

Review

Deep learning in fault detection and diagnosis of building HVAC systems: A systematic review with meta analysis

Fan Zhang ^{a,b}, Nausheen Saeed ^a, Paria Sadeghian ^{a,*}^a School of Technology and Business Studies, Sweden^b Department of Energy Technology, Dalarna University, Falun, 79188, Sweden

ARTICLE INFO

ABSTRACT

Keywords:

HVAC
Deep learning
Systematic review
Meta analysis
Fault detection and diagnosis

Building sector account for significant global energy consumption and Heating Ventilation and Air Conditioning (HVAC) systems contribute to the highest portion of building energy consumption. Therefore, the potential for energy saving by improving the efficiency of HVAC systems is huge and various fault detection and diagnosis (FDD) methods have been studied for this purpose. Although amongst all types of existing FDD methods, data-driven based ones are regarded as the most effective methods. As a relatively new branch of data-driven approaches, deep learning (DL) methods have shown promising results, a comprehensive review of DL applications in this area is absent. To fill the research gap, this systematic review with meta analysis analyses the relevant studies both quantitatively and qualitatively. The review is conducted by searching Web of Science, ScienceDirect, and Semantic search. There are 47 eligible studies included in this review following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol. 6 out of the 47 studies are identified as eligible for meta analysis of the effectiveness of DL methods for FDD. The most used DL method is 2D convolutional neural network (CNN). Results suggest that DL methods show promising results as a HVAC FDD. However, most studies use simulation/lab experiment data and real-world complexities are not fully investigated. Therefore, DL methods need to be further tested with real-world scenarios to support decision-making.

1. Introduction

According to the latest statistics [1] of 2022, building sector accounts for one-third of global energy consumption and has been identified as one of the main climate changers. Heating, Ventilation and Air Condition (HVAC) systems contribute to the highest portion (38%) of building energy consumption. However, HVAC systems can be faulty [2] leading to both system performance degradation and energy penalty [3,4]. However, energy saving of 5% - 30% is achievable by applying Fault Detection Diagnosis (FDD) [5].

Kim and Katipamula [6] provided a general review of methods for HVAC FDD, which were classified into three categories, quantitative model-based, qualitative model-based, and process history-based methods. According to the authors, data-driven approaches were the most suitable to be applied to complicated systems such as HVACs, where detailed knowledge of the physics of the system is not available. Third, they are robust to noise and can extract the underlying structure of a data set. The main drawback of their review is that the focus area of

this study is not clear due to the broad scope. Therefore, only a few methods were selectively reviewed by the authors without deep analysis, the selection criteria used in this review was not justified either. In addition, since modern techniques such as DL methods were not well developed at that point, the methods reviewed in this study were outdated.

Shi and O'Brien [7] separated the HVAC FDD process into several sub-tasks. The main tasks identified by the authors were feature generation, fault detection, symptom generation, fault diagnosis, etc. These relevant tasks and components were reviewed in this study. However, the scope of this study is also broad since its focus area is the entire development process of HVAC FDD. As a result, each component involved was discussed on a general level, and a detailed discussion of various methods was not available in this review.

Zhao et al. [8] reviewed the FDD methods for building energy systems up to the year 2018, with a focus on algorithms of Artificial Intelligence (AI). It was found in this review that physical models played a dominant role in FDD in the early years, while the use of AI models

* Corresponding author.

E-mail address: psd@du.se (P. Sadeghian).

increased significantly in the recent decade. In addition, the application of Deep Learning (DL) methods was not found in this area until 2018 and therefore, there was one DL relevant paper by Guo et al. [9] reviewed in this study.

Similar to Zhao et al. [8], data-driven FDD methods were reviewed by Mirnagh et al. [10]. With regard to faults, chiller and air handling unit (AHU) systems were identified as the two major components of interest in terms of the severity and the cost. The potential of using DL for FDD was identified in this study. However, they found that the application of DL in this area was limited, since only DL-relevant studies [9,11] were covered in this review.

Li et al. [12] reviewed various algorithms used for feature engineering involved in the development of an FDD approach. Five DL papers [13–16,9] were reviewed in this study. They concluded that DL methods were useful feature extractors. However, the focus area was limited to feature engineering. Therefore, the perspective from this study regarding DL methods for FDD was not somewhat narrow.

