Energy & Buildings 268 (2022) 112207

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

Integrated generative networks embedded with ensemble classifiers for fault detection and diagnosis under small and imbalanced data of building air condition system

Jianxin Zhang [a], Zhengfei Li [b], Huanxin Chen [b,*], Hengda Cheng [b], Lu Xing [c], Yuzhou Wang [b], Li Zhang [b]

[a] China-EU Institute for Clean and Renewable Energy, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[b] School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[c] Mechanical and Construction Engineering, Northumbria University, Newcastle Upon Tyne NE1 8ST, United Kingdom

ARTICLE INFO

Article history:
Received 16 February 2022
Revised 18 May 2022
Accepted 23 May 2022
Available online 27 May 2022

Keywords:
Variable refrigerant flow
Fault diagnosis
Generative Adversarial Network
Small and imbalanced data

ABSTRACT

Faults in building Heating, Ventilation, and Air-condition (HVAC) system create an uncomfortable indoor environment and cause energy waste. The data-driven method has been widely applied for Fault Detection and Diagnosis (FDD) in the complex building HVAC system. This method relies on the availability of many fault data which is difficult to collect. This makes it quite challenging to apply the data-driven methods for the FDD of the HVAC system. Thus, a novel data-driven FDD method that only utilizes small fault data collected from a Variable Refrigerant Flow air condition system has been proposed. Under different conditions, the fault and normal data are collected in an enthalpy difference laboratory to create small and imbalanced data. A generative network is developed by combining Wasserstein Generative Adversarial Network with Gradient Penalty and Variational Auto-Encoder. To improve the FDD classifier's accuracy and to train an end-to-end network model using small and imbalanced data, two ensemble classifiers are embedded into the generative network. The dataset includes normal and fault data have been applied to train the modified generative network, and two ensemble classifiers are used to detect and diagnose the fault, respectively. The performance indexes show that the proposed method is much better than the SMOTE-based methods in almost all training groups. Besides, the comparison between the proposed method and generative network with a single classifier indicates that the ensemble classifiers can improve the F1-score of fault detection and the accuracy of fault diagnosis.

1. Introduction

China's rapid urbanization significantly increases great energy and resource demands [1]. In the past decades, building accounts for almost 30% of the total energy consumption in China. Moreover, the Heating, Ventilation, and Air Condition (HVAC) systems account for 55% of public building energy consumption [2]. The variable Refrigerant Flow (VRF) system is a centralized HVAC system widely used in commercial buildings since it can provide high thermal comfort with lower energy consumption [3–5].

The regular operation of the components ensures that the VRF system can distribute the refrigerant effectively and avoid potential danger. Fault can damage components, waste energy, and cause an uncomfortable indoor environment [6]. Faults in VRF systems can be classified into two types: system-based and component-based [7]. Refrigerant undercharge and overcharge [8] can be regarded as the system-based fault. The component-based fault includes compressor liquid refrigerant flood back[9], valve stuck and leakage[10], fouling, etc. The Fault Detection and Diagnosis (FDD) method can accurately and quickly judge whether the VRF system is faulty, determine which type of fault is, and provide maintenance suggestions.

According to related literature, existing FDD methods for HVAC systems can be classified into data-driven and knowledge-driven

Abbreviations: CE, Cross Entropy; DNN, Deep Neural Network; EXV, Electronic Expansion Valve; FCL, Full Connection Layer; FDD, Fault Detection and Diagnosis; FRV, Four-way Reversing Valve; GAN, Generative Adversarial Network; HVAC, Heating, Ventilation, and Air condition; IU, Indoor Unit; OU, Outdoor Unit; PCA, Principal Component Analysis; RF, Random Forest; SMOTE, Synthetic Minority Oversampling Technique; SVM, Support Vector Machine; VAE, Variational Auto-Encoder; VRF, Variable refrigerant flow; WGAN-GP, Wasserstein Generative Adversarial Network with Gradient Penalty.

* Corresponding author at: Department of Refrigeration and Cryogenic Engineering, School of Energy and Power Engineering, Huazhong University of Science and Technology.

E-mail address: chenhuanxin@tsinghua.org.cn (H. Chen).

https://doi.org/10.1016/j.enbuild.2022.112207
0378-7788/

[11,12]. Benefiting from the improvement of computers, the data-based method has been widely studied for FDD of HVAC systems [13]. However, small fault data makes it challenging to extract enough feature information to train fault diagnosis classifiers, and the imbalance between normal and fault data also has a negative impact on training data-driven classifiers for fault detection [14].

In recent years, the data-driven FDD method under small and imbalanced data can be divided into classifier optimization, feature learning, and data augmentation[15]. Classifier optimization mainly changes the structure of the existing algorithm to be applicable to be trained under imbalanced data. The cost-sensitivity analysis is one of the most widely used methods for optimizing the classifier[16–18]. It can be embedded into many existing classifiers, such as Support Vector Machine (SVM) and Random Forest (RF), by setting different misclassification punishments between small and big data to force the classifier to learn more feature information from small data. Yan et al. [17,19] used cost sensitivity analysis for early FDD of air handling units. However, the performance improvement by using cost sensitivity analysis is lower than other methods. Besides, the results in some papers indicated that classifiers based on decision-tree and SVM are slightly influenced by imbalanced data[20–23]. Ebrahimifakhar et al. [24] compared several algorithms with SVM to detect and diagnose the fault of packaged rooftop units, indicating that SVM shows the best performance compared with mentioned methods. An optimized classifier always works with data augmentation for better performance under small and imbalanced data.

Feature learning means extracting fault features as much as possible by designing powerful machine learning algorithms, including Deep Neural Network (DNN) [25,26] and Transfer learning[27–29]. Jia et al. [26] proposed a deep normalized convolutional neural network framework for imbalanced fault classification of machinery. In the transfer learning scenario, the model learns fault features from the source domain, which contains many data and is micro-adjusted with small data from the target domain so that the model can perform well in the target domain. Lu and Tao [27] proposed a novel two-stage transferable feature space mining method that consists of a common feature network and compare network for bear fault diagnosis. The results indicated that the proposed method could efficiently learn feature information for transfer fault diagnosis with all labeled and unlabeled data. Fan et al. [28] applied prior knowledge transferred from the centrifugal chiller to the screw chiller, and the results indicated that the overall diagnostic performance of the screw chiller with much fewer data was improved.

Data augmentation mainly includes data sampling[30] and data generation. Data sampling is the traditional method to process imbalanced data, including oversampling[28,31], under sampling [32], and mixed sampling[33]. The Synthetic Minority Oversampling Technique (SMOTE) is one of the most widely used methods [34]. Zhou et al. [20] used Principal Component Analysis (PCA) to reduce the dimension of features, and SMOTE was used to augment fault data for the VRF system, including valve fault, fan fouling, etc. Fan et al. [28] used SMOTE to generate synthetic fault data of screw chiller, including refrigerant undercharge, condenser fouling, non-condensable gases, etc. The absolute accuracy of fault diagnosis based on SMOTE method can reach 96.7%. Besides, Wang et al. [32] used Random under sampling to improve the FDD strategy for the VRF system. Although data sampling requires fewer computing resources to get more fault data, such sampling methods find data distribution to make synthetic data, and it could blur the classification boundary without exploring the hidden features of data.

Compared with data sampling, data generation pays more attention to extracting features from small data to generate similar data. Generative Adversarial Network (GAN) [35–38] and Variational Auto-Encoder (VAE) [39] have been the most popular generative networks because of their ability to rebuild target data from random noise and extract feature information from original data. Generative networks have been widely used to generate synthetic fault data for the rotatory machine in recent years [40–43], but only a few related studies on the HVAC system can be found. Yan and Zhong [44–47] used GAN to generate fault data of chiller and air handling units for the first time and designed a general FDD strategy under small and imbalanced data for the HVAC system. Besides, Fan et al. [48] used VAE to enhance imbalanced data for air handling units, and several data sampling methods for data augmentation are compared with the VAE-based method to make an overall quantitative analysis. Studies above-mentioned have proved that synthetic data from the generative network does improve the performance of FDD. Although generative network still has vast potential to handle small and imbalanced data, there are several problems required to be solved:

a) The FDD classifier for the HVAC system trained under data generated from a generative network shows worse performance than that generated from SMOTE.
b) The existing FDD method for HVAC systems under small and imbalanced data consists of three parts: data augmentation, data evaluation, and classifier training, which is complex and time-consuming to design each process compared with end-to-end learning.
c) Although a related study compares the FDD performance of the HVAC system under small data between data sampling and VAE methods [48], is still a lack of clear comparison of the FDD performance of VRF system between GAN and SMOTE methods under different types of small data.

In this study, the structure of this paper is as follows: Section 2 illustrates the background of the method and the proposed FDD method. The VRF experiment setup and data preprocessing are described in Section 3. Moreover, in Section 4, the comprehensive discussion of FDD results. The conclusions are summarized in Section 5. What is more, the main contributions of this paper can be listed as follows:

a) A combined conditional VAE and Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) generative network is built for more stable data generation.
b) Two modular ensemble classifiers embedded into the generative network are trained simultaneously for faster, more accurate, and stable performance of FDD.
c) An end-to-end FDD method for the VRF system is introduced and validated. The detailed comparison between the proposed FDD method and the SMOTE-based method is discussed comprehensively under different sizes of fault data.

