

# Enhanced CNN with Global Features for Fault Diagnosis of Complex Chemical Processes

Qiugang Lu <sup>a,1</sup>, and Saif S. S. Al-Wahaibi <sup>a</sup>

<sup>a</sup> Department of Chemical Engineering, Texas Tech University, Lubbock, TX 79409, USA

## Abstract

Convolutional neural network (CNN) models have been widely used for fault diagnosis of complex systems. However, traditional CNN models rely on small kernel filters to obtain local features from images. Thus, an excessively deep CNN is required to capture global features, which are critical for fault diagnosis of dynamical systems. In this work, we present an improved CNN that embeds global features (GF-CNN). Our method uses a multi-layer perceptron (MLP) for dimension reduction to directly extract global features and integrate them into the CNN. The advantage of this method is that both local and global patterns in images can be captured by a simple model architecture instead of establishing deep CNN models. The proposed method is applied to the fault diagnosis of the Tennessee Eastman process. Simulation results show that the GF-CNN can significantly improve the fault diagnosis performance compared to traditional CNN. The proposed method can also be applied to other areas such as computer vision and image processing.

## Keywords

Convolutional neural network, Global features, Fault diagnosis, Chemical process.

## 1. Introduction

Deep learning methods have prevailed in various fault diagnosis applications due to their unique effectiveness in automatic feature extractions directly from raw data Zhang et al. (2020). Research has been reported on using deep learning methods, e.g., deep belief network Shao et al. (2015), recurrent neural network Jiang et al. (2018), auto-encoders Zheng and Zhao (2020), and convolutional neural networks (CNN) Wen et al. (2017). Exemplary applications include fault diagnosis of wind turbine gearbox Jing et al. (2017), rotating machine bearings Zhao et al. (2019), and complex chemical processes Huang et al. (2022). Such deep learning models can extract abstract features from the data and capture non-linearity, thereby exhibiting superior performance than traditional machine learning methods Wen et al. (2017).

Among the variety of deep learning methods that have been explored, CNN has attracted wide attention due to its excellent performance in complex feature learning and classification Zhang et al. (2017). CNN was firstly used by Janssens et al. (2016) for fault detection of rotating machinery. Further advancements include Wen et al. (2017); Zhang et al. (2018) and the references therein, where a number of strategies were proposed for converting time-series data into images. To capture different levels of features, multi-scale CNN has been put forward where different kernel sizes are proposed, e.g., Chen et al. (2021). Along this line, wide kernels have been employed as the first few layers of CNN,

followed by small kernels for improving the feature representation Zhang et al. (2017); van den Hoogen et al. (2020); Song et al. (2022). However, these reported studies mainly focus on the fault diagnosis of rotating machinery where the number of variables is small. Study on CNNs and their variants for fault diagnosis of complex chemical processes still remains limited Wu and Zhao (2018).

Complex chemical processes are featured by high dimensions, with strong spatial and temporal correlations among variables Lu et al. (2019). CNN model-based fault diagnosis for chemical processes has only been preliminarily attempted Wu and Zhao (2018); Huang et al. (2022). The obtained diagnosis performance is still far from satisfaction for real-world applications Shao et al. (2019). Moreover, existing research mainly focuses on extracting *local features* from process data, despite of the efforts on multi-scale CNN to enlarge the receptive field Song and Jiang (2022). For chemical processes, *global features* are also critical due to the complex interconnection of multiple units and thus the widespread coupling between process variables. Specifically, when forming images from multivariate time-series data, variables that are far apart may also possess strong correlations (e.g., see Fig. 3 below). Traditional CNN methods, including multi-scale CNN, cannot directly capture global features and often require deep layers to expand the receptive field to the entire image Song and Jiang (2022). This motivates us to develop a novel CNN architecture that can extract both global and local features in images to improve the fault diagnosis rate.

