High-Efficiency Generative Adversarial Network Model for Chemical Process Fault Diagnosis

Ruoshi Qin*, Jinsong Zhao*,**

** State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing, China (e-mail: qrs19@mails.tsinghua.edu.cn)*

***Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing, China (e-mail: jinsongzhao@tsinghua.edu.cn)*

Abstract: Fault diagnosis techniques are essential to ensure the long-term reliability of industrial process systems. Current deep learning methods mainly rely on a large quantity of training data. Generative Adversarial Network (GAN) model has started to be utilized for diagnostic problems with small sample size and data imbalance in recent years, but the fault diagnosis performance heavily depends on the experience of the model builder. In this work, a high-efficiency GAN model (HGAN) is proposed for chemical process fault diagnosis. HGAN integrates the advantages of Wasserstein GAN and Auxiliary Classifier GAN to promote the generating model training stability and the discriminative model training efficiency with Bayesian optimization. Experiments on the benchmark Tennessee Eastman process under the circumstance of small samples show that the presented model can achieve a satisfactory fault diagnosis accuracy without the assistance of a redundant deep neural network classifier and with reduced effort in model tuning.

Keywords: Auxiliary Classifier Generative Adversarial Network, Wasserstein distance, Gradient penalty, Bayesian optimization, Fault diagnosis, Tennessee Eastman process

1. INTRODUCTION

Modern chemical processes are highly automatic with the implementation of advanced computer and information technologies. Since the equipment used in manufacturing is dangerous and the operating condition is complex, process monitoring has drawn significant attention from academia and industry, and intelligent fault diagnosis methods are the emphasis. Over the last decade, data-driven fault diagnosis methods have attained considerable growth. Taking benefit from machine learning, they are proved to outperform the traditional model-based ones.

Data-driven fault diagnosis methods are generally made up of statistical models and deep learning models. Statistical models which are commonly applied include principal component analysis (PCA), partial least squares (PLS), Fisher discriminant analysis (FDA), k-nearest neighbors (kNN), etc., and their variants. Recently, deep learning models become a hot topic of interest both in academia and industry. Zhang proposed an extensible deep belief network (DBN) model and achieved a diagnostic accuracy of 82.1% for all the 20 fault types in the benchmarked Tennessee Eastman process (Zhang et al., 2017). Wu firstly adopted a deep convolutional neural network (DCNN) and improved the diagnosis performance (Wu et al., 2018). Zheng utilized convolutional stacked autoencoder (SAE) to conduct unsupervised data mining and reached a satisfying classification accuracy (Zheng et al., 2020). Some of the latest deep learning achievements have also been tested for chemical process fault diagnosis, such as transfer learning (Wu et al., 2020), self-adaptive algorithm (Wu et al., 2020), graph convolutional network (Wu et al., 2021), self-training algorithm (Zheng et al, 2021) and self-attentive variational autoencoder (Bi and Zhao, 2021) in order to enhance the model robustness under multiple production modes or the model explainability. However, the majority of the deep learning models require large amounts of data samples to be invested in training no matter whether they possess corresponding labels. Their performances are not acceptable enough under small sample conditions.

Generative adversarial network (GAN) was created in 2014, which has achieved lots of successful applications in the generation of images, texts, and audio to address the trouble of data imbalance and small sample size (Goodfellow et al, 2020). Composed by generating model and discriminative model, GAN can be configured with various neural networks. These two parts confront each other and strike a dynamic balance in the end. In the last three years, the application of GAN and its different forms appeared in fault diagnosis. Wang combined stacked denoising autoencoders (SDAE) with GAN for planetary gearbox fault pattern recognition with limited fault samples and strong noise interference (Wang et al, 2018). Shao developed an auxiliary classifier GAN (ACGAN) model to learn from mechanical sensor signals and generate realistic one-dimensional raw data and used a CNN classifier to output the machine fault diagnosis result (Shao et al., 2019). Yan introduced a supervised classifier framework with GAN to increase the number of faulty training samples and re-balance the training dataset and performed robust fault diagnosis for air handling units (Yan et al, 2020). Peng built an enhanced ACGAN with deep neural network (DNN) to solve imbalanced problems and classify the chemical process faults (Peng et al, 2020). But these current GAN-based fault

diagnosis methods still have an obvious drawback. They regard GAN only as a tool of data generation and overlook the important role of the discriminative model inside. Redundant classifiers are added to make the diagnosis framework more cumbersome.