2. Methods

In this section, the background of basic methods will be illustrated comprehensively, including variational Auto-Encoder, Wasserstein Generative Adversarial Network with Gradient Penalty, and ensemble classifier. Then, the structure of the proposed network and all loss functions are described in detail. Finally, the complete FDD strategy and related performance indexes of validation are listed.

2.1. Background

2.1.1. Wasserstein Generative Adversarial Network with Gradient Penalty

The Generative Adversarial Network (GAN) consists of a generator and a discriminator. The generator learns a mapping from random noise to target data, and the discriminator learns to discriminate between real data and generated data [35], such as in Fig. 1. Because the generator and discriminator are interacting with each other, the generator can produce more realistic data.

However, traditional GAN cannot be trained stably because of the game relationship between generator and discriminator. Therefore, Arjovsky et al. [49] modified the loss function of traditional GAN and applied Wasserstein distance to measure the difference between generative data and real data, where Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) is proposed. Gradient Penalty takes the place of weight clipping in WGAN to avoid gradient vanishing and exploding problems [38]. The modified loss functions of the generator and discriminator are displayed as follows:

$$L_D = D(\hat{x}) - D(x) + \lambda\left(\|\nabla_{\hat{x}} D(\bar{x})\|_2 - 1\right)^2 \quad (1)$$

$$L_G = -D(G(r)) \quad (2)$$

Where $\hat{x}$ is the generative data from random noise; x is the real data; $\bar{x}$ is the linear interpolation between generated and real data; λ is the gradient penalty coefficient; r is the random noise. The $\bar{x}$ can be calculated by x and $\hat{x}$ in this formula:

$$\bar{x} = \epsilon x + (1 - \epsilon)\hat{x} \quad (3)$$

Where $\hat{x}$ is calculated by $G(r)$; ϵ is a random number.

2.1.2. Variational Auto-Encoder

Kingma proposed the Variational Auto-Encoder (VAE) in 2013 [39], consisting of an encoder and a decoder. The encoder tries to learn a mapping from real data to normal distributions, and the decoder tries to reconstruct data from random sampling of normal distributions, such as in Fig. 2. Therefore, the decoder can be considered as a generator in GAN.

The encoder is updated by reducing the Kullback-Leibler divergence between a standard normal distribution and latent normal distribution:

$$KL(N(\mu, \sigma^2)\|N(0, 1)) = \frac{1}{2}\left(-log\sigma^2 + \mu^2 + \sigma^2 - 1\right) \quad (4)$$

The $N(\mu, \sigma^2)$ is the learned normal distribution from the encoder. Besides, the decoder is updated by reducing the reconstruction loss between reconstructed data and real data:

$$L_D = \sqrt{\frac{1}{N}\sum\|\tilde{x} - x\|_2} \quad (5)$$

Where $\tilde{x}$ and x are reconstructed and real data, respectively; N is the number of batch sizes. To make the model trainable, a reparameterization trick is used to sample from latent normal distribution as input of decoder:

$$Z = \mu + \varepsilon \times \sigma \quad (6)$$

The ε is a random sampling from a standard normal distribution. The KL divergence and reconstruction loss are interactive and adversarial so that encoder and decoder work in a dynamic balance.

2.1.3. Ensemble classifier

An ensemble classifier is a learning algorithm that combines the prediction results of several multi-base classifiers by bagging, boosting or some other methods to get more accurate classification results [50]. Bagging is first proposed by Breiman [51], and RF is one of the most wide-used bagging ensemble classifiers. In our study, several neural networks are combined as an ensemble classifier by bagging, as shown in Fig. 3.

Every classifier will give each prediction of the testing dataset, and the most occurrence of prediction is chosen as the final prediction. If two or more predictions occur the same number of times, the ensemble classifier will choose one of them randomly as the final output.

2.2. Modified generative network with ensemble classifier

VAE can generate stable data from random sampling but has trouble reconstructing more similar data. GAN can learn the difference between generated and real data, but it is challenging to generate stable data. Therefore, Larsen et al. [52] proposed a combined VAE-GAN model at first. Then, Bao et al. [53] added a classifier into the generative network to generate more stable and real data. Based on the model mentioned above, WGAN-GP replaces GAN, and two bagging network classifiers are embedded into the generative network to get more stable and accurate classification results under small data. Besides, to control the label of generated data, the label information is embedded in all input data except two ensemble classifiers. The combined generative and classifier network is shown in Fig. 4.

The loss of encoder consists of two parts: KL divergence and the mean-square error between real data and reconstructed data:

$$L_E = KL(N(\mu, \sigma^2)\|N(0,1)) + \sqrt{\frac{1}{N}\sum\|x - \tilde{x}\|_2} \quad (7)$$

The x and $\tilde{x}$ are real data and reconstructed data, respectively. The loss of discriminator is the same as that in WGAN-GP.

The loss of generator consists of four parts: one is the formula (2); the second is the formula (5); the third is the average of mean-square error between real data and reconstructed data in

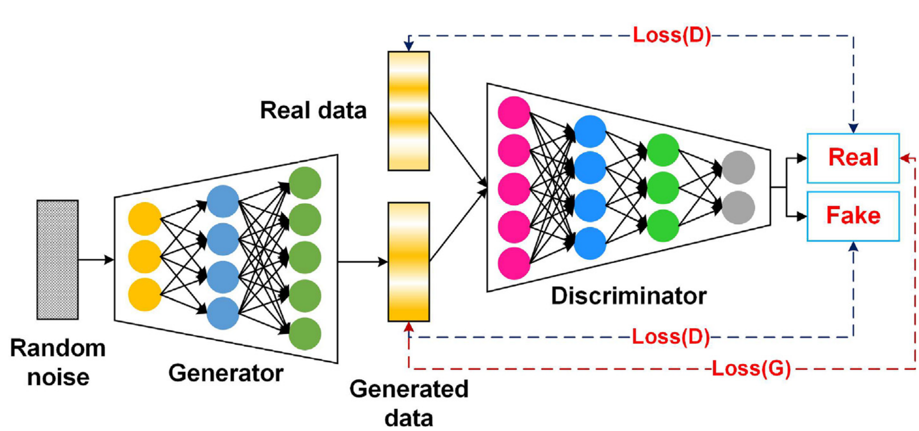

Fig. 1. The structure of GAN.

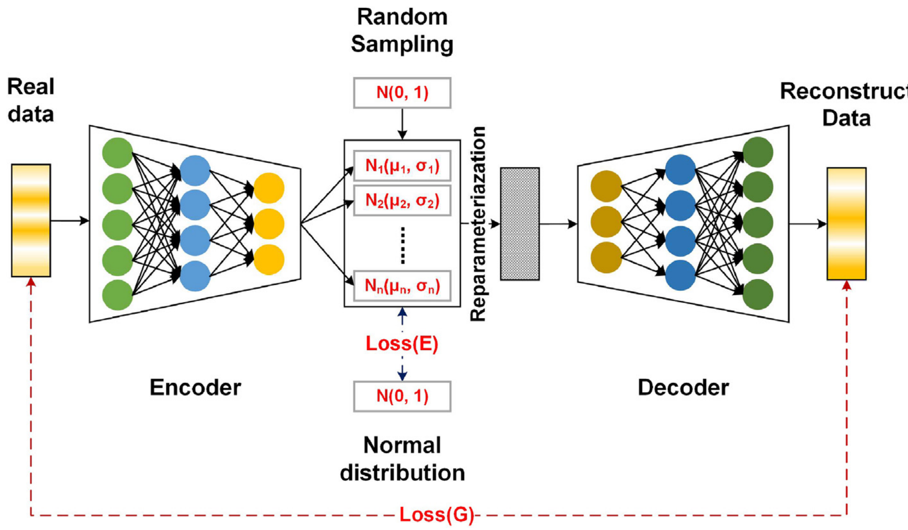

Fig. 2. The structure of VAE.

