direction (the columns) is used, as shown in Figure 4I. The preliminary model does not need to give out specific diagnosis information for HTD faults, which greatly reduces the training difficulty. Therefore, the preliminary model is designed to have a regular structure to lighten the computational burden.

3.2. The Secondary CNN Fault Diagnosing Model

After the normal status and ETD faults are precisely diagnosed by the preliminary model, HTD faults are still hard to diagnose from each other due to the non-linear relationship, noise, redundant information and the similar fault features inherent in data. Therefore, the discrete wavelet transform (DWT) method is used to filter the raw data first, as shown in Figure 4II. Additionally, multiple rectangular convolution layers and pooling layers are used to contract a deeper and more sophisticated CNN model, shown in Figure 4III, to diagnose HTD correctly.

In WCNN, DWT is only performed on the signal in a fixed time window at every turn to maintain the correlation of the time series in each sample matrix to the greatest extent. The critical problem in wavelet application is the selection of the wavelet base according to the characteristics of the analyzed data. Compactness, vanishing moment, truncation error, computational burden, orthogonality, symmetry and other properties exclude Meyer, Haar and some other wavelet bases out of our consideration [37]. Consequently, DB (Daubechies) wavelet is chosen in our study.

Like other chemical engineering processes, the informative information of R-22 process are mainly contained in a low-frequency signal [38–40]. In addition, the time-varying fault characteristics are also mainly reflected in the low frequency signal. Therefore, the low-frequency component A extracted by DWT method is used as the input of the secondary CNN model.

3.3. Online Queue Assembly Updating Method

Apart from the time consumed by model training, due to the mechanism of CNN and other deep-learning methods, there is a minimum requirement on the size of the input samples. Therefore, it may lead to a time delay of fault detection in online industrial process monitoring. For example, in our study, the number of new samples for prediction must be no less than 50. This would cause an hour's delay even if the abnormal cases can be diagnosed. Because CNN is still widely used in deal with the image recognition, and the online updating issue only matters for industrial applications, it has not received enough attention. Only Zhang suggested to shorten the sampling interval to accelerate the assembly of required new samples [29]. However, shorter sampling interval will lead to higher computational cost. Therefore, we propose a new prediction queue assembly updating method. This method is shown in Figure 3. The detailed procedure is:

1. Initiating the updating matrix with a training matrix including only normal samples;
2. Every time, a new sample is available, adding it to the end of the queue and removing one old sample;
3. Input this matrix to the WCNN model to obtain the FDD prediction information;
4. Repeat steps 2–4.

In this way, it is not necessary to wait until a certain number of new samples are available. Therefore, it may shorten the time delay in FDD. Moreover, this procedure isn't limited to CNN, it can be applied in any deep learning method based monitoring to short the time delay.

4. Application Results and Discussion

4.1. The Monitoring Performance of Fluorochemical Engineering Processes

In order to test the effectiveness of the WCNN based multi-model dynamic monitoring method, it was applied to monitoring the R-22 producing process. All data were collected from a fluorination plant located in East China.

Due to the confidentiality agreement, data from only the reaction R-301, which has the biggest impact on the entire production process, and the relevant process were used. The variables are listed in Table 1. All data were sampled with the sampling interval of 1 min from May to November 2019.

Table 1. Reactor R-301 related variables in R-22 producing process.

No.	Type	Tag Number
1	PV	FIQ-3002
2	PV	TICA-3002
3	PV	WIA-3001
4	PV	LICA-3003
5	CPV	FIQ-3017
6	PV	TRCA-3001A
7	PV	TRCA-3001B
8	PV	TRCA-3001C
9	PV	TRCA-3001D
10	PV	TRCA-3001E

The data of the normal case and five abnormal cases (because they were not serious enough to cause any damage) were selected to verify our WCNN multi-model method. For each case, 8000 samples were used to train the model while 2000 samples were used to test it.

Among these five types of abnormal case, abnormal case 1, 2 and 3 were hard to diagnose even using the regular CNN model. The two-dimensional clustering results of K-means shown in Figure 5I partly showed how badly they were mixed up with each other. Therefore, they were labeled as HTD cases.