In this paper, a high-efficiency chemical process fault diagnosis method based on Auxiliary Classifier GAN combined with Wasserstein distance, gradient penalty, and Bayesian optimization (HGAN for brevity) is proposed. Depending on the time-domain characteristics of process sensing signals, one-dimensional convolutional layers are constructed to improve the efficiency of feature extraction. Wasserstein distance and gradient penalty are used to measure the distance between two distributions and get rid of the pattern collapse problem in generator training. The Bayesian optimization strategy is employed to adaptively regulate the discriminator parameters. The entire GAN model is no longer limited to generating more data for training but directly used for fault diagnosis under the circumstance of small raw samples without adding another classification module.

The rest of the paper is organized as follows. In Section 2, the basic theory of HGAN and the diagnosis framework are introduced in detail. Section 3 is a case study and provides concrete analysis of the experiments. The conclusion and outlook of the proposed model are summarized in Section 4.

2. METHODOLOGY

2.1 Auxiliary Classifier generative adversarial network

The structure of GAN is built based on the idea of two-person zero-sum game. The discriminator learns to identify between real samples and fake samples generated by neural networks, and the generator aims to generate the pseudo data that can fool the discriminator. However, the primitive GAN model has defects such as poor stability, training difficulty, and lack of diversity in generating samples. ACGAN is created as an efficient derivative model which takes strengths of label information as well for sample generation and distinction.

To be specific, the generator maximums the mutual information of random noise z and class label set $c{\sim}p_c$ simultaneously to generate new samples $X_{generated}{\sim}G(z,c)$, and the discriminator has a comprehensive ability to distinguish the true degree and the sample classifications. Through the help of game mechanism, ACGAN is constantly optimized in terms of the improved capacity for sample generation with corresponding labels and class recognition.

Given both the real samples and label information need to be processed, the objective function consists of two loss functions: the log-likelihood of the correct source L_s and the log-likelihood of the correct class L_c, which can be represented as

$$L_s = E\big[\log P\big(S = fake|X_{generated}\big)\big] + E[\log P(S = real|X_{real})] \quad (1)$$

$$L_c = E\big[\log P\big(C = c|X_{generated}\big)\big] + E[\log P(C = c|X_{real})] \quad (2)$$

The goal of the discriminator training is to maximize $L_s + L_c$ while the generator is trained to maximize $L_s - L_c$.

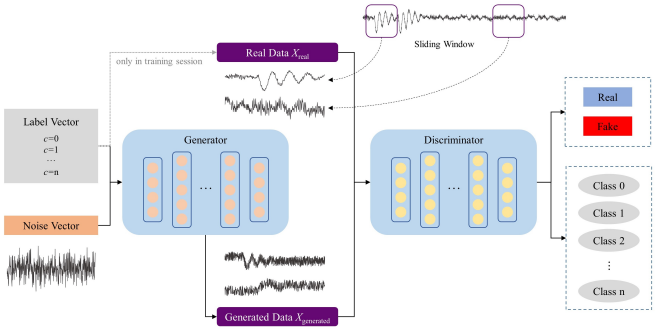

Figure 1. ACGAN model structure.

2.2 Wasserstein distance and gradient penalty

Facing the problem of GAN training instability, Arjovsky and Bottou devoted their endeavor and proposed Wasserstein GAN (WGAN) to avert vanishing gradient and mode collapse (Arjovsky et al, 2017). The unstable training of GAN is due to the objective function in the form of Kullback-Leibler (KL) divergence or Jensen–Shannon (JS) divergence. Wasserstein distance is applied to replace them to evaluate the distribution distance between real samples and generated samples.