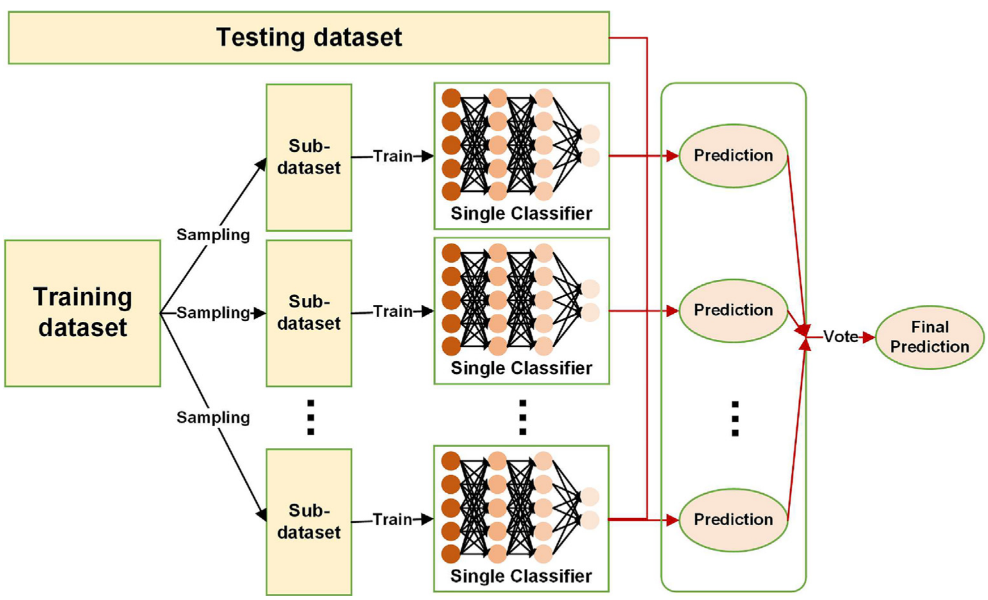

Fig. 3. The structure of the ensemble classifier.

the final layer without being processed by SoftMax of each classifier for fault detection, and the final one is the average of mean-square error between real data and reconstructed data in a final layer without being processed by SoftMax of each classifier for fault diagnosis. Therefore, the whole loss function of the generator can be as follow:

$$\begin{aligned} L_G &= -D(\widehat{x}) + \sqrt{\tfrac{1}{N}\sum \|\widetilde{x} - x\|_2} \\ &+ \tfrac{1}{N_{DT}}\sum_i^{N_{DT}} \sqrt{\tfrac{1}{N}\sum \|\widehat{DT}(x) - \widehat{DT}(\widehat{x})\|_2} \\ &+ \tfrac{1}{N_{DG}}\sum_i^{N_{DG}} \sqrt{\tfrac{1}{N}\sum \|\widehat{DG}(x) - \widehat{DG}(\widehat{x})\|_2} \end{aligned} \tag{8}$$

The N_{DT} and N_{DG} are the number of classifiers for fault detection and fault diagnosis, respectively. The $\widehat{DT}()$ is the final layer without being processed by SoftMax of each classifier for fault detection. The $\widehat{DG}()$ is the final layer without being processed by SoftMax of each classifier for fault diagnosis.

The loss of each classifier is calculated by Cross Entropy (CE) as follows:

$$L_{DT,i} = CE\left(Label(\widetilde{x})_{DT}, DT(\widetilde{x})\right)_i + CE\left(Label(x)_{DT}, DT(x)\right)_i, \quad i = 1, 2, 3 \cdots, N_{DT} \tag{9}$$

$$L_{DG,i} = CE\left(\mathrm{Label}(\widetilde{x})_{DG}, DG(\widetilde{x})\right)_i + CE\left(\mathrm{Label}(x)_{DG}, DG(x)\right), \quad i = 1, 2, 3 \cdots, N_{DG} \tag{10}$$

The DT() and DG() are the final output of the classifier for fault detection and fault diagnosis, respectively. The $\mathrm{Label}(\widetilde{x})_{DT}$ and $\mathrm{Label}(x)_{DT}$ are the real labels for fault detection. The $\mathrm{Label}(\widetilde{x})_{DG}$ and $\mathrm{Label}(x)_{DG}$ are the real labels for fault diagnosis. The training and testing process are shown in Fig. 5. The order of optimization is an encoder, generator, discriminator, and classifiers for fault detection from 1 to N_{DT}, and classifiers for fault diagnosis from 1 to N_{DG}.

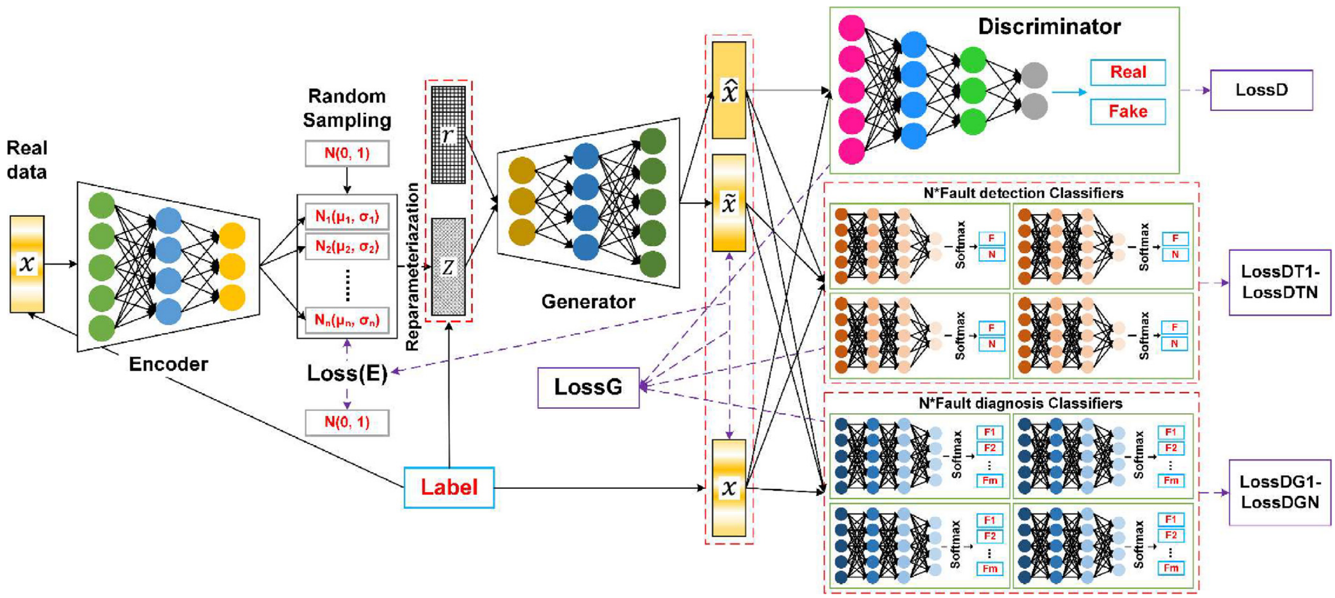

Fig. 4. The structure of the proposed network.

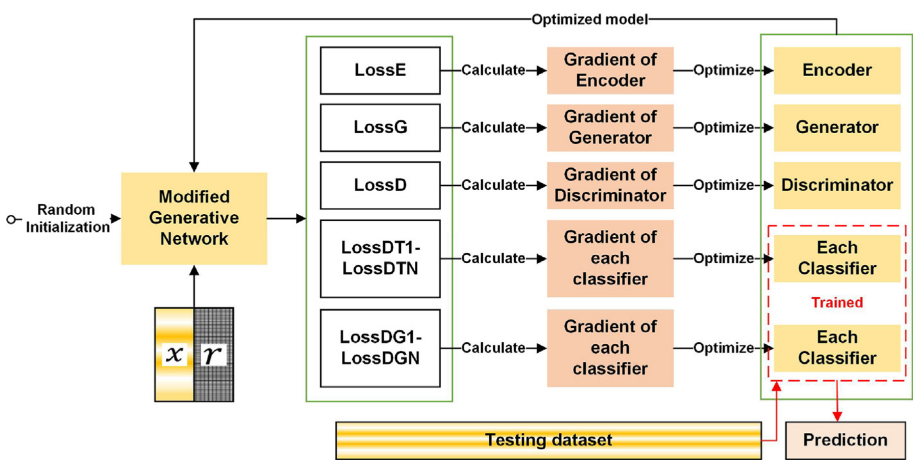

Fig. 5. The training and testing process of modified generative network.

2.3. Proposed FDD strategy

Based on the modified generative network, an FDD strategy is proposed in this study, as shown in Fig. 6. The workflow is composed of three parts. Firstly, the experiment is conducted on the VRF device to collect fault and normal data in data collection and preprocessing. All data is transmitted into the database and manually labeled. After preliminary parameters selection and normalization, the data is divided into training and testing datasets. Secondly, the training dataset is sent into the configured proposed network for training. Thirdly, the trained ensemble classifiers of the proposed network are extracted for testing FDD performance under the testing dataset. Besides, the extra fault data can be used to upgrade trained ensemble classifiers. Several details in the process require to be explained:

a) The experiment data comes from the VRF system in the enthalpy difference laboratory. To create imbalanced and small data, the number of fault training data is much less than normal training data.
b) The parameters of the VRF system are selected based on existing studies and expert knowledge.
c) The generative network is built based on Full Connection Layers (FCL). The whole network is an end-to-end model because data augmentation, data evaluation, and classifier training are compressed into one whole model.
d) The FDD is realized by two bagging ensemble classifiers, where ensemble learning is used to improve the accuracy and generalization.