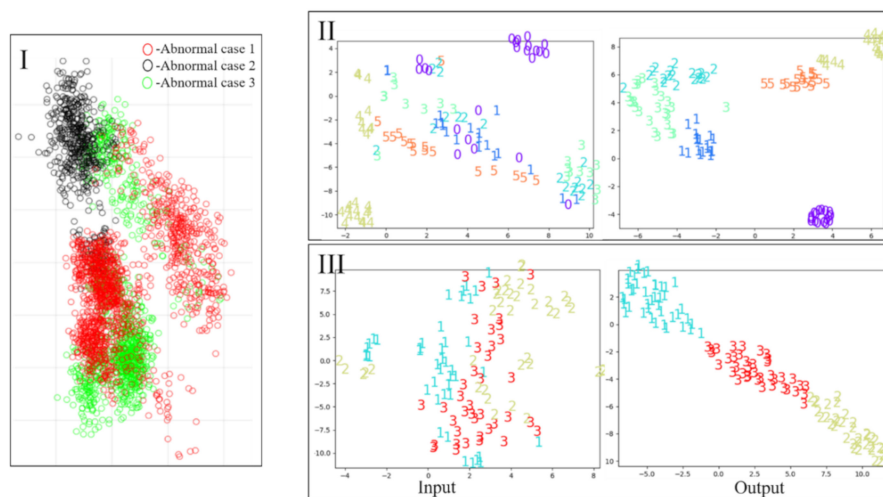


Figure 5. Visualization results of R-22 processes obtained by K-means and t-distributed stochastic neighbor embedding (t-SNE). (I) K-means clustering results. (II) and (III) t-SNE visualization results of the preliminary fault diagnosing model and the secondary fault diagnosing model, respectively.

As mentioned in Section 3.2, the DB wavelet family should be adopted according to the time-varying characteristics of fluoride industrial data signals. It provides compact support and regularity, which means that it has good local performance and can obtain good smoothing effect in signal or image reconstruction [41]. The ordinal number of the DB wavelet family represents its vanishing moment, which should be selected according to the signal to be processed. Basically, a large order of vanishing moment of a DB wavelet can focus on lower frequency to avoid a high-frequency interference [42]. With the consideration for the large-scale time-varying characteristics of HTD cases in a fluorine chemical process, DB3 should be selected as the mother wavelet to highlight the low frequency time-varying characteristics without too much detail information lost.

For both preliminary and secondary CNN models, the input size of one sample matrix was 50×10 , where “50” represented the length of the time series and “10” represented the number of variables. The relevant parameters of both CNN models were described in Section 3.

The major difficulty in designing a convolutional neural network is that there are no guidelines. We have repeatedly tested and optimized the network structures based on research by Hao and Zhao [29]. Table 2 shows the diagnostic accuracy of various structures of the secondary CNN model. The performance was measured by fault diagnosis rate (FDR) as:

$$\text{FDR} = \frac{p}{p + q} \quad (4)$$

where p and q are the number of faults which are diagnosed correctly and incorrectly, respectively [43].

Table 2. The secondary CNN candidate models for fault diagnosis of the R-22 process.

Model	Architecture	FDR (%)
1	Conv(64)-Pool-FC(1024)-FC(3)	88.3
2	Conv(32)-Conv(64)-Pool-FC(1024)-FC(3)	90
3	Conv(32)-Conv(64)-Conv(128)-Pool-FC(1024)-FC(3)	90
4	Conv(64)-Pool-Conv(128)-Pool-FC(1024)-FC(3)	91.7
5	Conv(32)-Conv(64)-Pool-Conv(128) -Conv(128)-Pool-FC(1024)-FC(3)	93.3
6	Conv(32)-Conv(64)-Conv(128)-Pool-Conv(128)-Conv(128)-Pool-FC(1024)-FC(3)	91.7

FDR: fault diagnosis rate.