Wasserstein distance formula is as follows:

$$W\big(p_{data}(x), p_z(z)\big) = \inf_{\gamma \sim \Pi(p_{data}(x), p_z(z))} f \; E_{(x,z)\sim\gamma}[\|x - z\|] \quad (3)$$

where γ implicates an aggregation of all of the possible joint distributions $p_z(z)$ combined for real data $p_{data}(x)$ and generated data. The distance z is calculated to obtain the expected value of the sample to the distance under the joint distribution.

Referring to the above preliminaries, the objective function of WGAN can be expressed as follows:

$$\min_G \max_{D\epsilon\Omega} E_{x\sim p_{data}}[D(x)] - E_{z\sim p_z}\big[D\big(G(z)\big)\big] \quad (4)$$

where Ω represents the set of 1-Lipschitz functions, and the weights are located in a compact space $[-w, w]$. Thus, on the aspect of the entire model, minimizing the value function concerning the generator parameters equals minimizing the Wasserstein distance in possession of an optimal discriminator (Li et al, 2021).

Moreover, in order to handle the convergence difficulty arisen by adding the Lipschitz constraint on the discriminator, gradient penalty is introduced into the WGAN algorithm. The loss function and the objective function are presented as follows:

$$L = E_{z\sim p_z}\big[D_s\big(G(z)\big)\big] - E_{x\sim p_{data}}[D_s(x)] + \lambda E_{\hat{s}\sim p_{\hat{s}}}[(\|\nabla_{\hat{s}} D(\hat{s})\|_2 - 1)^2] \quad (5)$$

$$\min_{G}\max_{D\in\Omega} E_{x\sim p_{data}}[D(x)] - E_{z\sim p_z}\big[D\big(G(z)\big)\big] - \lambda E_{\hat{s}\sim p_{\hat{s}}}[(\|\nabla_{\hat{s}} D(\hat{s})\|_2 - 1)^2] \quad (6)$$

Since the optimization of the generator gets effortless to be optimized after the WGAN value function produces a better critical function relative to the input gradient, the balance between the generator and the discriminator is much easier to maintain and the training process becomes more stable and faster, which is of great benefit to the solution of small sample problem.

2.3 Bayesian optimization

Hyperparameters in GAN models are crucial to their performance of generation and discrimination, but they cannot be learned from the training process directly which is also a part of the algorithm. Hyperparameter selection based on empirical setting and grid search is a tedious process that is time-consuming and may encounter the problem of gradient explosion. The Bayesian Optimization algorithm is an appropriate solution to obtain optimal values rapidly.

The basic procedure of Bayesian optimization is divided into two steps: probabilistic model selection and acquisition function iteration. First, the objective function can be represented by a modified probabilistic model. Gaussian process is extensively adopted due to its excellent flexibility, and the sampled black-box objective function is as follows:

$$f(x)\sim \mathrm{GP}(m(x), k(x,x')) \quad (7)$$

where $f(x)$ is the evaluated value of the objective function corresponding to samples and $m(x)$ represents a mean function. $k(x,x')$ refers to a covariance function of $f(x)$ and $f(x')$.

$$k(x,x') = E[(f(x) - E[f(x)])(f(x') - E[f(x')])] \quad (8)$$

The optimal hyperparameters are derived by maximizing the marginal likelihood distribution. Then, based on the estimated posterior distribution of the objective function, the hyperparameter combination of the next sample needs to be selected with acquisition function iteration. Bayesian optimization requires a trade-off between exploration and exploitation in order to avoid local optima. Expected improvement is employed as the acquisition function which offers a standard of the usefulness of each given test point (Xu et al., 2020).

$$EI(x) = \begin{cases} \big(\mu(x) - f(x')\big)\phi(Z) + \sigma(x)\phi(Z), \sigma(x) > 0 \\ 0 \qquad\qquad\qquad\qquad\quad , \sigma(x) = 0 \end{cases} \quad (9)$$

$$Z = \frac{\mu(x) - f(x')}{\sigma(x)} \quad (10)$$

The algorithm takes advantage of previous samples to adjust the prior distribution and the parameters that minimize the probability result to the global minimum. Finally, the optimal observed value and hyperparameters are gained by continuous iterations. For the purpose of improving the capability of the auxiliary classifier in the discriminator, its objective function is set to be the opposite of the accuracy of classification output. In each iteration of Bayesian optimization, the hyperparameters of the generator are fixed and only the ones of the discriminator are adjusted. At the same time, the training performance of both models can be improved jointly by game mechanism and reach the best at Nash equilibrium (Li et al, 2021).