Buffa et al. [17] reviewed the control and fault detection strategies for District Heating and Cooling (DHC) systems. The focus was to compare traditional control strategies with more advanced control methods that have been applied to the fourth and fifth generation DHC systems. With regard to FDD methods and their applications to the DHC systems were less discussed. In addition, there were only two DL studies [16,18] reviewed in this study.

It is concluded from the existing reviews published in the last five years that DL methods for HVAC FDD is a relatively new that just started in 2018, though DL methods have been reported to be more effective form HVAC FDD in recent studies.

There are two main reasons make DL methods perform better than conventional machine learning methods for HVAC FDD. First, due to components coupling, changing thermodynamics, local, and long-term temporal decencies, etc., operational data of HVAC systems are complex in nature, while deep learning methods have been proved to be more effective in terms of modelling such complexity and multi-scale temporal-dependencies of HVAC faults [19]. Second, data scarcity is a common problem in the field of HVAC FDD, which leads to performance degradation of conventional machine learning methods, deep learning based generative models have been successfully in recent HVAC FDD studies [20] to mitigate the data scarcity issue.

However, none of the existing reviews regarding HVAC FDD was conducted systematically, and no quantitative analysis was conducted to investigate the effectiveness of DL method for HVAC FDD.

On the other hand, a systematic review delivers a clear and comprehensive overview of available evidence on DL methods for HVAC FDD. Moreover, it helps to identify research gaps in the current understanding of this field. It can highlight methodological concerns in relevant studies that can be used to improve future work in this area. Lastly, it can be used to identify questions for which the available evidence provides clear answers and thus for which further research is not necessary.

Against this background, the aim of this study is to provide a systematic review of DL methods for HVAC FDD as well as evaluate their effectiveness through meta analysis. The following research questions are the focus of this review.

- Which DL methods have been used for HVAC FDD?
- What are the advantages and disadvantages?
- Are DL methods effective for HVAC FDD?
- What are the potential areas of improvement?

To answer questions 1,2,4, a qualitative analysis of the included papers in the systematic review is conducted, to answer question 3, a quantitative meta analysis based on an eligible subset of reviewed papers is conducted. The scope is for complete HVAC systems as well as the main components, e.g., chiller, air handling unit, etc.

The rest of the paper is organized as following, Section 2 describes

the methods used for systematic review and meta analysis. Section 3 provides an overview of the 47 included studies based on the method in Section 2.1, while Section 4 compares different DL methods used in the 47 studies qualitatively. Section 5 performs the meta analysis based on 6 eligible studies filtered according to the method used in Section 2.2, which analyses the effectiveness of DL methods for HVAC FDD quantitatively. The current status and improving areas are discussed in Section 6, followed by the conclusions. A full list of abbreviations used in this paper is summarized in the appendix.

2. Methods

In this study a systematic review is conducted following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) protocol [21]. Three independent researchers have been involved in collecting and reviewing the data according to the protocol. Web of Science, Semantic Scholar, and Scopus are used to identify studies regarding fault detection and diagnosis of building HVAC systems using DL methods. In addition, reference checking is performed to retrieve the relevant papers that are potentially missed out via database searching. This review is complemented with a Meta-analysis that analyses the effectiveness of DL methods using a subset of eligible studies.

2.1. Search strategy

The process starts with defining keywords to build the search strings. Keyword combinations, the corresponding synonyms, and abbreviations relevant to three subjects: “building HVAC”, “fault detection and diagnosis”, and “deep learning” are used for searching. The detailed search query is specified below:

(LSTM OR CNN OR convolutional OR long short OR deep OR GAN OR adversarial OR transfer OR graph OR DNN OR DANN OR BDLSTM OR recurrent OR generative OR GNN OR RNN OR GRU) AND (fault detection and diagnosis OR fault diagnosis OR fault detection and) AND (HVAC OR air conditioning OR ventilation OR heating OR cooling OR district heating OR building OR heat pump OR chiller OR air handling unit OR heat pumps OR chillers OR air handling units)

To assure the quality of search results, the keyword list is reviewed and agreed upon by three independent reviewers. In addition, an external domain expert is involved in the process of keywords discussion and double-checking.