The fifth model was selected as the final secondary CNN model. The structure of the preliminary CNN was optimized as: Conv(64)-Pool-Conv(128)-Pool-FC(1024)-FC(4). It can be seen that compared with the preliminary CNN model, the secondary CNN model had two more convolution layers and two more pooling layers. The reason for a simpler structure of the preliminary CNN was to save training or updating time in order to detect all faults as soon as possible. But the secondary CNN structure was much more complex than the preliminary one in order to achieve the purpose of accurately diagnosing HTDs.

The kernel and the pooling layers in both CNN models were optimized to 1×2 and 2×1 , respectively. This design can highlight the correlation among different variables and extract the feature information contained in the time series to the greatest extent considering the computational burden. The “padding” parameter of the convolution layer was set to “SAME” to avoid the adverse effect on feature extraction caused by the limitation of convolution kernel size in considering the corner information. Maximum pooling was utilized to obtain the most important features. The “dropout” was set to 0.5 to avoid overfitting.

The features learned by convolutional neural networks are often high-dimensional, which is difficult to observe and understand intuitively. For example, the array size is $13 \times 10 \times 128$ in our study. In order to visualize the classification results, t-distributed stochastic neighbor embedding (t-SNE) [44] was used to map the high-dimensional feature output of the convolutional neural network to a two-dimensional space to visualizing the effect of the convolution process. Visualization results of t-SNE is shown in Figure 5II,III. The digital numbers 0, 1, 2, . . . , 5 represent samples in normal case and abnormal cases 1 to 5, respectively.

From Figure 5II, we can see that all samples were badly mixed up. The samples of abnormal cases 4, 5 (two ETDs) and normal case, were distinguished well by the preliminary CNN model. But the samples of HTD cases were partially overlapped in the feature space. After being preprocessed by DWT, as what Figure 5III shows, they were completely distinguished by the secondary CNN model.

To comprehensively verify the monitoring performance of our proposed method, the monitoring results of these abnormal cases obtained by optimized SVM [45], DBN [46] and the traditional CNN [29] are listed in Table 3. All models were trained and tested on a computer (Windows 10, I7-4980HQ CPU, RAM 16G).

Table 3. The monitoring performance of R-22 process obtained by support vector machine (SVM), deep belief network (DBN), CNN and WCNN.

	FDR-SVM (%)	FDR-DBN (%)	FDR-CNN (%)	FDR-WCNN (%)
Normal case	100	95	100	100
Abnormal case 1	65	70	75	90
Abnormal case 2	90	85	95	100
Abnormal case 3	85	90	85	90
Abnormal case 4	100	100	100	100
Abnormal case 5	100	100	100	100
Average	90	90.83	92.5	96.7

SVM is a shallow learning method. It was not appropriate to include SVM in this deep-learning application study. But since SVM is a widely used non-linear method, it was used as a reference in our study to demonstrate the improvement the deep learning methods made in FDD. Without any surprise, for ETDs, all methods can detect and diagnose them with 100% FDR. But for HTDs, especially for abnormal case 1, the FDR of SVM was only 65%, which was the lowest. By contrast, the FDR of our WCNN was the best at 90%. For the rest two HTDs diagnosis, it could be seen that even traditional CNN, as a deep learning method, was still not very effective. In contrast, WCNN got the highest FDR (100% and 90%), which mainly due to its unique multi-model strategy and the filtering of redundant information through DWT. The average FDR of WCNN was the highest also, 4.2% higher than the traditional CNN and 6.7% higher than SVM and DBN.

Table 4 shows the difference in average time complexity between traditional CNN and WCNN (the whole procedure was repeated five times). To deal with all the faults including ETDs and HTDs, the traditional CNN not only took a longer time for each single epoch, but also had poorer convergence performance. For WCNN, the preliminary model was easier to converge and only needed 30 epochs. The secondary model was targeted to diagnose only HTDs, whose single epoch needed less training time because of the smaller size of data.

Table 4. Model training and inference time of CNN and WCNN in fluorochemical engineering processes.