2.4 Framework of HGAN-based fault diagnosis method

Combined with the introductions of the utilized models above, a framework of the HGAN-based fault diagnosis method is shown in Figure 2.

At the offline stage, raw data of the chemical process is collected including operation parameters and process parameters, which is subsequently divided into different sets for training and testing. After normalization and other necessary pre-processing steps, the training dataset is used to train the HGAN model and acquire the optimized parameters of the generator and the discriminator. The training model is saved and the parameters are migrated to the online stage.

At the online stage, the normalized real-time data is fed into the trained discriminative model and the fault diagnosis results with corresponding labels are output. Under the situation of data imbalance, the generative model can also expand the training set with generated samples to enhance the capability of diagnosis.

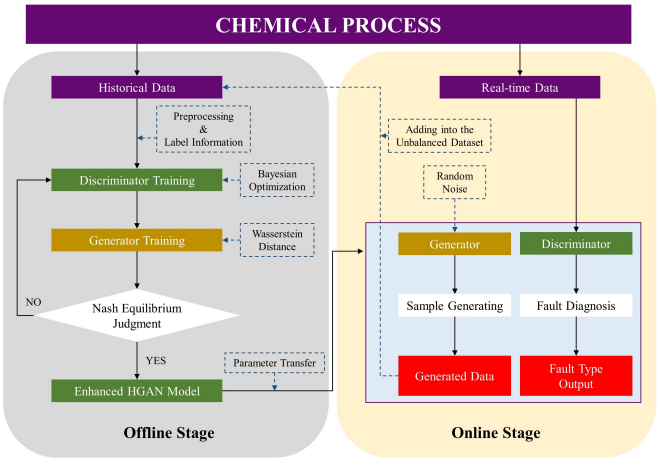

Figure 2. The framework of the proposed fault diagnosis method.

The objective function of the generator and the discriminator in HGAN can be described in the following form.

$$L_{Generator} = E_{z\sim p_z}\big[D_s\big(G(z)\big)\big] - E_{x\sim p_{data}}[D_s(x)] + E_{x\sim p_{data}}[D_c(x|c)|c] + E_{z\sim p_z}[D_c(G(z|c)|c)] + \lambda E_{\hat{s}\sim p_{\hat{s}}}[(\|\nabla_{\hat{s}} D(\hat{s})\|_2 - 1)^2] \quad (11)$$

$$L_{Discriminator} = E_{z\sim p_z}\big[D_s\big(G(z)\big)\big] + E_{x\sim p_{data}}[D_s(x)] + E_{x\sim p_{data}}[D_c(x|c)|c] + E_{z\sim p_z}[D_c(G(z|c)|c)] \quad (12)$$

The architecture of the generator and the discriminator is shown in Fig.3 and Fig.4. The generative model is composed of dense, up-sampling, and convolutional layers with ReLU and tanh selected as the activation functions. The four convolutional layers have 128, 64, 32, and 1 filter respectively

with the stride set to 2. The discriminative model consists of convolutional layers, batch normalization, max-pooling, and fully connected layers accompanied by the LeakyReLU activation function. The output layer provides a source label and a class label in the end.

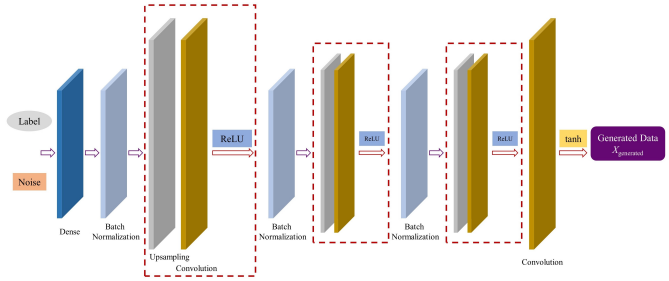

Figure 3. The architecture of the generator.