In this work, we present a novel global feature-enhanced CNN (GF-CNN) for the fault diagnosis of complex chemi-

<sup>1</sup> Corresponding author: Q. Lu (E-mail: jay.lu@ttu.edu).

cal processes. Specifically, in parallel with the convolutional and pooling layers in CNN, we employ a multi-layer perceptron (MLP) network for dimension reduction. That is, the MLP directly maps the vectorized input images to a low-dimensional feature vector, which is then concatenated with the first fully-connected layer of the CNN. Similar CNN-MLP architecture has been employed in Sinitsin et al. (2022); Ahsan et al. (2020). However, those works are towards using CNN and MLP to handle multiple input data types (e.g., images and numerical data), instead of using MLP for dimension reduction to assist CNN to capture global features. Moreover, those works focus on bearing fault diagnosis and COVID-19 detection, in contrast to our work on fault diagnosis of complex chemical processes.

The outline of this paper is as follows. Section 2 presents the fundamentals of CNN relevant to this work. Section 3 describes the proposed GF-CNN in detail, followed by a case study on the Tennessee Eastman process in Section 4. The conclusion is given in Section 5.

## 2. Fundamentals of CNN

For a typical CNN network, convolution and pooling are the two critical operations. In addition, dropout is often employed for regularization to prevent overfitting.

### 2.1 Convolutional Layers

For typical convolutional layers, consider a generic 2D feature map  $x_i^l \in \mathbb{R}^{h \times d}$  on layer  $l \in \{1, \dots, L\}$  and channel  $i \in \{1, \dots, N_l\}$ , where  $h$  and  $d$  are the height and width of the feature map. The output after a convolutional kernel  $k_{i,j}^l$  is shown to be Zhao et al. (2019)

$$x_j^{l+1} = f \left( \sum_{i=1}^{N_l} k_{i,j}^l * x_i^l + b_j^l \right), \quad j = 1, \dots, N_{l+1}, \quad (1)$$

where  $x_j^{l+1}$  is the  $j$ -th feature map in layer  $l+1$ ,  $*$  is the convolution operation, and  $N_{l+1}$  is the number of kernel filters, i.e., the number of channels, in layer  $l+1$ . Commonly used activation functions  $f(\cdot)$  include ReLU, leaky ReLU, or sigmoid function. In this work, we choose to use the ReLU function as the nonlinear activation. When conducting the convolution, a kernel (often square) slides through the entire feature map with a stride  $s$ . With each stride, the convolution operation above is carried out. In addition, if needed, zero padding can be added to the input map to carry its dimension to the new feature maps.

### 2.2 Pooling Layers

Pooling is an operation that is often used after the convolution layers. The purpose of adding the pooling layers is to extract the major features in local regions of the new feature maps after convolutional layers. Therefore, unnecessary details or noise can be filtered out. Moreover, the dimension of the feature maps can be significantly reduced after pooling, and so do the computation time and the number of parameters in the network. Common pooling techniques include

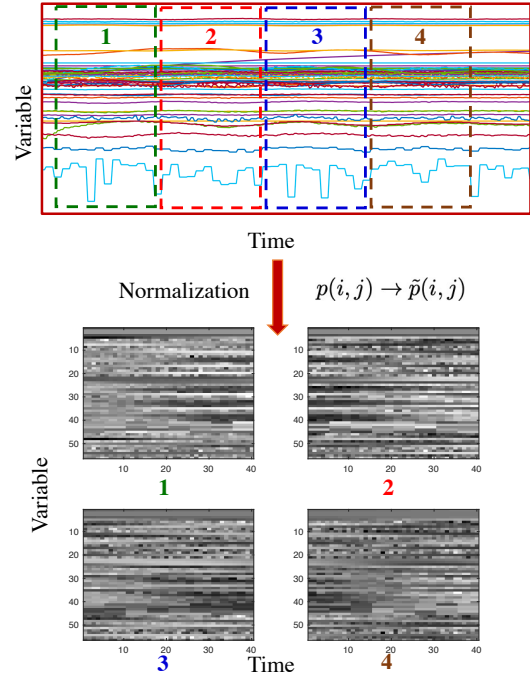


Figure 1: Illustration of signal-to-image conversion. Top: plots of process variables over time; Bottom: the obtained gray images after conversion.

average pooling, weighted pooling, and max pooling. In this work, we adopt the max pooling technique Jing et al. (2017)

$$P_j^{l+1} = \max_{x_j^{l+1} \in S} x_j^{l+1}, \quad (2)$$

where  $S$  is pooling block size,  $P_j^{l+1}$  is the output of the  $j$ -th feature map in the  $(l+1)$ -th layer after pooling. The dimension of the feature map is then reduced  $S$  times.