	CNN	The Preliminary Model	The Secondary Model
Training time for one epoch (s)	2.36	1.77	0.88
Epoch Times of convergence	160	30	200
Total training time (s)	377	53	176
Inference frames per second	30.2	49.3	90.9

The total training time for our multi-model method was 349 s including 53 s for the preliminary model, 120 s for WT and 176 s for the secondary model. All detection and diagnosis information can be available after 349 s, which is 28 s faster than the traditional CNN method. Considering both the consumed time and the performance, we can see that WCNN obtained 6.7% higher FDR with 28 s faster than the traditional CNN method. It was indeed a comparatively large improvement. More importantly, the training time for the preliminary model was only 53 s, while the training time of the traditional CNN was 377 s. The preliminary model in our multi-model framework still can detect HTD faults even though it could not diagnosis them. It means that it only took 53 s to detect all abnormal status and even to diagnose 40% of them. However, with the traditional CNN, after 377 s the abnormal status can be detected and diagnosed.

Apart from the time consumed by model training as discussed above, inference time is another important criterion. Frames per second (FPS), representing the number of inferences made in a second, is used to define inference speed. Bigger FPS means a faster inference speed of a method. In 10 trials, the inference FPS of the traditional CNN was 30.2. The total FPS of WCNN including the preliminary and the secondary models was 31.8 which were slightly better than that of the traditional CNN. However, the FPS of the preliminary CNN model was 49.3, which means that WCNN can detect an

abnormal case 1.6 times faster than traditional CNN. Additionally, the FPS of the secondary CNN model was 90.9, which was three times faster than the traditional one.

For a real industrial process, it is undeniable that we cannot only consider the time of model-making inference. In the R-22 process, the sampling interval is 1 min. In order to meet the requirements of single input matrix, the traditional CNN needs to collect data within 1 h before inputting it for diagnosing. However, WCNN can output the inferential result within 1 min due to its queued updating method, whose details can be found in Section 3.3. In this instance, the unique method of queue updating in WCNN made a great improvement. The effect of this method was verified by abnormal case 4. WCNN can diagnose and trigger an alarm after 37.2 min and update new results within 1 min. It greatly reduced the time delay in FDD compared with the 60 min of the traditional CNN.

4.2. The Monitoring Performance for the Tennessee Eastman Process

To further verify the generalization performance of our method, it was applied to monitor the Tennessee Eastman (TE) process, a widely used simulation benchmark. The TE process, as a simulation program for real chemical processes, can provide massive amounts of simulated industrial data for advanced process control studies. Figure 6 illustrates the diagram of the TE process. It contains five major units: a reactor, a stripper, a condenser, a recycle compressor and a separator. A detailed process description including the process variables and the specific plant-wide closed-loop system can be found in the research of Bathelt et al. [47].