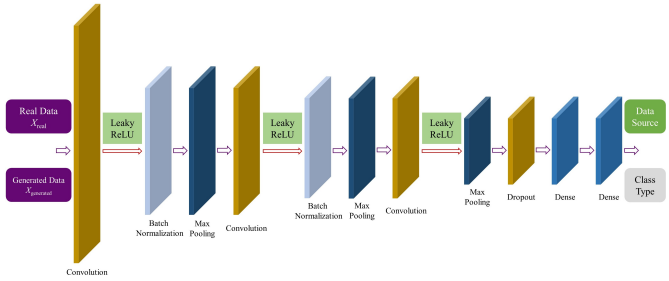

Figure 4. The architecture of the discriminator.

3. CASE STUDY

In this section, the benchmark Tennessee Eastman process (TEP) simulator is applied to evaluate the performance of our proposed model. The experiments are conducted with small size of training samples under data balance and data imbalance conditions separately. The diagnosis results of HGAN are compared with a traditional statistical model, SVM, and other deep learning models, including CNN, GAN-CNN, and ACGAN for ablation experiments.

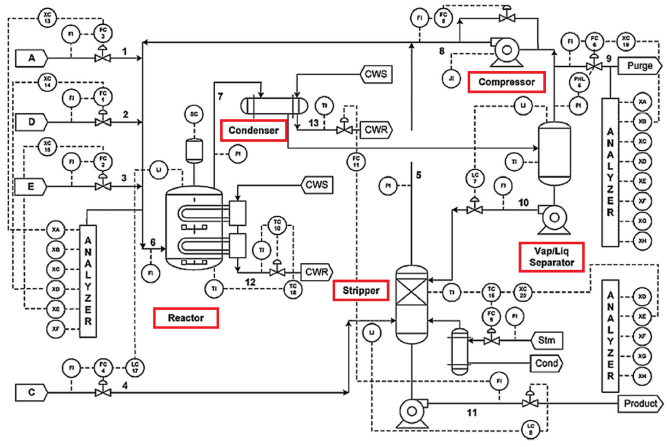

Figure 5. P&ID of TE process.

3.1 Tennessee Eastman process

TEP has been a widely used case for various studies including process modeling and monitoring. It consists of five main operation units: reactor, condenser, recycle compressor, separator, and stripper (Downs and Vogel, 1993). Figure 5 displays the P&ID of TEP. There is a set of 21 programmed fault modes which are defined in Table 1 with 41 measurements and 11 manipulated variables under observation. The simulation data is publicly available and can be downloaded from the following website, https://github.com/camaramm/tennessee-eastman-profBraatz (Chiang et al, 2001). Each training dataset collects 500 sampling points from a simulation of 25 hours with a sampling interval of 3 minutes, and the disturbance of each fault mode is introduced after one-hour running. Each testing dataset consists of 960 sampling points with 160 normal data and 800 fault data.

For purpose of highlighting the circumstance of the limited training dataset, the sampling sliding window size is fixed to 50 data points and only 100 normal samples are selected to be input for model training. The testing dataset includes 16800 fault samples in total. The values of the samples are normalized to have zero mean and unit variance.

Table 1. Process Disturbances in the TEP

IDVs	Variable Description	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 variation
IDV(9)	D feed temperature (stream 2)	Random variation
IDV(10)	C feed temperature (stream 4)	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Sticking
IDV(15)	Condenser cooling water valve	Sticking
IDV(16) - IDV(20)	Unknown	Unknown
IDV(21)	Valve (Stream 4)	Constant position

3.2 Fault diagnosis with data-balanced training

The semi-supervised fault diagnosis experiment on real-time data with small and balanced training samples is carried out. 100 fault samples of each failure mode are also put into the

training model at the offline stage. The fault diagnosis accuracy results between the proposed model and other popular models are compared in Table 2.