### 2.3 Fully-connected (FC) Layers

The FC layers are located after all convolutional and pooling layers, serving to classify the extracted features from images. Prior to FC layers, the obtained feature maps are flattened into vectors. For the FC part, the output of the  $j$ -th neuron in the  $l$ -th layer is

$$x_j^l = f \left( \sum_{i=1}^{M_{l-1}} x_i^{l-1} w_{i,j}^l + b_j^l \right), \quad j = 1, \dots, M_l, \quad (3)$$

where  $M_{l-1}$  and  $M_l$  are the input and output dimensions for the  $l$ -th layer, respectively.  $w_{i,j}^l$  is the weight parameter. The output from the last layer of the FC network is then passed into the softmax function for generating the probability of the input image being classified to each class.

### 2.4 Dropout

To prevent overfitting, dropout is a common technique used in CNN. When the dropout is appended to a layer, the output of each neuron will be set to zero with a given probability. For instance, dropout  $p_{dp} = 0.5$  means that each neuron's output value is set to 0 with probability 0.5. The

dropout method can impair the co-adaptation of hidden neurons and disable the contribution of those neurons in both backward and forward propagation Wu and Zhao (2018).

### 3. Proposed Methods for Fault Diagnosis

In this section, we will demonstrate the proposed methods for establishing the GF-CNN model for fault diagnosis.

#### 3.1 Signal-to-Image Conversion

For typical chemical processes, the process variables are measured constantly and thus the dataset is in the form of multivariate time series. For 2D-CNN models to consume these data, it is necessary to convert them into images. We will use similar techniques as in Cao et al. (2018); Wen et al. (2017) for the signal-to-image conversion.

Consider a chemical process with  $n$  variables and  $m$  samples. As shown in Fig. 1, we choose a non-overlapping moving window of width  $w$  to slide in the temporal direction. Each sliding window contains  $w$  measurements of all the  $n$  variables. Then each moving window is considered as an image of dimension  $n \times w$  and the measured values of variables are treated as pixels. To ensure the pixel values ranged between  $[0, 255]$ , the standardization below can be performed:

$$\tilde{p}(i, j) = \text{round} \left( \frac{p(i, j) - p_{\min}}{p_{\max} - p_{\min}} \right) \times 255, \quad (4)$$

$$i = 1, \dots, n, \quad j = 1, \dots, w,$$

where  $p(i, j)$  is the  $j$ -th measurement of the  $i$ -th variable, and  $p_{\min}$  and  $p_{\max}$  represent the minimum and maximum signal values of the underlying window, respectively. The pixel intensity at location  $(i, j)$  is denoted as  $\tilde{p}(i, j)$ . Note that the converted image is a gray image with one channel.

#### 3.2 The Proposed GF-CNN Model

The proposed GF-CNN combines MLP and CNN for classification so that both global and local features can be extracted from the raw data. Fig. 2 shows the structure of the proposed GF-CNN model. Specifically, the top part shows the traditional CNN consisting of multiple convolutional and pooling layers. Denote the mapping from input images to the extracted feature maps in the first FC layer (the purple color in Fig. 2) as:

$$x_{fc,cnn} = f_{\theta_{conv}}(x), \quad (5)$$

where  $\theta_{conv} \in \mathbb{R}^{n_{conv}}$  is the collection of training parameters in all convolutional layers.  $x \in \mathbb{R}^{n \times w}$  is the input image and  $x_{fc,cnn} \in \mathbb{R}^{n_{x,cnn}}$  is the first FC layer feature vector obtained from the CNN. For the bottom part of Fig. 2, we only construct *one layer* of MLP for the purpose of dimension reduction. As a result, the amount of additional parameters to be trained is fairly limited and also far less than that if multiple layers are employed. We further define the mapping from input images to the reduced-order states (the green color in Fig. 2) from the one-layer MLP as:

$$x_{fc,mlp} = f_{\theta_{mlp}}(\text{vec}(x)), \quad (6)$$

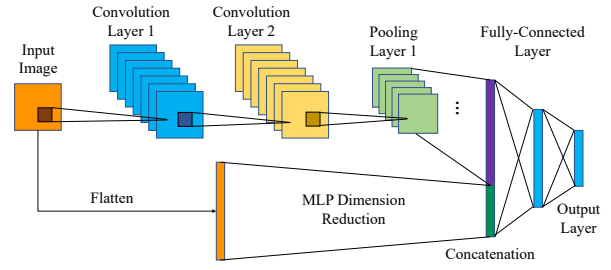


Figure 2: The proposed GF-CNN architecture where the MLP serves to extract global features directly.

where  $\theta_{mlp} \in \mathbb{R}^{n_{mlp}}$  represents the training parameters from the one-layer MLP network for dimension reduction.  $x_{fc,mlp} \in \mathbb{R}^{n_{x,mlp}}$  is the reduced-order state. For the proposed GF-CNN method, vectors  $x_{fc,cnn}$  and  $x_{fc,mlp}$  are concatenated as the first layer of the FC network. The FC network afterwards can be represented mathematically as

$$\hat{y} = f_{\theta_{fc}}(x_{fc}^1), \quad x_{fc}^1 = [x_{fc,cnn}^T \quad x_{fc,mlp}^T]^T \in \mathbb{R}^{n_{x,cnn} + n_{x,mlp}}, \quad (7)$$

where  $\hat{y} \in \mathbb{R}^C$  ( $C$  is the number of classes) is the output of the GF-CNN and  $\theta_{fc} \in \mathbb{R}^{n_{fc}}$  is the training parameter associated with the FC layers.  $x_{fc}^1$  is the concatenation of the flattened feature map from CNN and reduced-order state from MLP.

With the GF-CNN model above, the overall training problem can be formulated as (cross-entropy criterion):

$$\min_{\theta} J(\theta) = -\frac{1}{N} \left[ \sum_{i=1}^N \sum_{c=1}^C \mathbb{I}(y_i = c) \log \mathbb{P}(\hat{y}_i = c | x_i; \theta) \right], \quad (8)$$

where  $\theta = \{\theta_{conv}, \theta_{mlp}, \theta_{fc}\}$  gathers all training parameters from each block;  $y_i$  is the one-hot true label for the input image  $x_i$ ;  $\mathbb{P}(\hat{y}_i = c | x_i; \theta)$  is the predicted probability of classifying  $x_i$  into class  $c$  from the softmax function;  $N$  is the total number of training samples; and  $\mathbb{I}(\cdot)$  is the indicator function.

#### 3.3 Model Complexity

Incorporating a single-layer MLP as the dimension reduction into CNN introduces additional parameters. Here we will show that such a GF-CNN network will not significantly increase the number of parameters (i.e., model complexity) compared with the pure CNN. In general, for process data, the number of dominant features  $n_{x,mlp}$  is usually much less than the data dimensionality Lu et al. (2019). That is,  $n_{x,mlp} \ll n \cdot w$  and usually we also have  $n_{x,mlp} \ll n_{x,cnn}$ . Note that the parameters in all FC layers  $n_{fc}$  consist of those induced by CNN  $n_{fc,1}$  and those by single-layer MLP  $n_{fc,2}$ , i.e.,  $n_{fc} = n_{fc,1} + n_{fc,2}$ . From the argument above that  $n_{x,mlp} \ll n_{x,cnn}$ , we thus have  $n_{fc,2} \ll n_{fc,1}$ . In addition, since we only utilize one-layer MLP for dimension reduction, it can be inferred that, in general, the total number of induced parameters from the MLP is much less than that from the CNN:  $n_{mlp} + n_{fc,2} \ll n_{conv} + n_{fc,1}$ . In other words, adding the global feature extraction will not introduce many parameters. This observation will be verified in the case study below. In summary, compared with traditional CNN, the GF-CNN can boost the performance (will be shown below) by only slightly increasing the model complexity.