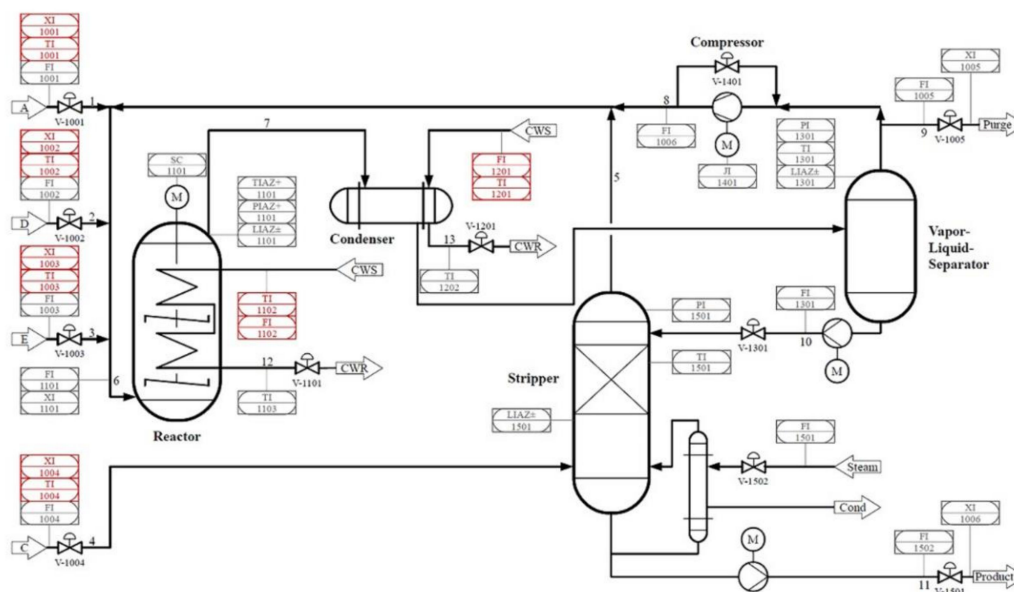


Figure 6. The flowchart of the Tennessee Eastman (TE) process.

The simulator used in our study is based on the revised version which is available online [48]. The variables include 12 process-manipulated variables, 22 continuous process measurements and 19 component analysis measurements. Even though there are 28 process fault types in the revised version, IDV1-IDV20 (vector of disturbance flags) in mode 3 (listed in Table 5) were used in the research of Hao and Zhao [29]. Therefore, they were also used by us for a fair comparison with the published results using similar algorithms.

Table 5. IDVs (vector of disturbance flags) category of TE process.

Category	Process Variable	Type
IDV(1)	A/C feed ratio, B composition constant (stream 4)	Step
IDV(2)	B composition, A/C ratio constant (stream 4)	Step
IDV(3)	D feed temperature (stream 2)	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed loss (stream 1)	Step
IDV(7)	C header pressure loss-reduced availability (stream 4)	Step
IDV(8)	A,B,C feed composition(stream 4)	Random
IDV(9)	D feed temperature (stream 2)	Random
IDV(10)	C feed temperature (stream 4)	Random
IDV(11)	Reactor cooling water inlet temperature	Random
IDV(12)	Condenser cooling water inlet temperature	Random
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condensor cooling water valve	Sticking
IDV(16)	Unknown	Unknown
IDV(17)	Unknown	Unknown
IDV(18)	Unknown	Unknown
IDV(19)	Unknown	Unknown
IDV(20)	Unknown	Unknown

Because two variables in mode 3 of the TE process are constants, only the remaining 51 variables were used for monitoring. The sampling period was set to 50 samples/hr. Each sample matrix contained data sampled in one hour, therefore, the matrix size was 50×51 . The data were collected as follows:

1. To cover the normal data distribution as comprehensive as possible, the simulator ran in a normal state 10 times with 10 different set points, respectively. For each normal state run, the simulator continued to run for 50 h to collect the normal data for each normal state. Therefore, 25,000 ($50 \text{ h} \times 50 \text{ samples} \times 10 \text{ times}$) normal samples were collected in total.
2. For each IDV state, except for IDV6, the disturbance was introduced after 10 h of normal operation. Then the simulator kept running for another 40 h to collect the IDV data. This simulation process was repeated for 10 times with different production set points. Therefore, 20,000 ($40 \text{ h} \times 50 \text{ samples} \times 10 \text{ times}$) samples for each IDV were collected.
3. Because the simulator automatically shut down about 6h after IDV6 was introduced. Only 3000 ($6 \text{ h} \times 50 \text{ samples} \times 10 \text{ times}$) samples were collected for it.

The total number of IDV samples was 383,000 and of normal samples was 25,000. Eighty percent of them were used to train the model while the other 20% were used to test it. IDV3, IDV9, IDV15 and IDV16 were considered as HTD IDVs in this paper, because they were hard to diagnose even using deep-learning methods according to the results of published work [29,43,46].

Because there were 17 types of ETD IDV that needed to be diagnosed, the structure of the preliminary CNN was correspondingly complicated. The structures of both the preliminary CNN and the secondary CNN were (Conv (32)-Conv(64)-Pool-Conv(128)-Conv(128)-Pool-FC(1024)-FC (17 or 4)). The “Padding” parameter of the first two convolution layers were set as “VALID”, and the latter were set as “SAME”. Additionally, db5 was selected as the mother wavelet according to the monitoring performance.