According to the previous studies on the TEP, the 3rd, 9th, and 15th faults are hard to distinguish which is also displayed in this experiment (Luo et al, 2020). Except for the above fault types, the proposed model demonstrates satisfactory diagnostic performances in most of the others. Furthermore, HGAN defeats the other models in terms of more than 15 kinds of faults. It strongly confirms that the enhanced model has more advantages over the original framework on diagnosis work of a small balanced dataset. The accuracy of some fault types is comparable to that of using deep neural network model with sufficient training samples.

Figure 6 illustrates the confusion matrix of the proposed model's diagnosis results, where the main diagonal numbers are the percentage of samples correctly identified. This diagram reveals the powerful capability of HGAN in feature learning and classification.

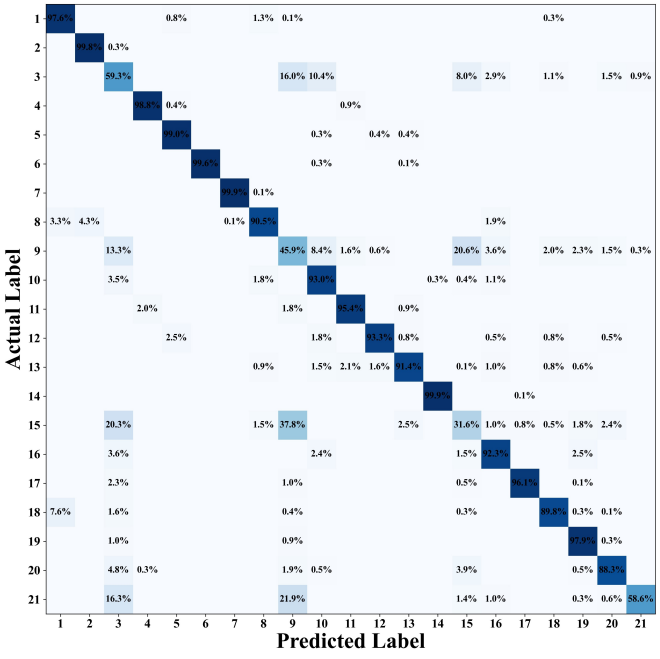

Figure 6. Confusion matrix for diagnosis task with HGAN.

Table 2. Fault diagnosis accuracy with small and balanced training data by CNN, GAN-CNN, ACGAN, and HGAN

IDVs	SVM	CNN	GAN -CNN	ACGAN	HGAN
IDV(1)	85.3%	84.1%	93.8%	92.9%	**97.6%**
IDV(2)	83.8%	86.3%	99.3%	99.5%	**99.8%**
IDV(3)	16.5%	15.5%	33.3%	24.6%	**59.3%**
IDV(4)	88.1%	83.4%	97.0%	93.9%	**98.8%**
IDV(5)	75.9%	79.3%	98.3%	91.5%	**99.1%**
IDV(6)	91.6%	90.9%	**99.8%**	99.5%	99.6%
IDV(7)	94.5%	91.8%	99.8%	**99.9%**	**99.9%**
IDV(8)	73.9%	52.1%	86.6%	77.8%	**90.5%**
IDV(9)	12.0%	10.4%	23.1%	18.3%	**45.9%**
IDV(10)	60.6%	21.1%	85.9%	86.0%	**93.0%**
IDV(11)	73.5%	80.9%	94.5%	88.9%	**95.4%**
IDV(12)	77.8%	65.8%	**94.1%**	80.4%	93.3%
IDV(13)	63.6%	62.0%	87.4%	72.3%	**91.4%**
IDV(14)	85.6%	93.4%	**99.9%**	99.8%	**99.9%**
IDV(15)	14.8%	19.1%	25.5%	21.1%	**31.6%**
IDV(16)	46.5%	22.4%	64.3%	34.6%	**92.3%**
IDV(17)	73.3%	83.1%	93.3%	91.0%	**96.1%**
IDV(18)	64.9%	71.6%	85.4%	85.8%	**89.8%**
IDV(19)	69.0%	92.6%	96.1%	**98.8%**	97.9%
IDV(20)	61.1%	58.5%	67.6%	64.9%	**88.3%**
IDV(21)	25.4%	35.6%	54.0%	49.8%	**58.6%**
average	63.7%	61.9%	80.0%	74.8%	**86.6%**