After finalizing the keywords, eligibility criteria are specified for filtering out non-relevant studies. To be more specific, the studies are included in this review only if all the following criteria are met: 1) the scientific work is written in English; 2) published in peer-reviewed scientific journals, conference proceedings, or books; 3) the studied HVAC systems are used in buildings, not in other domains; 4) implementation of fault diagnosis should be present; 5) deep learning methods are used. 6) not a review article.

The literature search is performed on 25th March 2022. The overall process and results are shown in Fig. 1.

From Fig. 1, 139, 2121, and 184 articles are retrieved from Web of Science, Semantic search, and Scopus, respectively. After removing duplicates, there are 2264 articles left for the first stage screening based on the titles and abstracts, amongst which, 2173 articles are excluded, the reason for most excluded articles is because they are not relevant to DL or FDD or building HVAC. After this stage, there are 90 articles available for the second stage screening, the full text of each article is checked thoroughly at this stage, and 47 articles fulfil the eligibility criteria. In addition, reference lists of previously excluded review articles at the screening phase are re-examined to avoid any potentially relevant studies that are missed out. All three reviewers discuss and agree upon results at each stage from the flowchart. Finally, there are 47 eligible articles included in this review.

Among these 47 articles, if one article does not report the confusion matrix or its dataset is not accessible, this article is included for