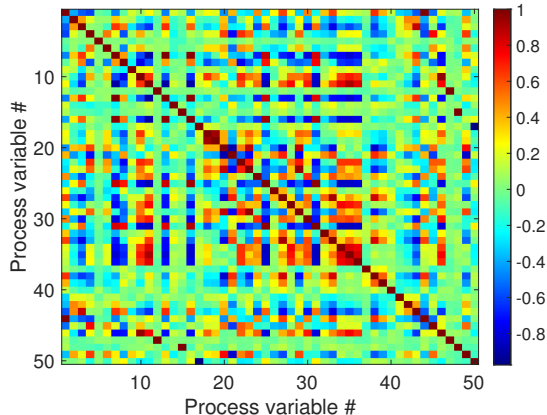


Figure 3: The correlation coefficient colormap for all 50 process variables from TEP.

#### 4. Case Study on the Tennessee Eastman Process

We will use the benchmark Tennessee Eastman process (TEP) as the platform for validating the fault diagnosis performance of our GF-CNN model. The TEP simulates a complex chemical process consisting of five units: a two-phase reactor, a vapor/liquid extractor, a condenser, a compressor, and a stripper. The closed-loop operation data of TEP contains 52 variables, including 41 process measurements and 11 manipulated variables. Normal and different types of faults are programmed into the simulator to generate the corresponding dataset. In this work, we use the large-sample TEP data available from Harvard University Rieth et al. (2017).

##### 4.1 Data Description and Global Features

The selected dataset contains 20 different types of faults. Similar to Wu and Zhao (2018), we drop XMV(5) (compressor recycle valve) and XMV(9) (stripper steam valve) from the data due to their constant values in some simulations. As a result, the number of process variables  $n = 50$ . The sampling period is 3 min. For the training data, each fault class is simulated for 24 hours under 40 random seeds, which gives a total number of 19,200 samples. For the dataset of each fault, we choose the moving window size  $w = 20$  without overlapping, yielding 960 gray-scale images. The total number of images for all 20 faults is 19,200. For the test data, each fault is simulated for 40 hours under 7 different random seeds. The window size is also  $w = 20$  and this gives us 5,600 samples of data, or equivalently 280 images for each fault. Thus, the total number of test images for all 20 faults is 5,600.

The colormap in Fig. 3 shows the correlation coefficients among all 50 variables, corresponding to the 50 rows of each image. It is observed that not all strong correlations exist around the diagonal line. Instead, many strong correlations are dispersed across the entire colormap. In other words, different from traditional images where the textures are often localized, for the images converted from multivariate time series, *non-local features* are common and sometimes dominant. In fact, the order of process variables into the rows of the image can make two variables strongly correlated, although they are far apart.

##### 4.2 Data Preprocessing

Prior to the signal-to-image conversion, both training and test data are normalized. Specifically, for the training data, the overall mean  $\bar{x}_{tr} \in \mathbb{R}^n$  and standard deviation  $\sigma_{tr} \in \mathbb{R}^n$  vectors are computed. Each training and test data sample is normalized as below:

$$\tilde{x}_{tr}^i = \frac{x_{tr}^i - \bar{x}_{tr}}{\sigma_{tr}}, \quad \tilde{x}_{te}^i = \frac{x_{te}^i - \bar{x}_{tr}}{\sigma_{tr}}. \quad (9)$$

Note that here we are using the mean and standard deviation from the training data for normalizing the test data. The purpose is to keep the separability between different faults; Otherwise, some separating features may be lost if the test data uses its own mean and standard deviation for normalization. After normalization, both training and test data will be passed to the signal-to-image conversion in Section 3.1 to obtain images.

##### 4.3 Model Design and Training

To comprehensively assess the fault diagnosis performance of the proposed method, we design 6 model architectures with increasing complexity for CNN and GF-CNN, shown in Table 2. For GF-CNN models, the reduced-order feature vector from the MLP has a dimension  $n_{x,mlp} = 10$ , much less than the dimension  $n_{x,cnn}$  from the last feature map of the CNN. Consequently, each GF-CNN model only has slightly more parameters than the corresponding CNN model, consistent with the analysis in Section 3.3. Each model is trained on the training images and validated against the test images. The selected hyperparameters are shown in Table 1 below. All models are trained on the NVIDIA Tesla V100 GPU with 16 GB memory.