3.3 Fault diagnosis with data-imbalanced training

The diagnostic effectiveness of the model proposed in this work is also tested when the categories of training data turn unbalanced. The number of fault samples for training decreases to 20 per type. The imbalance ratio of normal samples to fault samples of each class is 5:1. The results of the above models are revealed in Table 3.

Table 3. Fault diagnosis accuracy with small and imbalanced training data by CNN, GAN-CNN, ACGAN, and HGAN

IDVs	SVM	CNN	GAN -CNN	ACGAN	HGAN
IDV(1)	61.0%	63.4%	87.4%	81.8%	**96.6%**
IDV(2)	62.8%	66.6%	93.6%	90.1%	**97.0%**
IDV(3)	10.8%	9.4%	22.5%	12.8%	**46.6%**
IDV(4)	51.6%	56.8%	92.1%	86.4%	**96.5%**
IDV(5)	43.9%	48.0%	94.6%	83.3%	**97.9%**
IDV(6)	82.3%	74.5%	96.8%	92.0%	**97.5%**
IDV(7)	85.5%	79.3%	**98.8%**	94.5%	**98.8%**
IDV(8)	39.6%	22.9%	77.9%	71.4%	**85.3%**
IDV(9)	8.4%	7.3%	16.4%	12.4%	**32.0%**
IDV(10)	38.0%	13.3%	76.8%	72.1%	**89.1%**
IDV(11)	58.3%	56.6%	87.0%	82.4%	**93.5%**
IDV(12)	53.6%	36.8%	88.5%	67.5%	**88.8%**
IDV(13)	31.5%	35.5%	75.1%	64.6%	**84.9%**
IDV(14)	77.1%	80.4%	**98.9%**	93.0%	98.6%
IDV(15)	9.6%	11.0%	18.3%	17.8%	**27.8%**
IDV(16)	33.8%	17.8%	47.5%	25.1%	**87.9%**
IDV(17)	52.5%	54.6%	90.9%	77.0%	**93.6%**
IDV(18)	43.9%	53.4%	80.5%	74.5%	**86.4%**
IDV(19)	60.1%	80.5%	93.3%	90.3%	**95.1%**
IDV(20)	48.4%	37.1%	59.5%	41.4%	**74.0%**
IDV(21)	19.3%	14.6%	36.8%	25.8%	**53.3%**
average	46.3%	43.8%	73.0%	64.6%	**82.0%**

Though the fault diagnosis experiment with small and imbalanced training samples is harsh, it is gratifying to note that HGAN still achieves stable classification. This experiment illustrates that the generative model is able to generate pseudo-samples with accurate features consistently and the discriminative model is capable of efficient feature recognition.

4. CONCLUSION AND OUTLOOK

In this paper, a novel fault diagnosis model named High-Efficiency Generative Adversarial Network (HGAN) is developed. The main contributions of this work are as follows.

(1) A novel efficient fault diagnosis method is presented based on GAN model which has no requirement for another deep neural network classifier to be attached and less time-consuming for model tuning.

(2) The proposed HGAN model incorporates the advantages of WGAN and ACGAN, which utilizes Wasserstein distance and gradient penalty to empower the generator training process, and adopts the Bayesian optimization strategy to enhance the discriminator performance.

(3) The experiments with limited samples are conducted with data-balanced training and data-imbalanced training. The outcomes verify the validity and robustness of HGAN.

Owing to the limitation of time and article pages, there are still some aspects that are not mentioned in this paper. On the one hand, in case of sufficient training data, the comparison between HGAN and the latest deep learning models should proceed. On the other hand, the diagnosis performance with multiple faults or multimode switches under small sample conditions is desired to be discussed. These issues should be further studied and explored in the future to explore the potential of this model.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from the National Science and Technology Innovation 2030 Major Project of the Ministry of Science and Technology of China (2018AAA0101605) and the National Natural Science Foundation of China (No. 21878171).