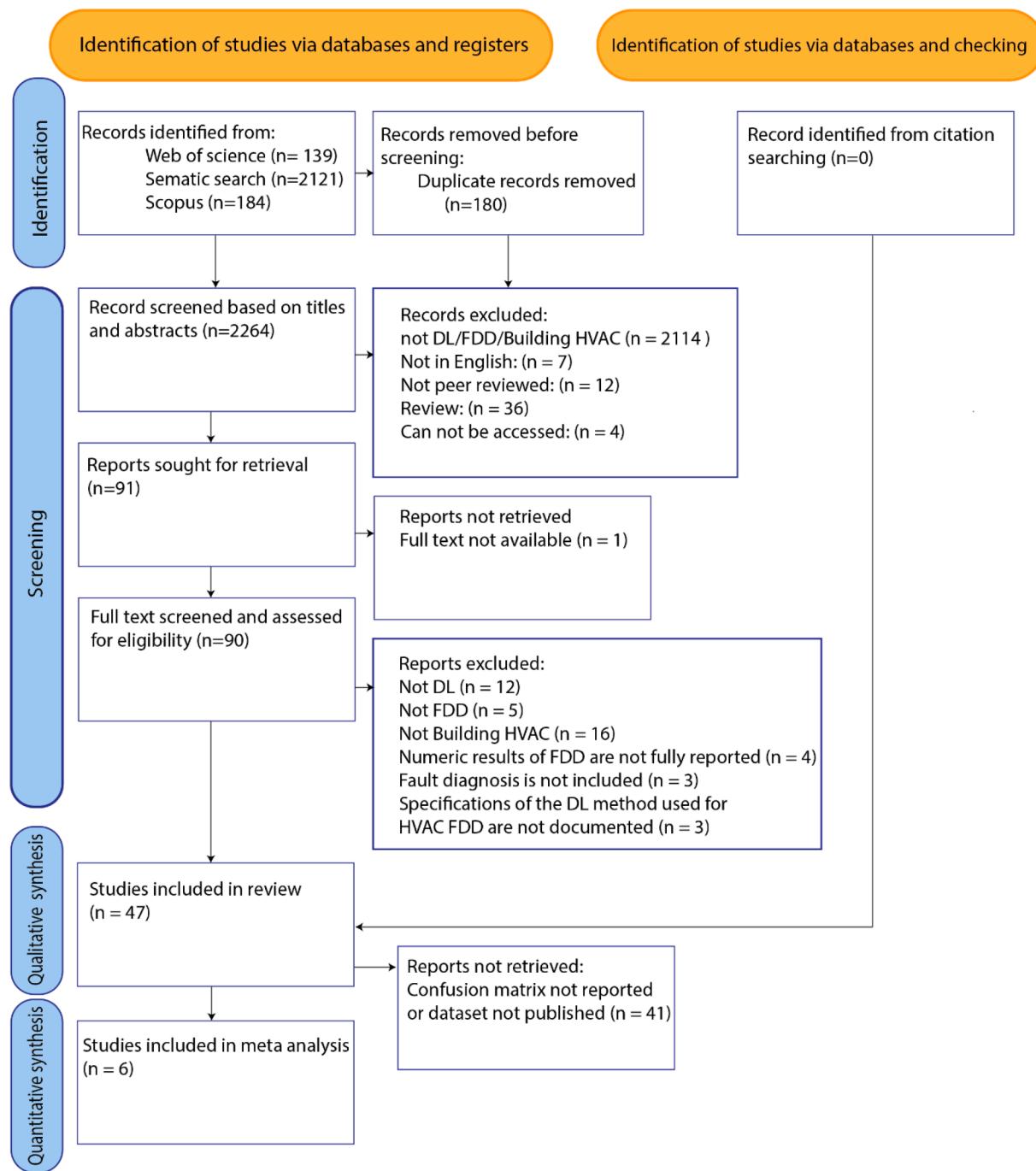


Fig. 1. Flowchart of the systematic review process.

qualitative analysis but excluded for meta analysis. As a result, 6 studies [22–27] ,are identified as eligible for meta analysis.

2.2. Qualitative synthesis

The 47 eligible papers for qualitative synthesis are reviewed and summarized in detail from various dimensions. To be more specific, information extracted from these papers includes the DL building blocks used, benchmarked methods, and the network structure type, methods-wise. Fault-wise and HVAC system-wise, information about the faults and system types is extracted. Data-wise, the description of the dataset, fault generation methods, and data format are summarized. In addition, other supplementary information such as whether the proposed approach is validated using a real-world dataset, whether FDD of faults

at different severity levels is performed, etc. is extracted.

2.3. Data for quantitative synthesis

The American society of heating, refrigerating and air-conditioning engineers research report 1043 (ASHRAE RP-1043) dataset [28] is used in all six studies eligible for meta analysis, which included seven types of faults, i.e., condenser fouling, excess oil, refrigerant leak, refrigerant overcharge, reduced condenser water flow, reduced evaporator water flow, non-condensable gas in refrigerant of four severity levels, amongst which, confusion matrix of all severity levels are reported in [22]. The severity level of the confusion matrix used in [23,24] is unknown, the confusion matrix of severity level one is reported in [25], and that of severity level two is reported in [26]. Although the

confusion matrix of all four levels is reported in [27], the severity level four confusion matrix is not correctly constructed due to one missing category. Therefore, only the results of the other three confusion matrix reported in this study are used for meta-analysis.

To fulfil the requirement that the input format has to be a 2×2 confusion matrix, the results of the included studies are categorized into two classes and aggregated across all severity levels using the one versus the rest approach [29]. For the quantitative analysis of fault detection, results of the seven faults are consolidated together into one faulty class versus the normal class, while performing quantitative analysis for diagnosing a specific type of fault, results of the rest classes are consolidated together.

2.4. Quantitative synthesis

There are two types of meta analytic summary: summary points, e.g., summary sensitivity, specificity, diagnostic odds ratio (DOR), and summary lines, i.e., summary receiver operating characteristic (SROC) curve, which shows summary of test performance, visual assessment of threshold effect, and heterogeneity of data in ROC space between sensitivity and specificity [30]. Summary points are based on true positive (TP), true negative (TN), false positive (FP), and false negative (FN) calculated from the confusion matrix.

To be more specific, sensitivity is the proportion of positive results out of TP results and specificity, is the proportion of negative results out of TP results. In the context of DL methods for HVAC FDD, sensitivity represents the ability of DL methods to correctly identify the system with faults. Specificity represents the ability of DL methods to correctly identify the system without faults.

They are given in Eqs. (1, 2).

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2)$$

DOR is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease. In terms of HVAC FDD, it describes the odds of positive fault detection in HVAC systems with faults relative to the odds of positive fault detection in those without any faults. This measure incorporates information about both sensitivity and specificity and tends to be reasonably constant despite the diagnostic threshold and is regarded as one of the most useful single indicators for diagnostic testing [31]. DOR is calculated using Eq. (3).

$$\text{DOR} = \text{Sensitivity} * \text{Specificity} / (1 - \text{Sensitivity}) * (1 - \text{Specificity}) \quad (3)$$

One common problem of ratio summary statistics is that the number

scale is not symmetric; e.g., a DOR of 0.5 (a halving) and a DOR of 2 (a doubling) are opposites such that they should average to no effect, the average of 0.5 and 2 is not a DOR of 1 but a DOR of 1.25. Therefore, log transformation is commonly used in the literature to avoid the asymmetric problem, and it is adopted in this study for the same reason.