Table 1: Selected hyperparameters for training the models

Hyperparameter	Value	Hyperparameter	Value
Batch size:	128	Epochs:	50
Learning rate:	0.001	Dropout rate:	0.5

##### 4.4 Results

In this subsection, we will demonstrate the fault diagnosis performance of the proposed method for the TEP, and also compare it with the standard CNN without directly considering global features. For each model, we repeat the training and testing 10 times to account for the randomness from the optimization algorithm and data shuffling. The fault diagnosis rate (FDR), for the  $i$ -th class, is defined as

$$FDR = \frac{P}{P+B}, \quad (10)$$

where  $P$  represents the number of image samples correctly classified into the  $i$ -th class,  $B$  is the number of image samples classified into the other classes, and  $P+B$  is the total number of images in class  $i$ , i.e., the ground truth.

Fig. 4 shows the comparison of the overall FDR between 6 different CNN and GF-CNN models averaged over 20 faults. It is observed that for simple model architectures (e.g., model

Table 2: The designed CNN and GF-CNN architectures for the case study

Models	CNN		GF-CNN	
	Structure	Parameter #	Structure	Parameter #
# 1	C(16)-P(2)-F(100)*	347,880	C(16)-P(2)-G(10)-F(100)*	358,890
# 2	C(16)-P(2)-C(32)-P(2,1)-F(300)*	644,720	C(16)-P(2)-C(32)-P(2,1)-G(10)-F(300)*	657,730
# 3	C(32)-P(2)-C(64)-P(2,1)-F(300)*	1,292,336	C(32)-P(2)-C(64)-P(2,1)-G(10)-F(300)*	1,305,346
# 4	C(64)-P(2)-C(64)-C(128)-P(2,1)-F(300)*	1,653,744	C(64)-P(2)-C(64)-C(128)-P(2,1)-G(10)-F(300)*	1,666,754
# 5	C(32)-C(64)-P(2)-C(128)-P(2,1)-F(300)*	2,518,192	C(32)-C(64)-P(2)-C(128)-P(2,1)-G(10)-F(300)*	2,531,202
# 6	C(64)-C(64)-P(2)-C(128)-C(256)-P(2,1)-F(300)*	3,331,312	C(64)-C(64)-P(2)-C(128)-C(256)-P(2,1)-G(10)-F(300)*	3,344,322

C( $n$ ): Convolution with  $n$  kernels of  $3 \times 3$ ; P( $n$ ):  $n \times n$  max pooling; P( $n,m$ ):  $n \times m$  max pooling;

F( $n$ ): FC layer with  $n$  neurons; G( $n$ ): Global feature of dimension  $n$  from MLP; \*: Dropout with rate  $p_{dp} = 0.5$ .

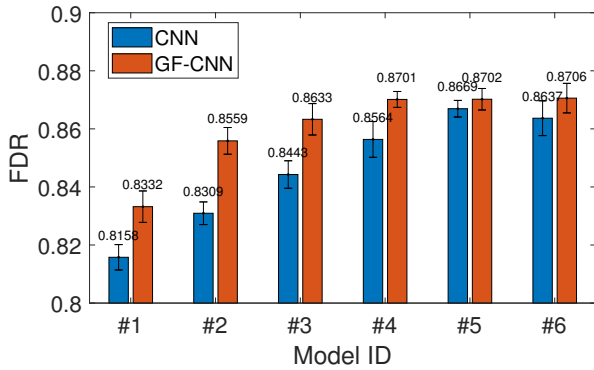


Figure 4: Fault diagnosis performance of the six designed CNN (blue) and GF-CNN (red) models on the TEP data.