REFERENCES

Arjovsky, M., Chintala, S. and Bottou, L., 2017. Wasserstein GAN. *arXiv e-prints*, arXiv:1701.07875.

Bi, X. and Zhao, J., 2021. A novel orthogonal self-attentive variational autoencoder method for interpretable chemical process fault detection and identification. *Process Safety and Environmental Protection*, 156, pp.581–597.

Chiang, L.H., Russell, E. L. and Braatz, R.D., 2001. Fault Detection and Diagnosis in Industrial Systems. *Measurement science & technology*, 12(10), pp.1745.

Downs, J. and Vogel, E., 1993. A plant-wide industrial process control problem. *Computers & Chemical Engineering*, 17, pp.245-255.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2020. Generative adversarial networks. *Communications of the ACM*, 63, pp.139–144.

Li, D., Liu, Y., Zhao, Y. and Zhao, Y., 2021. Fault Diagnosis Method of Wind Turbine Planetary Gearbox Based on Improved Generative Adversarial Network. *Proceedings of the Chinese Society of Electrical Engineering*, 41 (21), 7496-7506.

Li, Z. et al., 2021. A Novel Method for Imbalanced Fault Diagnosis of Rotating Machinery Based on Generative Adversarial Networks. *IEEE transactions on instrumentation and measurement*, 70, pp.1–17.

Luo, L., Xie, L. and Su, H., 2020. Deep Learning With Tensor Factorization Layers for Sequential Fault Diagnosis and Industrial Process Monitoring. *IEEE access*, 8, pp.105494–105506.

Peng, P., Wang, Y., Zhang, W., Zhang, Y. and Zhang, H., 2020. Imbalanced Process Fault Diagnosis Using Enhanced Auxiliary Classifier GAN. *2020 Chinese Automation Congress*, pp.313-316.

Shao, S., Wang, P. and Yan, R., 2019. Generative adversarial networks for data augmentation in machine fault diagnosis. *Computers in Industry*, 106, pp.85–93.

Wang, Z., Wang, J. and Wang, Y., 2018. An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition. *Neurocomputing (Amsterdam)*, 310, pp.213–222.

Wu, D. and Zhao, J., 2021. Process topology convolutional network model for chemical process fault diagnosis. *Process Safety and Environmental Protection*, 150, pp.93–109.

Wu, H. and Zhao, J., 2018. Deep convolutional neural network model based chemical process fault diagnosis. *Computers & Chemical Engineering*, 115, pp.185–197.

Wu, H. and Zhao, J., 2020. Fault detection and diagnosis based on transfer learning for multimode chemical processes. *Computers & Chemical Engineering*, 135, p.106731.

Wu, H. and Zhao, J., 2020. Self-adaptive deep learning for multimode process monitoring. *Computers & chemical engineering*, 141, p.107024.

Xu, Y., Yang, X., Gan, Y., Zhou, W., Cheng, H. and He, X., 2021. A Music Generation Model Based on Generative Adversarial Networks with Bayesian Optimization. *Proceedings of 2020 Chinese Intelligent Systems Conference*, vol 705. Springer, Singapore.

Yan, K. et al., 2020. Unsupervised learning for fault detection and diagnosis of air handling units. *Energy and Buildings*, 210, p.109689.

Zhang, Z. and Zhao, J., 2017. A deep belief network based fault diagnosis model for complex chemical processes. *Computers & Chemical Engineering*, 107, pp.395–407.

Zheng, S. and Zhao, J., 2020. A new unsupervised data mining method based on the stacked autoencoder for chemical process fault diagnosis. *Computers & Chemical Engineering*, 135, p.106755.

Zheng, S. and Zhao, J., 2021. A Self-Adaptive Temporal-Spatial Self-Training Algorithm for Semi-Supervised Fault Diagnosis of Industrial Processes. *IEEE transactions on industrial informatics*.