2.4. Performance indexes

To evaluate the performance of the proposed FDD strategy, four indexes are introduced as follows. The fault detection is considered a binary classification, as shown in Table 1. All faulty data is considered positive in the fault detection phase, and all normal data is negative[44]. Precision shows how many real fault data are in all predicted positives. Recall shows how many fault data are detected correctly in all fault data. F1-score is an overall indicator of precision and recall. Accuracy shows how many data are cor-

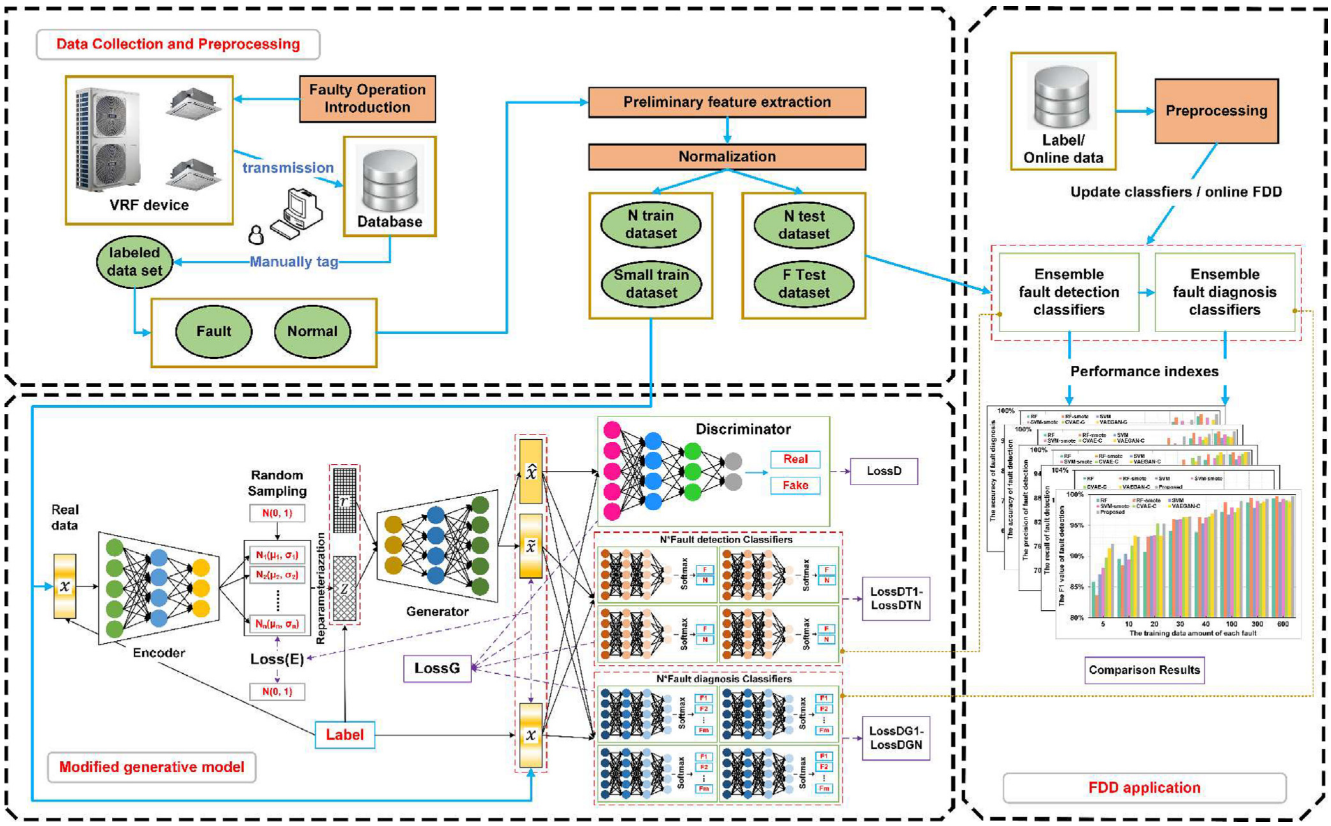

Fig. 6. The proposed FDD strategy.

Table 1
The confusion matrix of binary classification.

		Real data	
		Positive(fault)	Negative(normal)
Predicted data	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

rectly classified in all data. The calculation formulas of the above indexes are listed as follows:

$$Precison = \frac{TP}{(TP+FP)} \tag{9}$$

$$Recall = \frac{TP}{(TP+FN)} \tag{10}$$

$$F1_{score} = \frac{2}{\left(\frac{1}{recall}+\frac{1}{precision}\right)} \tag{11}$$

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{12}$$

The fault diagnosis is considered multi-classification, and overall accuracy is used to compare the performance of different FDD methods.

3. Data preprocessing and experiments

In this section, the experiment setup and data preprocessing will be described, especially about fault introduction, the parameters of the VRF system for FDD, and the detailed dataset information. Besides, the hyperparameters of proposed networks are illustrated.

3.1. VRF system and sensors

The VRF device is placed in an enthalpy difference laboratory [20], which can adjust the indoor and outdoor temperature and humidity in a setting range. The experiment VRF system comprises one outdoor unit (OU) and five indoor units (IU), as shown in Fig. 7. The scroll compressor and air-cooled condenser are located in OU. The rated refrigerating capacity of the whole system is 28 kW, and those of five indoor evaporators numbered from IU1 to IU5 are 3.6, 11.2, 2.2, 5.0, and 7.1 kW, respectively. The refrigerant in the circulatory system is R410A.

There are various sensors placed in different parts to record and store real-time operating data, and some of them have been marked in the system drawing. The operating data are recorded and transmitted every three seconds. All sensors are checked and calibrated carefully, and there is no data missing.

There are more than 800 parameters recorded from the system, including the sensor parameters, control signal, electrical signal, switch signal, et al. The sensor parameters can be divided into IU and OU parameters. The first step is to clean out the useless parameters simply. The remaining parameters are selected by expert knowledge and according to previous studies [10,20,32]. The final parameters used for FDD are displayed in Table 2.

3.2. Dataset descriptions

There were seven types of fault simulated in the experiment VRF system, as shown in Table 3. The refrigerant undercharge and Refrigerant overcharge are system-based faults. The Compressor liquid back, EXV stuck, refrigerant leak, and FRV power-off are component-based faults. A series number from F1 to F7 is used to mark each fault.

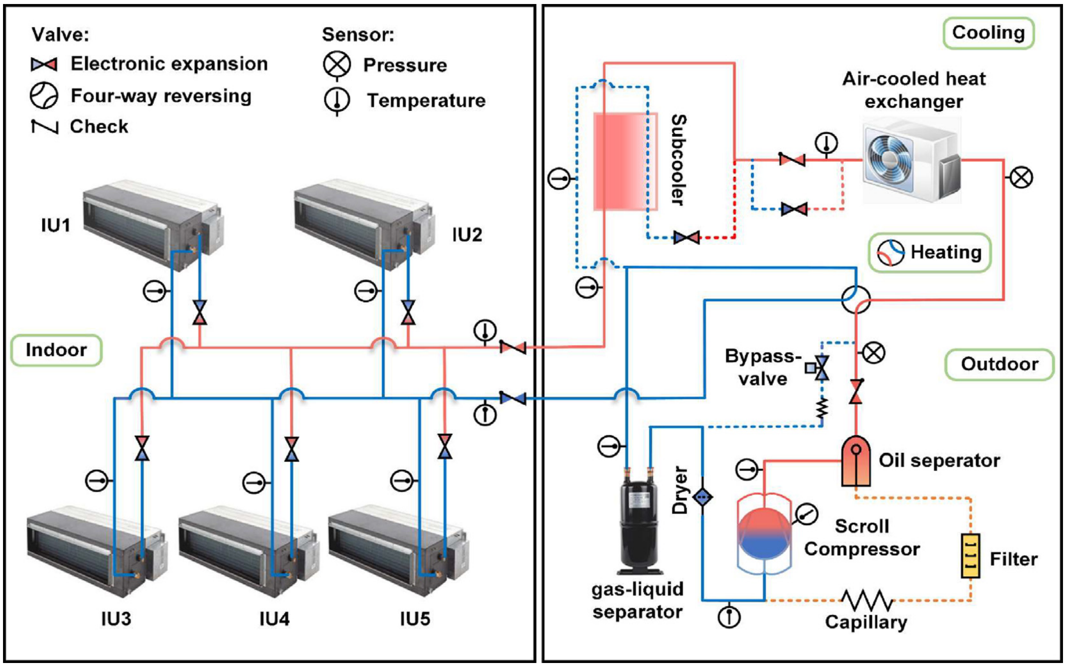

Fig. 7. The configuration of the VRF system.

Table 2
The selected parameters for FDD.