Because DOR alone does not reflect the variance [32], an SROC curve is constructed to analyse the relation between sensitivities and specificities.

After data extraction and formatting, separate effective sizes from each study are pooled together using random-effect models to incorporate the heterogeneity between different studies [33,34]. Sensitivity analysis is performed to evaluate the influence of a single study on the pooled results using the leave one out approach [35,36]. Funnel plots [37] are used to analyse the potential publication bias caused by the selective publication of favourable results or statistically significant results. The overall process of meta-analysis is shown in Fig. 2.

3. Results

This section summarizes all the information extracted from the 47 eligible papers remaining after the systematic review process of Section 2. Characteristics of included studies are summarized in Table 1.

From Table 1, DL based hybrid methods were used in nine studies, while the remaining 38 studies used single DL models. Of these 38 studies, eleven applied 2D CNN, and the least applied methods were convolutional variational autoencoder (CVAE), bi-directional gated recurrent unit (BDGRU), and deep bidirectional long short-term memory neural networks (BDLSTM); there was one study per method. There are six studies each for deep MLP and LSTM methods, followed by generative adversarial network (GAN), 1D CNN and deep belief network (DBN), which are 5, 4, and 3 studies, respectively. Different DL methods in terms of HVAC FDD are discussed in the below.

4. Discussions of different DL methods for HVAC FDD

4.1. Deep MLP based methods

From the reviewed literature, vanilla deep MLP-based is less used in recent years; three out of six deep MLP studies were conducted in 2019. Because training a deep MLP model using the conventional back-propagation approach with random initialization suffers from the gradient vanishing problem. In addition, the deep multilayer perceptron (MLP) model can get stuck in a local optimal if the amount of training data is small [71], which is common in terms of HVAC FDD since operation data of HVAC systems with labelled faults are difficult to be acquired. Therefore, the performance of a deep MLP model can be even

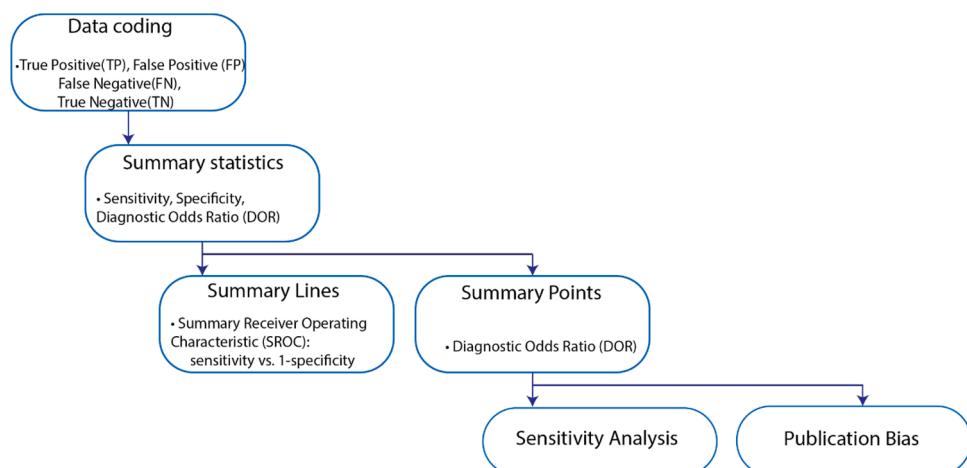


Fig. 2. Flowchart of meta-analysis.

Table 1

Characteristics of the reviewed studies.