#1, #2, and #3), the GF-CNN outperforms the CNN significantly, with a slight increase in the number of learnable parameters. In addition, a simple GF-CNN can achieve a comparable or even better performance than a complex CNN model. For instance, GF-CNN model #2 can yield comparable performance to CNN models #4, but the number of parameters is far less (657,730 vs. 1,653,744). This is because CNN usually requires deep layers to make the receptive field to cover the entire image. Thus, shallow CNN may not capture global features. However, for GF-CNN, it can directly acquire global features without unnecessarily building deep layers. For deeper models (e.g., model #5, #6), the performance difference between CNN and GF-CNN becomes less apparent because deep CNN can extract global features.

Among those model architectures, we select GF-CNN model #4 and CNN model #5 for further comparison. Note that GF-CNN model #4 has much less trainable parameters than CNN model #5. Table 3 shows the FDR across all 20 faults averaged from 10 runs. It is clear that for most faults, the GF-CNN yields better fault detection rates, except for Faults 9, 12, and 16, where CNN gives a slightly higher FRD. In particular, for Faults 3, 8, 13, 15, and 20, the GF-CNN outperforms the CNN significantly. Fig. 5 shows the confusion map of fault diagnosis based on GF-CNN model #4, where most faults can be correctly diagnosed. It also indicates that Faults 3, 9, 15, and 18 are difficult to classify, consistent

Table 3: FDR of GF-CNN model #4 and CNN model #5

Fault	CNN (%)	GF-CNN (%)	Fault	CNN (%)	GF-CNN (%)
1	99.9 ± 0.3	<b>100</b>	11	96.7 ± 0.4	<b>96.8 ± 1.0</b>
2	99.2 ± 1.2	<b>100</b>	12	<b>79.1 ± 3.6</b>	74.5 ± 2.8
3	67.0 ± 4.2	<b>73.1 ± 4.2</b>	13	92.3 ± 0.9	<b>93.1 ± 1.1</b>
4	99.7 ± 0.4	99.7 ± 0.4	14	100	100
5	99.7 ± 0.3	<b>99.9 ± 0.2</b>	15	39.9 ± 8.9	<b>41.1 ± 3.8</b>
6	100	100	16	<b>76.6 ± 3.0</b>	76.0 ± 2.4
7	99.9 ± 0.2	<b>100</b>	17	96.1	96.1
8	91.0 ± 2.6	<b>91.5 ± 2.5</b>	18	94.6 ± 0.8	<b>95.0 ± 1.1</b>
9	<b>30.1 ± 4.6</b>	28.5 ± 5.0	19	98.9 ± 0.6	<b>99.3 ± 0.7</b>
10	83.0 ± 3.2	<b>83.8 ± 3.4</b>	20	89.6 ± 1.6	<b>92.0 ± 1.2</b>

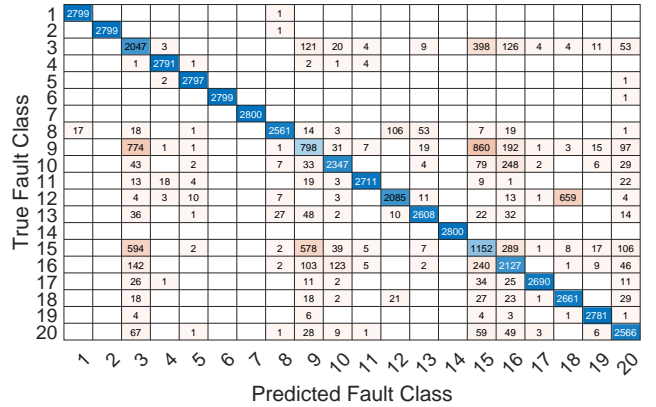


Figure 5: The confusion matrix of the fault diagnosis based on the GF-CNN model #4.

with the observations in literature Lu et al. (2019). These results validate the effectiveness of GF-CNN in improving FDR with a much simpler structure than CNN.

## 5. Conclusion

This article presents a novel global feature-enhanced CNN with an application to the fault diagnosis of complex chemical processes. For the proposed method, a simple single-layer MLP network is utilized to reduce the dimension of input images and extract global features. Such features are then integrated with the local features obtained from CNN to improve the classification performance. Different architectures of CNN and GF-CNN are designed and validated through the

benchmark Tennessee Eastman process dataset. Fault diagnosis results reveal that the GF-CNN with a simple structure can greatly improve the fault diagnosis performance by accounting for global features. The proposed method can be easily extended to other areas such as computer vision and image processing. Future work includes combining multi-scale CNN with GF-CNN, applying the GF-CNN method to real industrial data, and optimizing the model structure with hyperparameter tuning techniques.