Parameters	Abbreviation	Description
Outdoor unit	T_{out}	Outdoor environment temperature
	$T_{com,d}$	Compressor discharge temperature
	$T_{com,s}$	Compressor shell temperature
	T_{hp}	High-pressure temperature
	T_{def}	Condenser outlet temperature
	$T_{sub,l}$	Super-cooler liquid outlet temperature
	$T_{sub,g}$	Super-cooler gas inlet temperature
	T_{lp}	Low-pressure temperature
	$T_{sep,in}$	Separator inlet temperature
	$T_{sep,out}$	Separator outlet temperature
	C_{now}	Current capability
	f_{com}	Compressor frequency
	f_{fan}	Fan frequency
	I_{com}	Compressor current
	I_{fan}	Fan current
	U_{com}	Compressor bus voltage
	U_{fan}	Fan bus voltage
Indoor unit (i = 1–5)	$T^{i}_{t,in}$	Indoor unit tube inlet temperature
	T^{i}_{in}	Indoor environment temperature
	N^{i}_{ret}	Return air fan rotate speed
	T^{i}_{ret}	Indoor unit return air temperature
	$T^{i}_{t,out}$	Indoor unit tube outlet temperature

The indoor and outdoor conditions under heating/cooling mode and different IU working states combine various operating conditions, as shown in Table 4. The refrigerant leak means that the EXV of non-operating IU is not closed so that only three EXV are working while one non-operating EXV is set as a fault component. The FRV power-off means that the valve cannot be switched in heating because the power is off.

The dataset under a different number of training data has been set, as shown in Table 5. By random sampling from each type of dataset, eight training groups are created.

3.3. Hyperparameters setting of proposed networks

The detailed structural information of the generative network is displayed in Fig. 8. In this experiment, the learning rate of each sub-network is 1e*-3 except discriminator, whose learning rate is 1e*-6 for more stable training. The number of latent normal distributions in the encoder is set as 8. Each training batch dataset is set as 25 fault data and 25 normal data. The ensemble classifier consists of ten base classifiers by bagging, and 6/7 of each batch of data is selected randomly to train each base classifier.

The proposed network is built based on python 3.7 with Tensorflow 2.4 package and run by one GPU GTX1080Ti and two CPUs E5-2620 v4. To keep the final results more representative, each group repeated ten times, and the most stable result was chosen for analysis.

Table 3
The description of seven types of faults.

Fault types	Fault description
Normal (N)	The normal operating of the VRF system
Refrigerant undercharge (F1)	The refrigerant is set as 65% of normal level manually
Refrigerant overcharge (F2)	The refrigerant is set as 130% of normal level manually
Compressor liquid back (F3)	A little liquid refrigerant is injected into compressor inlet manually
EXV stuck at 0% (F4)	The EXV of IU5 is fixed at 0% opening manually
EXV stuck at 100% (F5)	The EXV of IU5 is fixed at 100% opening manually
refrigerant leak (F6)	The EXV of IU3 is set at 50% opening manually
FRV power-off (F7)	The FRV keeps position of cooling mode when heating

Table 4
The different operating conditions.

Operating types	Mode	Working IU	Outdoor condition/°C	Indoor condition/°C
F1	Cooling	Single/Three/All	28–42	20–32
	Heating		−9–9	10–22
F2	Cooling	Single/Three/All	28–42	20–32
	Heating		−9–9	10–22
F3	Cooling	Three	33–34	25–29
	Heating	All	0–12	19–23
F4	Cooling	All	33–42	24–29
	Heating		7–11	20–22
F5	Cooling	All	35–39	25–29
	Heating		7–11	17–22
F6	Heating	Three	7–8	19–22
F7	Heating	All	11–14	17–21
N	Cooling	Single/Three/All	28–42	20–32
	Heating		−9–9	10–22

Table 5
The dataset under different numbers of training data.

Dataset	F1	F2	F3	F4	F5	F6	F7	N
Training group1	5	5	5	5	5	5	5	20,000
Training group2	10	10	10	10	10	10	10	20,000
Training group3	20	20	20	20	20	20	20	20,000
Training group4	30	30	30	30	30	30	30	20,000
Training group5	40	40	40	40	40	40	40	20,000
Training group6	100	100	100	100	100	100	100	20,000
Training group7	300	300	300	300	300	300	300	20,000
Training group8	600	600	600	600	600	600	600	20,000
Test	1200	1200	1200	1200	1200	1200	1200	10,000

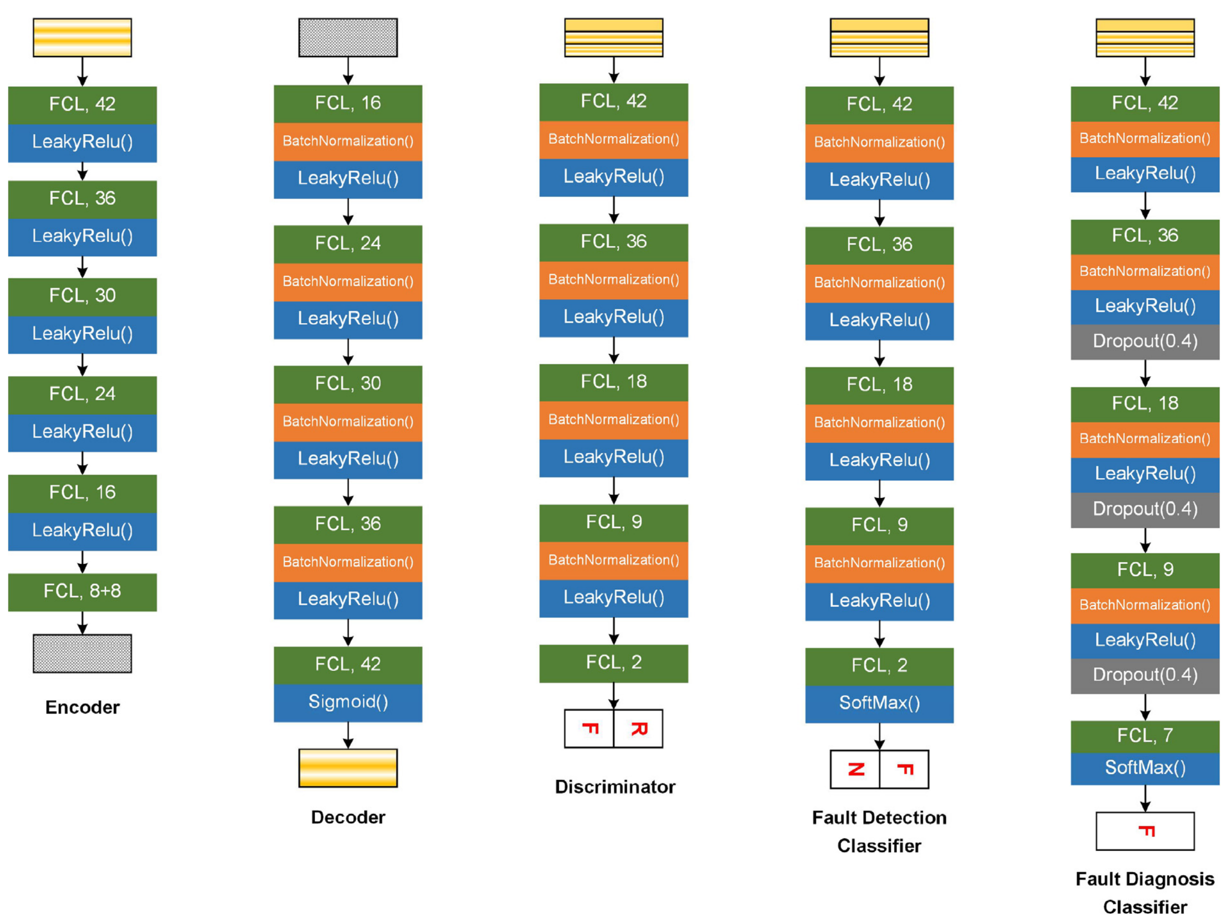

Fig. 8. The parameters setting of generative networks.

4. Analysis and discussion of the proposed FDD method.

In this section, the FDD results of the proposed method compared with six other methods are discussed comprehensively. The comparison groups include RF, RF-SMOTE, SVM, SVM-SMOTE, CVAE with a single classifier, and a proposed generative network with a single classifier.

4.1. The results of VRF fault detection

The VRF fault detection performance indexes are shown in Figs. 9-12. A decreasing trend in the performance improvements can be found in training groups 1 to 8. The SVM-based method shows better performance than RF-based method data when the number of training data is smaller than 40 but performs worse with more data. Besides, generative network with classifier methods have much higher fault detection accuracy than SVM and RF-based methods, especially under training groups 1–3 in Fig. 9. The generative network with classifier methods shows clear improvement compared with RF and SVM-based methods on recall, as shown in Fig. 11. The improvements between the average recall of three generative network with classifier methods and four RF and SVM-based methods are 11.91%, 6.79%, 5.06%, 2.62%, 3.81%, 2.44%, 1.45%, and 0.74% from training group 1 to 8. Such results indicate that generative network with classifier methods can detect more faults from fault data.