In order to verify the performance of this method, diagnosis results were compared with the best results applied to the TE process obtained by other deep-learning algorithms like DBN and the traditional CNN, which were reported by Hao and Zhao and Zhang and Zhao [29,46]. The results are listed in Table 6.

1. For IDV5, IDV12 and IDV 18, only DBN obtained FDRs for IDV5 and IDV12 testing samples lower than 90%. All other deep-learning methods can diagnose them correctly;

- Even IDV3 was considered as one of HTD IDVs, but the performance of all deep-learning methods were all higher than 90%;
- For IDV9, a HTD IDV, the test performance for neither DBN nor the regular CNN were good enough. But for WCNN, it was improved to 70%;
- For IDV15, another HTD IDV, neither DBN nor the regular CNN could diagnose it. The train performance of WCNN was as high as 98%, but the test performance was only 63%.
- For IDV16, the forth HTD IDV, neither DBN nor the regular CNN could diagnose it. However, both training and testing performance of WCNN were good enough (99% and 81%, respectively).

Table 6. The monitoring performance of the TE process obtained by DBN, CNN and WCNN.

Status Index	FDR (%)					
	Train-DBN	Train-CNN	Train-WCNN	Test-DBN ¹	Test-CNN ²	Test-WCNN
Nomal	-	91.6	97	-	97.8	91
IDV 01	100	99.8	100	100	98.6	100
IDV 02	100	99.6	99	99	98.5	97.5
IDV 03	99	99.6	99	95	91.7	93
IDV 04	98	99.9	100	98	97.6	100
IDV 05	90	99.8	100	86	91.5	100
IDV 06	100	99.8	100	100	97.5	100
IDV 07	100	99.9	100	100	99.9	100
IDV 08	96	98.5	96	78	92.2	91.3
IDV 09	65.5	97.3	97	57	58.4	70
IDV 10	97.5	97.7	98.6	98	96.4	95
IDV 11	97.5	99.5	99	87	98.4	95
IDV 12	85.5	99.2	98.7	85	95.6	91
IDV 13	96.5	97.8	100	88	95.7	98.8
IDV 14	96	99.8	99	87	98.7	95
IDV 15	0	99.7	98	0	28	63
IDV 16	0	91.2	99	0	44.2	81
IDV 17	100	98.8	99	100	94.5	95
IDV 18	100	97	99	98	93.9	94.3
IDV 19	97	99.6	100	93	98.6	98
IDV 20	98.7	97.1	97	93	93.3	95
Average	85.9	98.6	98.8	82.1	88.2	93

¹ [46]; ² [29].

The best average FDR in both training and testing samples obtained by WCNN were the highest ones, which strongly proved its superiority in monitoring.

Using IDV4 as an example, with the queue assembly updating method, the alarm of a fault detection was triggered at the 27.6 min for the first time and was stable after 46.8 min, which was much faster than the 60 min of the traditional CNN.

5. Conclusions

In order to improve the monitoring performance of the non-linear and complicated fluorochemical processes as well as to minimize the time delay in fault detection and diagnosis (FDD) caused by the training and updating of deep-learning models, we proposed a wavelet assisted convolution neural network based (WCNN) multi-model monitoring method. The multi-model diagnosis framework was developed to reduce the computational burden of the model and to integrate the valuable prior knowledge of the black box process monitoring. Rectangular convolution kernels and pooling functions were used in fluorine chemical data to improve the feature extraction ability of WCNN, which can be extended to other industrial data. The wavelet pretreatment algorithm was used to extract the fault characteristics under different frequencies and scales. Furthermore, we proposed a queue assembly updating method for online monitoring to overcome the inherent time-delay problem of deep-learning

methods and to realize rapid FDD of industrial processes. This method has been verified by its application on the monitoring of a R-22 production process, an actual fluorination process located in East China, and by the TE process. The comparison results revealed that the method has promising industrial application prospects.