### Acknowledgment

Q. Lu acknowledges the startup funds from Texas Tech University. S. Al-Wahaibi acknowledges the Distinguished Graduate Student Assistantships from Texas Tech University.

### References

- Ahsan, M. M., T. E. Alam, T. Trafalis, and P. Huebner (2020). Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and Non-COVID-19 patients. *Symmetry* 12(9), 1526.
- Cao, S., L. Wen, X. Li, and L. Gao (2018). Application of generative adversarial networks for intelligent fault diagnosis. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pp. 711–715. IEEE.
- Chen, X., B. Zhang, and D. Gao (2021). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *Journal of Intelligent Manufacturing* 32(4), 971–987.
- Huang, T., Q. Zhang, X. Tang, S. Zhao, and X. Lu (2022). A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems. *Artificial Intelligence Review* 55(2), 1289–1315.
- Janssens, O., V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufer, S. Verstockt, R. Van de Walle, and S. Van Hoecke (2016). Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration* 377, 331–345.
- Jiang, H., X. Li, H. Shao, and K. Zhao (2018). Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network. *Measurement Science and Technology* 29(6), 065107.
- Jing, L., M. Zhao, P. Li, and X. Xu (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* 111, 1–10.
- Lu, Q., B. Jiang, and E. Harinath (2019). Fault diagnosis in industrial processes by maximizing pairwise Kullback–Leibler divergence. *IEEE Transactions on Control Systems Technology* 29(2), 780–785.
- Rieth, C. A., B. D. Amsel, R. Tran, and M. B. Cook (2017). Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation.
- Shao, B., X. Hu, G. Bian, and Y. Zhao (2019). A multichannel LSTM-CNN method for fault diagnosis of chemical process. *Mathematical Problems in Engineering* 2019.
- Shao, H., H. Jiang, X. Zhang, and M. Niu (2015). Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology* 26(11), 115002.
- Sinitsin, V., O. Ibryaeva, V. Sakovskaya, and V. Ereemeeva (2022). Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model. *Mechanical Systems and Signal Processing* 180, 109454.
- Song, Q. and P. Jiang (2022). A multi-scale convolutional neural network based fault diagnosis model for complex chemical processes. *Process Safety and Environmental Protection* 159, 575–584.
- Song, X., Y. Cong, Y. Song, Y. Chen, and P. Liang (2022). A bearing fault diagnosis model based on CNN with wide convolution kernels. *Journal of Ambient Intelligence and Humanized Computing* 13(8), 4041–4056.
- van den Hoogen, J., S. Bloemheuvel, and M. Atzmueller (2020). An Improved wide-kernel CNN for classifying multivariate signals in fault diagnosis. In *2020 International Conference on Data Mining Workshops (ICDMW)*, pp. 275–283. IEEE.
- Wen, L., X. Li, L. Gao, and Y. Zhang (2017). A new convolutional neural network-based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics* 65(7), 5990–5998.
- Wu, H. and J. Zhao (2018). Deep convolutional neural network model based chemical process fault diagnosis. *Computers & Chemical Engineering* 115, 185–197.
- Zhang, S., S. Zhang, B. Wang, and T. G. Habetler (2020). Deep learning algorithms for bearing fault diagnostics – A comprehensive review. *IEEE Access* 8, 29857–29881.
- Zhang, W., G. Peng, C. Li, Y. Chen, and Z. Zhang (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* 17(2), 425.
- Zhang, W., F. Zhang, W. Chen, Y. Jiang, and D. Song (2018). Fault state recognition of rolling bearing based fully convolutional network. *Computing in Science & Engineering* 21(5), 55–63.
- Zhao, R., R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing* 115, 213–237.
- Zheng, S. and J. Zhao (2020). A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis. *Computers & Chemical Engineering* 135, 106755.