However, generative network with classifier methods has lower precision than other methods, as shown in Fig. 10, which means that the fault detection classifier in generative network regards more normal data as fault data. The precision of the generative net-

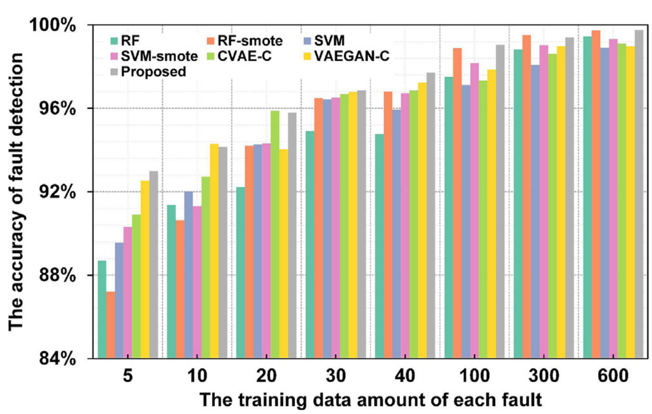

Fig. 9. The accuracy of fault detection.

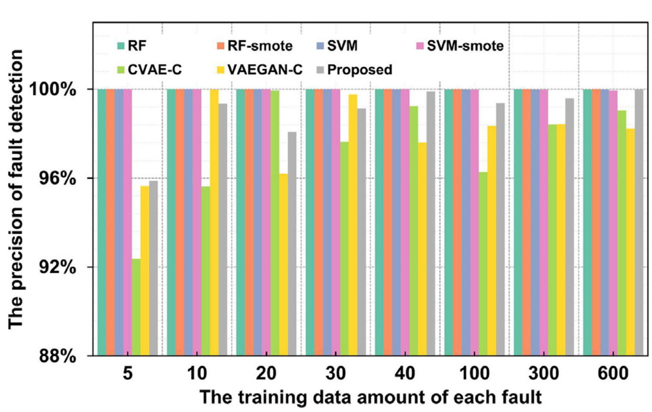

Fig. 10. The precision of fault detection.

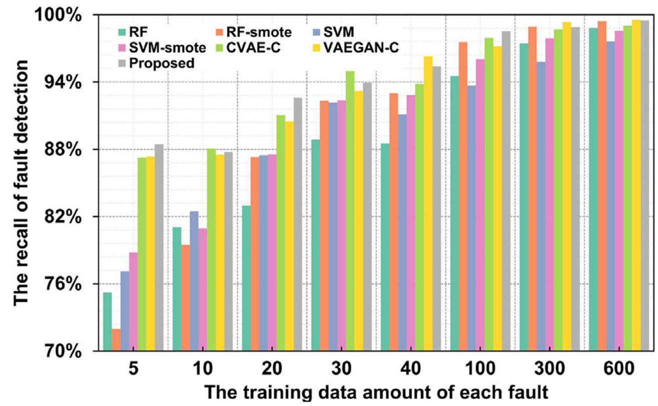

Fig. 11. The recall of fault detection.

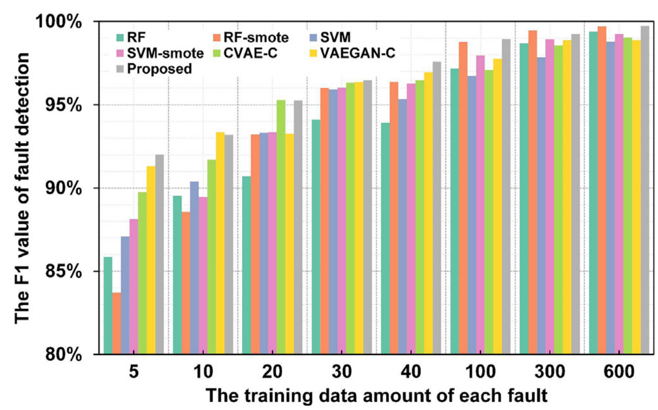

Fig. 12. The F1-score of fault detection.

work with classifier methods is almost 1.91% lower than that of SVM and RF-based methods. The ensemble classifier could improve the performance of the generative network on precision, especially when the number of fault data is higher than 40.

The improvements between generative network with classifier methods and the RF-SMOTE method are analyzed as shown in Fig. 13. This improvement of the F1-score is calculated by setting the RF-SMOTE method as a benchmark. In group 6, the F1-scores of CVAE and the proposed network with a single classifier are almost 1.69% and 1.01% lower than that of the RF-SMOTE method, respectively, while the proposed network with ensemble classifiers is 0.19% higher. Compared with the RF-SMOTE method, the proposed method with ensemble classifiers has an overall average performance improvement of 2.38% while the performance is 0.22% lower under training group 7 only.

The F1-score improvements of the proposed method compared with other methods are displayed in Table 6. The table shows that the proposed method has more than 2% average performance improvement when the number of fault training data is less than or equal to 20, more than 1.3% average performance improvement when the number of fault training data is equal to 40, and 100 and more than 0.5% average performance improvement when the number of fault training data is equal to 30, 300 and 600. This comparison indicates that the proposed method does perform better on fault detection in almost all training groups.

4.2. The results of VRF fault diagnosis

The performance of VRF fault diagnosis is shown in Fig. 14. The proposed method has the best performance on fault diagnosis than

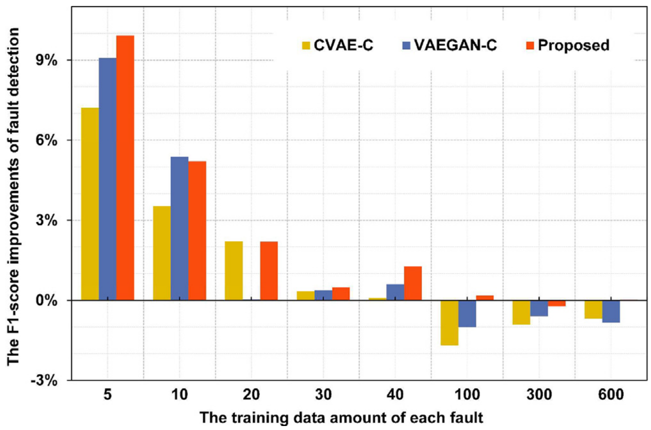

Fig. 13. The F1-score improvements of fault detection based on the RF-SMOTE method.

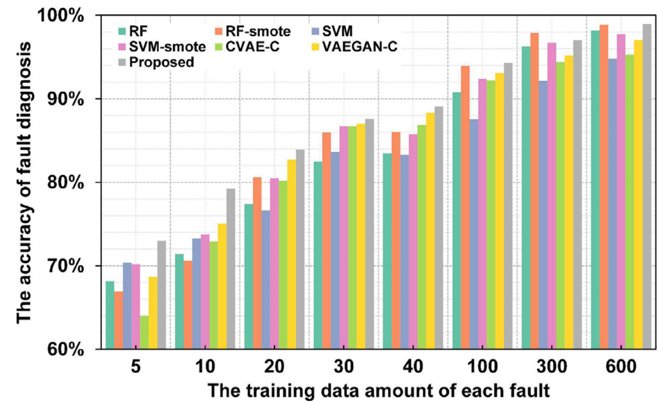

Fig. 14. The accuracy of fault diagnosis.

other methods except under training group 7. Although the CVAE and proposed network with a single classifier perform worst with 5 training fault data compared with the SVM-based method, the proposed network with ensemble classifiers could improve the accuracy of fault diagnosis significantly.

Fig. 15 shows the accuracy improvements of generative network methods compared with the RF-SMOTE method. The proposed method with ensemble classifiers has 12% higher accuracy in group 2 but only 0.86% lower in group 7, while the proposed method with a single classifier has only 6% higher accuracy in group 2 but almost 2.75% lower in group 7. Besides, the proposed network shows better performance than CVAE in all groups.

It is evident in Table 7 that the proposed method has more than 5% performance improvement when the number of fault training data is less than or equal to 20. Only when the number of fault training data is equal to 300 the performance of the proposed method is a little worse than that of the RF-SMOTE method. Compared with propose network with a single classifier, the ensemble classifiers improve fault diagnosis accuracy with the greatest improvement of 5.92% under training group 1 and the lowest improvement of 0.65% under training group 4. The results indicate that the combined network with ensemble classifiers could improve the data-driven fault diagnosis method under small data.

4.3. Diagnosis results comparison based on four methods

The VRF fault diagnosis results of RF-SMOTE, CVAE with a single classifier, proposed generative network with a single classifier, and ensemble classifiers are discussed. Table 8, Table 9, Table 10, and Table 11 show the accuracy of each fault diagnosis result based on four methods. The proposed method doesn't show the best performance on all faults but performs well on most faults. The RF-SMOTE method shows stable performance improvements with data increase, while the performance change of generative network methods is far more fluctuant.

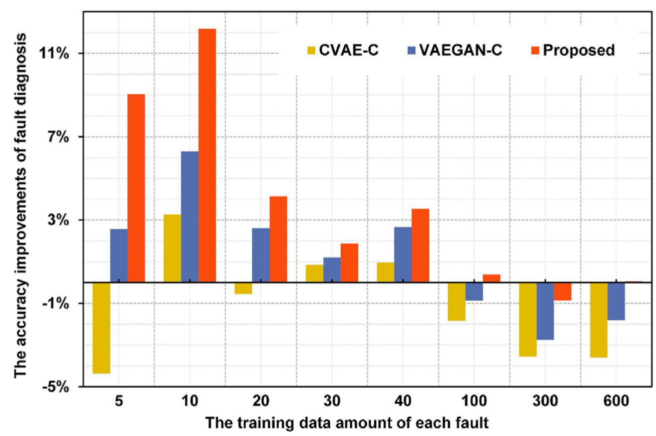

Fig. 15. The accuracy improvements of fault diagnosis based on the RF-SMOTE method.