Author Contributions: Conceptualization, X.L., K.Z. and K.S.; methodology, X.L., K.Z. and K.S.; software, X.L.; validation, X.L., F.X. and Z.C.; formal analysis, X.L. and K.Z.; investigation, X.L.; resources, Z.G.; data curation, X.L., F.X., Z.C. and K.Z.; writing—original draft preparation, X.L.; writing—review and editing, K.S.; visualization, F.X.; supervision, X.C.; project administration, K.S.; funding acquisition, K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Key Research and Development Program of China (Grant No. 2018YFC0808600).

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

References

1. Yin, L.; Parekh, A.; Selesnick, I.W. Stable principal component pursuit via convex analysis. *IEEE Trans. Signal Process.* **2019**, *67*, 2595–2607. [CrossRef]
2. Camarrone, F.; Van Hulle, M.M. Fast Multiway Partial Least Squares Regression. *IEEE Trans. Biomed. Eng.* **2018**, *66*, 433–443. [CrossRef]
3. Feng, L.; Zhao, C.; Huang, B. A slow independent component analysis algorithm for time series feature extraction with the concurrent consideration of high-order statistic and slowness. *J. Process. Control.* **2019**, *84*, 1–12. [CrossRef]
4. Pisner, D.A.; Schnyer, D.M. Chapter 6-Support vector machine. In *Machine Learning*; Mechelli, A., Vieira, S., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 101–121.
5. Rocha, A.D.; Lima-Monteiro, P.; Parreira-Rocha, M.; Barata, J. Artificial immune systems based multi-agent architecture to perform distributed diagnosis. *J. Intell. Manuf.* **2017**, *30*, 2025–2037. [CrossRef]
6. Shahnazari, H. Fault diagnosis of nonlinear systems using recurrent neural networks. *Chem. Eng. Res. Des.* **2020**, *153*, 233–245. [CrossRef]
7. Hinton, G.E. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]
8. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A Fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef] [PubMed]
9. Le Cun, Y.; Jackel, L.D.; Boser, B.; Denker, J.S.; Graf, H.P.; Guyon, I.; Henderson, D.; Howard, R.E.; Hubbard, W. Handwritten digit recognition: Applications of neural network chips and automatic learning. *IEEE Commun. Mag.* **1989**, *27*, 41–46. [CrossRef]
10. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]
11. Fukushima, K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **1980**, *36*, 193–202. [CrossRef]
12. LeCun, Y.; Kavukcuoglu, K.; Farabet, C. Convolutional networks and applications in vision. Proceedings of 2010 IEEE International Symposium on Circuits and Systems, Paris, France, 30 May–2 June 2010; pp. 253–256.
13. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.
14. Zeiler, M.D.; Fergus, R. Visualizing and Understanding Convolutional Networks. *arXiv* **2013**, arXiv:1311.2901. Available online: <https://arxiv.org/abs/1311.2901> (accessed on 17 November 2020).
15. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