The accuracy of fault diagnosis on F6 and F7 can reach over 80% with the proposed method or RF-SMOTE method when the number of training data is 5. The possible reason could be that the condition of refrigerant leak and FRV power-off is heating mode with a smaller temperature range compared with other faults. The accuracy of fault diagnosis on F2 can reach 90% only when the number of training data is more than 100, which indicates that the refrigerant overcharge in various conditions is more challenging to diagnose than other faults.

Comparing the accuracy of each fault diagnosis in Table 10 and Table 11, it can be found that the ensemble classifiers perform better on F6 and F7 compared with a single classifier. Although the generative network shows better overall performance on fault diagnosis, no obvious regular pattern can be found: which kind of factors can influence the performance of the generative network and how the influence changes.

Table 6
The F1-score improvements of proposed method compared with other methods.

Method	5	10	20	30	40	100	300	600
RF	6.68%	3.91%	4.79%	2.44%	3.77%	1.79%	0.55%	0.34%
RF-smote	9.03%	4.95%	2.14%	0.48%	1.25%	0.19%	−0.22%	0.02%
SVM	5.37%	2.99%	2.04%	0.56%	2.30%	2.24%	1.40%	0.96%
SVM-smote	4.22%	3.99%	2.00%	0.45%	1.34%	1.00%	0.31%	0.49%
CVAE-C	2.47%	1.60%	−0.02%	0.15%	1.16%	1.87%	0.69%	0.71%
VAEGAN-C	0.77%	−0.16%	2.10%	0.10%	0.65%	1.19%	0.37%	0.86%
Average	**4.76%**	**2.88%**	**2.18%**	**0.70%**	**1.75%**	**1.38%**	**0.52%**	**0.56%**

Table 7
The diagnosis accuracy improvements of proposed method compared with other methods.

Method	5	10	20	30	40	100	300	600
RF	6.62%	9.89%	7.81%	5.83%	6.29%	3.72%	0.77%	0.76%
RF-smote	8.29%	10.86%	3.97%	1.83%	3.42%	0.39%	−0.87%	0.06%
SVM	3.57%	7.54%	8.69%	4.54%	6.48%	7.16%	5.03%	4.15%
SVM-smote	3.82%	6.90%	4.11%	0.99%	3.73%	2.03%	0.33%	1.19%
CVAE-C	12.30%	7.95%	4.50%	1.01%	2.50%	2.22%	2.71%	3.66%
VAEGAN-C	5.92%	5.26%	1.46%	0.65%	0.84%	1.26%	1.90%	1.88%
average	**6.75%**	**8.07%**	**5.09%**	**2.48%**	**3.88%**	**2.80%**	**1.65%**	**1.95%**

Table 8
The accuracy of each fault diagnosis based on RF-SMOTE.

Dataset	F1	F2	F3	F4	F5	F6	F7	Total
5	8.25%	42.08%	86.58%	73.58%	80.17%	86.75%	**91.17%**	66.94%
10	65.83%	35.83%	87.50%	74.58%	60.83%	79.17%	**90.58%**	70.62%
20	89.08%	48.50%	93.17%	75.33%	78.50%	89.83%	89.83%	80.61%
30	96.42%	52.50%	**92.92%**	**87.00%**	88.17%	93.00%	91.92%	85.99%
40	97.00%	58.75%	90.50%	**90.92%**	89.00%	85.08%	91.08%	86.05%
100	98.58%	83.83%	97.25%	95.00%	86.50%	99.75%	96.58%	93.93%
300	**99.83%**	92.50%	98.67%	97.42%	96.92%	99.83%	**99.92%**	**97.87%**
600	**100.00%**	95.58%	**99.58%**	**98.17%**	98.83%	100.00%	**99.92%**	98.87%

Table 9
The accuracy of each fault diagnosis based on CVAE with single classifier.

Dataset	F1	F2	F3	F4	F5	F6	F7	Total
5	46.92%	33.75%	86.00%	37.25%	**86.42%**	90.00%	67.75%	64.01%
10	66.00%	**58.25%**	82.50%	72.33%	60.17%	83.17%	88.08%	72.93%
20	97.00%	46.33%	**94.33%**	60.42%	79.83%	**92.75%**	90.50%	80.17%
30	97.42%	**66.92%**	91.75%	80.92%	78.58%	96.42%	95.00%	86.71%
40	**97.17%**	68.25%	90.50%	71.00%	91.75%	**93.75%**	95.67%	86.87%
100	98.00%	90.17%	91.42%	75.58%	**96.00%**	99.17%	95.08%	92.20%
300	99.50%	90.92%	96.58%	**99.50%**	76.50%	98.75%	99.00%	94.39%
600	99.75%	92.25%	94.67%	98.17%	84.00%	100.00%	98.33%	95.31%

Table 10
The accuracy of each fault diagnosis based on proposed network with single classifier.

Dataset	F1	F2	F3	F4	F5	F6	F7	Total
5	55.08%	38.58%	**88.08%**	**79.33%**	56.08%	81.83%	81.67%	68.67%
10	73.25%	46.92%	85.75%	68.08%	**72.25%**	91.83%	87.33%	75.06%
20	**97.33%**	51.75%	86.50%	73.83%	**89.75%**	91.08%	88.75%	82.71%
30	**98.00%**	61.00%	90.50%	74.58%	**96.83%**	94.58%	93.67%	87.02%
40	96.33%	**80.25%**	90.00%	86.25%	85.75%	89.33%	90.50%	88.35%
100	**98.67%**	89.67%	**94.08%**	85.25%	89.67%	99.25%	95.17%	93.11%
300	99.67%	**95.58%**	97.50%	95.75%	79.17%	99.83%	98.75%	95.18%
600	99.00%	96.67%	99.00%	88.42%	98.17%	98.92%	99.33%	97.07%

Table 11
The accuracy of each fault diagnosis based on proposed network with ensemble classifier.

Dataset	F1	F2	F3	F4	F5	F6	F7	Total
5	**56.67%**	**47.83%**	85.67%	58.25%	81.67%	**90.42%**	90.42%	**72.99%**
10	**85.58%**	53.50%	**89.92%**	**89.50%**	53.67%	**92.42%**	90.00%	**79.23%**
20	97.25%	**54.25%**	87.00%	**87.17%**	80.25%	91.17%	**90.50%**	**83.94%**
30	94.50%	57.58%	91.83%	84.00%	91.17%	**96.75%**	**97.33%**	**87.60%**
40	93.92%	69.42%	**91.58%**	86.83%	**92.58%**	91.92%	**97.42%**	**89.10%**
100	97.92%	**92.00%**	93.00%	**92.83%**	87.50%	**99.92%**	**96.92%**	**94.30%**
300	99.75%	91.58%	**99.00%**	92.33%	**97.67%**	**100.00%**	98.83%	97.02%
600	99.42%	**97.17%**	99.33%	96.92%	**99.92%**	**100.00%**	99.75%	**98.93%**

5. Conclusion

The availability of small and imbalanced data influences the data-driven method's performance for fault detection and diagnosis. Thus, this paper proposes a novel fault detection and diagnosis method to solve this problem. Seven types of faults are applied to verify the proposed method's accuracy and feasibility, and six comparison methods are used to validate the performance improvement of the proposed method. The proposed method's fault detection and diagnosis results compared with other common methods are discussed comprehensively. The main conclusions are summarized below.

(1) The end-to-end generative network with a single classifier model could perform better on fault detection than SMOTE-based methods, according to results analysis for accuracy, recall, and F1-score. The average recall improvements between generative network with classifier methods and RF and SVM-based methods are 11.91%, 6.79%, 5.06%, 2.62%, 3.81%, 2.44%, 1.45%, and 0.74% from training group 1 to 8.
(2) The generative network with the classifier has lower fault detection recall than SMOTE-based methods. The precision of generative network with classifier methods is lower, almost 1.91%, than that of SVM and RF-based methods on average, while ensemble classifiers could improve fault detection precision, especially when the number of fault data is higher than 40.
(3) The embedded ensemble classifier can improve fault detection and diagnosis performance significantly. The greatest and lowest average improvements in fault detection F1-score between proposed and other methods are 4.76% and 0.52% from training group 1 and group 8, respectively. The greatest and lowest average improvements in fault diagnosis accuracy between proposed and other methods are 8.07% and 1.65% from training group 2 and group 7, respectively.

The study proposes a novel and explorable end-to-end data-driven method under small and imbalanced data. The research results validate that a combined generative network can improve fault detection and diagnosis performance compared with SMOTE. Improving the recall of the proposed method and finding the important factors that can control the generative network perform more stable on overall classification tasks will be an explorable task in the future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China. (No. 51876070).

References

[1] B. Li, R. Yao, Building energy efficiency for sustainable development in China: challenges and opportunities, Building Research & Information 40 (4) (2012) 417–431.
[2] S. Lang, Progress in energy-efficiency standards for residential buildings in China, Energy and Buildings 36 (12) (2004) 1191–1196.
[3] X. Yu, D.a. Yan, K. Sun, T. Hong, D. Zhu, Comparative study of the cooling energy performance of variable refrigerant flow systems and variable air volume systems in office buildings, Applied Energy 183 (2016) 725–736.
[4] X. Liu, T. Hong, Comparison of energy efficiency between variable refrigerant flow systems and ground source heat pump systems, Energy and Buildings 42 (5) (2010) 584–589.
[5] H. Wan et al., A review of recent advancements of variable refrigerant flow air-conditioning systems, Applied thermal engineering 169 (2020) 114893.
[6] T.N. Aynur, Variable refrigerant flow systems: A review, Energy and Buildings 42 (7) (2010) 1106–1112.
[7] Y. Li, Z. O'Neill, A critical review of fault modeling of HVAC systems in buildings, Building Simulation 11 (5) (2018) 953–975.
[8] W. Kim, J.E. Braun, Evaluation of the impacts of refrigerant charge on air conditioner and heat pump performance, International Journal of Refrigeration 35 (7) (2012) 1805–1814.
[9] J. Wang, G. Li, H. Chen, J. Liu, Y. Guo, Y. Hu, J. Li, Liquid floodback detection for scroll compressor in a VRF system under heating mode, Applied Thermal Engineering 114 (2017) 921–930.
[10] Z. Li et al., Simulated Annealing Wrapped Generic Ensemble Fault Diagnostic Strategy for VRF System, Energy and Buildings 224 (2020) 110281.
[11] Y. Zhao, T. Li, X. Zhang, C. Zhang, Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future, Renewable and Sustainable Energy Reviews 109 (2019) 85–101.
[12] W. Kim, S. Katipamula, A review of fault detection and diagnostics methods for building systems, Science & technology for the built environment 24 (1) (2018) 3–21.
[13] M.W. Ahmad, M. Mourshed, B. Yuce, Y. Rezgui, Computational intelligence techniques for HVAC systems: A review, Building Simulation 9 (4) (2016) 359–398.
[14] M.S. Mirnaghi, F. Haghighat, Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: A comprehensive review, Energy and Buildings 229 (2020) 110492.
[15] T. Zhang et al., Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions, ISA transactions 119 (2022) 152–171.
[16] Q. Xu, S. Lu, W. Jia, C. Jiang, Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning, Journal of Intelligent Manufacturing 31 (6) (2020) 1467–1481.
[17] K.e. Yan, L. Ma, Y. Dai, W. Shen, Z. Ji, D. Xie, Cost-sensitive and sequential feature selection for chiller fault detection and diagnosis, International journal of refrigeration 86 (2018) 401–409.
[18] Zhang, X. Speech separation by cost-sensitive deep learning. 2017: IEEE.
[19] K.e. Yan, C. Zhong, Z. Ji, J. Huang, Semi-supervised learning for early detection and diagnosis of various air handling unit faults, Energy and buildings 181 (2018) 75–83.
[20] Z. Zhou, H. Chen, G. Li, H. Zhong, M. Zhang, J. Wu, Data-driven fault diagnosis for residential variable refrigerant flow system on imbalanced data environments, International journal of refrigeration 125 (2021) 34–43.
[21] J. Wei et al., New imbalanced bearing fault diagnosis method based on Sample-characteristic Oversampling Technique (SCOTE) and multi-class LS-SVM, Applied soft computing 101 (2021) 107043.
[22] H. Han, X. Cui, Y. Fan, H. Qing, Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features, Applied Thermal Engineering 154 (2019) 540–547.
[23] Y. Gao, S. Liu, F. Li, Z. Liu, Fault detection and diagnosis method for cooling dehumidifier based on LS-SVM NARX model, International journal of refrigeration 61 (2016) 69–81.
[24] A. Ebrahimifakhar, A. Kabirikopaei, D. Yuill, Data-driven fault detection and diagnosis for packaged rooftop units using statistical machine learning classification methods, Energy and buildings 225 (2020) 110318.
[25] F. Jia, S. Li, H. Zuo, J. Shen, Deep Neural Network Ensemble for the Intelligent Fault Diagnosis of Machines Under Imbalanced Data, IEEE Access 8 (2020) 120974–120982.
[26] F. Jia, Y. Lei, N.a. Lu, S. Xing, Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization, Mechanical systems and signal processing 110 (2018) 349–367.
[27] N. Lu, T. Yin, Transferable common feature space mining for fault diagnosis with imbalanced data, Mechanical systems and signal processing 156 (2021) 107645.
[28] Y. Fan et al., Chiller fault diagnosis with field sensors using the technology of imbalanced data, Applied thermal engineering 159 (2019) 113933.
[29] W. Qian, S. Li, X. Jiang, Deep transfer network for rotating machine fault analysis, Pattern Recognition 96 (2019) 106993.
[30] R., M., R. J. and A. M, Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results, in 2020 11th International Conference on Information and Communication Systems (ICICS), 2020.
[31] Qi Hang, Jinghui Yang, Lining Xing, Diagnosis of Rolling Bearing Based on Classification for High Dimensional Unbalanced Data. IEEE, Access 7 (2019) 79159–79172.
[32] Y. Wang et al., Research on diagnostic strategy for faults in VRF air conditioning system using hybrid data mining methods, Energy and Buildings 247 (2021) 111144.
[33] S., C. and L. J. Hybrid Method of Undersampling and Oversampling for Handling Imbalanced Data. in 2018 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). 2018.
[34] Han, H., W. Wang and B. Mao, Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. 2005, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 878-887.
[35] I.J. Goodfellow et al., Generative Adversarial Nets, Advances in neural information processing systems 27 (2014).
[36] M. Lucic et al., Are GANs Created Equal?, A Large-Scale Study (2017).
[37] Arjovsky, M., S. Chintala and B. Léon. Wasserstein Generative Adversarial Networks. in International conference on machine learning. 2017: Proceedings of Machine Learning Research.
[38] I. Gulrajani et al., Improved Training of Wasserstein GANs. (2017).

[39] D.P. Kingma, M. Welling, Auto-Encoding Variational Bayes. (2013).
[40] Zhenxiang Li, Taisheng Zheng, Yang Wang, Zhi Cao, Zhiqi Guo, Hongyong Fu, A Novel Method for Imbalanced Fault Diagnosis of Rotating Machinery Based on Generative Adversarial Networks, IEEE transactions on instrumentation and measurement 70 (2021) 1–17.
[41] B. Zhao, Q. Yuan, Improved generative adversarial network for vibration-based fault diagnosis with imbalanced data, Measurement : journal of the International Measurement Confederation 169 (2021) 1.
[42] Tongyang Pan, Jinglong Chen, Cheng Qu, Zitong Zhou, A method for mechanical fault recognition with unseen classes via unsupervised convolutional adversarial auto-encoder, Measurement science & technology 32 (3) (2021) 035113, https://doi.org/10.1088/1361-6501/abb38c.
[43] Y. Wang, G. Sun, Q. Jin, Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network, Applied soft computing 92 (2020) 106333.
[44] K. Yan et al., Unsupervised learning for fault detection and diagnosis of air handling units, Energy and buildings 210 (2020) 109689.
[45] K., Y. et al., Chiller Fault Diagnosis Based on VAE-Enabled Generative Adversarial Networks, IEEE Transactions on Automation Science and Engineering (2020) 1–9.
[46] K. Yan, A. Chong, Y. Mo, Generative adversarial network for fault detection diagnosis of chillers, Building and Environment 172 (2020) 106698.
[47] Chaowen Zhong, Ke Yan, Yuting Dai, Ning Jin, Bing Lou, Energy Efficiency Solutions for Buildings: Automated Fault Diagnosis of Air Handling Units Using Generative Adversarial Networks, Energies (Basel) 12 (3) (2019) 527, https://doi.org/10.3390/en12030527.
[48] C. Fan et al., Quantitative assessments on advanced data synthesis strategies for enhancing imbalanced AHU fault diagnosis performance, Energy and buildings 252 (2021) 111423.
[49] M, A., C. S and B. L. Wasserstein Generative Adversarial Networks. in International conference on machine learning. 2017: PMLR.
[50] Omer Sagi, Lior Rokach, Ensemble learning: A survey, WIREs Data Mining and Knowledge Discovery 8 (4) (2018), https://doi.org/10.1002/widm.2018.8.issue-410.1002/widm.1249.
[51] BBEIMAN, L. Bagging Predictors. Machine Learning. 1996. 24(2). 123-140.
[52] Larsen, A.B.L. et al. Autoencoding beyond pixels using a learned similarity metric. 2015.
[53] Bao, J., et al. CVAE-GAN_Fine-Grained Image Generation through Asymmetric Training. In Proceedings of the IEEE International Conference on Computer Vision. 2017. ICCV.