17. Hoang, D.-T.; Kang, H.J. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cogn. Syst. Res.* **2019**, *53*, 42–50. [CrossRef]
18. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [CrossRef]
19. Arneodo, A.; Grasseau, G.; Holschneider, M. Wavelet Transform of Multifractals. *Phys. Rev. Lett.* **1988**, *61*, 2281–2284. [CrossRef] [PubMed]
20. Stein, E.M.; Weiss, G. *Introduction to Fourier Analysis on Euclidean Spaces (PMS-32)*; Princeton University Press: Princeton, NJ, USA, 1972.
21. Wu, J.; Li, J.; Guo, L.; Tang, X.; Wang, J.; Yang, L.; Bai, J. The differences analysis on filtering properties of wavelet decomposition and empirical mode decomposition. In Proceedings of the International Conference on Information Technology and Management Engineering (ITME 2011), Wuhan, China, 23–25 September 2011; pp. 23–26.
22. Liu, P.; Zhang, H.; Lian, W.; Zuo, W. Multi-Level Wavelet Convolutional Neural Networks. *IEEE Access* **2019**, *7*, 74973–74985. [CrossRef]
23. Li, B.; Chen, X. Wavelet-based numerical analysis: A review and classification. *Finite Elem. Anal. Des.* **2014**, *81*, 14–31. [CrossRef]
24. Chen, J.; Li, Z.; Pan, J.; Chen, G.; Zi, Y.; Yuan, J.; Chen, B.; He, Z. Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2016**, 1–35. [CrossRef]
25. Graps, A. An introduction to wavelets. *IEEE Comput. Sci. Eng.* **1995**, *2*, 50–61. [CrossRef]
26. Daubechies, I. Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **1988**, *41*, 909–996. [CrossRef]
27. Mallat, S.E.G. A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]
28. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [CrossRef]
29. Hao, W.; Zhao, J. Deep convolutional neural network model based chemical process fault diagnosis. *Comput. Chem. Eng.* **2018**, *115*, 185–197.
30. Li, C.; Zhao, D.; Mu, S.; Zhang, W.; Shi, N.; Li, L. Fault diagnosis for distillation process based on CNN-DAE. *Chin. J. Chem. Eng.* **2019**, *27*, 598–604. [CrossRef]
31. Wang, H.; Liu, Z.; Liu, Z.; Qin, Y. Understanding and learning discriminant features based on multiattention 1DCNN for wheelset bearing fault diagnosis. *IEEE Trans. Ind. Inform.* **2019**, *16*, 5735–5745. [CrossRef]
32. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 17 November 2020).
33. Dauphin, Y.N.; De Vries, H.; Bengio, Y. Equilibrated adaptive learning rates for non-convex optimization. *arXiv* **2015**, arXiv:1502.04390. Available online: <https://arxiv.org/abs/1502.04390> (accessed on 17 November 2020).
34. Zeiler, M.D. ADADELTA: An Adaptive Learning Rate Method. *arXiv* **2012**, arXiv:1212.5701. Available online: <https://arxiv.org/abs/1212.5701> (accessed on 17 November 2020).
35. Chakrabarty, S.; Habets, E.A.P. Multi-speaker DOA estimation using deep convolutional networks trained with noise signals. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 8–21. [CrossRef]
36. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
37. Motard, R.L. *Wavelet Applications in Chemical Engineering*; Springer: Boston, MA, USA, 1994.
38. Bakshi, B.R.; Stephanopoulos, G. Compression of chemical process data by functional approximation and feature extraction. *AIChE J.* **1996**, *42*, 477–492. [CrossRef]
39. Raghavan, V.K. Wavelet Representation of Sensor Signals for Monitoring and Control. Master's Thesis, Oklahoma State University, Oklahoma City, OK, USA, May 1995.
40. Thornhill, N.F.; Choudhury, M.S.; Shah, S.L. The impact of compression on data-driven process analyses. *J. Process. Control.* **2004**, *14*, 389–398. [CrossRef]
41. Cohen, A.; Daubechies, I.; Feauveau, J.-C. Biorthogonal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **1992**, *45*, 485–560. [CrossRef]
42. Daubechies, I.; Heil, C. Ten lectures on wavelets. *J. Comput. Phys.* **1998**, *6*, 1671.

43. Shen, Y.; Ding, S.X.; Haghani, A.; Hao, H.; Ping, Z. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *J. Process. Control.* **2012**, *22*, 1567–1581.
44. Laurens, V.D.M.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
45. Zhang, Y. Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM. *Chem. Eng. Sci.* **2009**, *64*, 801–811. [[CrossRef](#)]
46. Zhang, Z.; Zhao, J. A deep belief network based fault diagnosis model for complex chemical processes. *Comput. Chem. Eng.* **2017**, *107*, 395–407. [[CrossRef](#)]
47. Bathelt, A.; Ricker, N.L.; Jelali, M. Revision of the Tennessee Eastman process model. *IFAC-PapersOnLine* **2015**, *48*, 309–314. [[CrossRef](#)]
48. Tennessee Eastman Challenge Archive. Available online: <http://depts.washington.edu/control/LARRY/TE/download.html> (accessed on 17 November 2020).

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).