Study	Year	Proposed method(s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[9]	2018	DBN	DBN	Air Conditioning	✗	MLP	Dataset of a VRF system with five IDUs from a standard enthalpy difference laboratory	✓	✗	✓	✗	✓	✗	✗	four-way reversing valve fault, outdoor unit fouling fault, refrigerant undercharge and overcharge faults
[38]	2018	2D CNN	2D CNN	Chiller	✗	SVDD, PCA	ASHRAE RP-1043	✓	✗	✓	✓	✗	✗	✗	Refrigerant overcharge, reduced condenser water flow rate, evaporator water flow rate, refrigerant leakage, excess oil, condenser fouling, non-condensable gas
[39]	2019	2D CNN	2D CNN	Heat Source System With Water Thermal Storage Tanks	✗	2D CNN	simulation data + data from an office building in Tokyo, Japan	✓	✓	✓	✓	✓	✗	✗	Fault 1: chiller deterioration by condenser fouling, Fault 2: improper sewage pump set value, Fault 3: heat exchanger fouling, Fault 4: temperature sensor error at the supply side of the heat exchanger, Fault 1 + Fault 2, Fault 1 + Fault 3, Fault 1 + Fault 4, Fault 2 + Fault 3, Fault 2 + Fault 4, Fault 1 + Fault 2 + Fault 3, Fault 1 + Fault 2 + Fault 4, Fault 2 + Fault 3 + Fault 4, Fault 1 + Fault 2 + Fault 3 + Fault 4
[15]	2019	Convolution-sequence (C-S model) + 1D CNN	1D CNN + LSTM (C-S)	Heat Pump	✓	1D CNN, C-S model + MLP, MLP	Simulation data + winter	✓	✗	✓	✗	✓	✗	✗	Condenser fouling, evaporator

(continued on next page)

Table 1 (*continued*)

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method (s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[42]	2019	1D CNN+WaveCluster	1D CNN	Air Handling Units	×	AUX-NN, CB-NN	simulation data of an AHU in Ouagadougou, Africa	✓	✗	✓	✗	✓	✗	✗	temperature after throttling, Chilled-water supply temperature, Chilled-water return temperature, Compressor suction pressure, Compressor discharge pressure, Inlet pressure of the throttle device, Outlet pressure of the throttle device sensors Supply air temperature, supply water temperature, return water temperature, and chilled water flow rate sensors fault
[43]	2019	LSTM	LSTM	Ventilation And Heating System	×	LSTM	simulation data of a demand-controlled ventilation and heating system	✓	✗	✓	✗	✓	✗	✗	Gain Fault in Temperature Sensor, Offset Fault in Temperature Sensor, Damper Fault On, Damper Fault Off, Heater Fault On, Heater Fault Off, Battery Fault in Temperature Sensor, Battery Fault in CO2 Sensor, Gain Fault in CO2 Sensor and Communication Fault in Router Reduced condenser water flow, Reduced evaporator water
[44]	2019	LSTM	LSTM	Chiller	✗	GRU, RNN	ASHRAE RP-1043	✓	✗	✓	✗	✓	✗	✗	(continued on next page)

Table 1 (continued)

Study	Year	Proposed method (s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[45]	2019	LSTM	LSTM	Chiller	✗	LR, MLP	dataset of an air-cooled chiller system in the Honeywell Home and Building Control Laboratory in Tianjin University, China	✓	✗	✓	✗	✓	✗	✗	flow, Condenser fouling, Non-condensables in refrigerant, Refrigerant leak sensor faults of Compressor suction temperature, Compressor discharge temperature, Condenser-air temperature at the outlet, Refrigerant temperature before throttling, Refrigerant temperature after throttling, Chilled-water supply temperature, Chilled-water return temperature, Compressor suction pressure, Compressor discharge pressure, Inlet pressure of the throttle device, Outlet pressure of the throttle device
[46]	2019	WGAN-ensemble + KNN/DT/MLP/ SVM/RF	GAN	Air Handling Units	✗	WGAN-SVM/ GAN-ensemble/ GAN-SVM + KNN/DT/MLP/ SVM/RF, semi-supervised SVM	ASHRAE RP-1312	✓	✗	✓	✗	✓	✗	✗	Exhausted air damper stuck, Return fan at fixed speed, Cooling coil valve control unstable, Cooling coil valve partially closed, Outdoor air damper leak, AHU duct leaking

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method(s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[47]	2019	AANN	Deep MLP	HVAC System	✓	PCA	Data collected from the simulation model in TRNSYS for several modes of the HVAC system cooling operation	✓	✗	✓	✓	✓	✗	✗	Faults of two zones, the chilled water tank, the cooling coil return water, and the AHU output air temperature sensors
[48]	2019	DBM	DBN	Heating System	✗	MLP, SVM	Two datasets from the centralized heating pipe networks: the branch pipe network and the loop pipe network.	✓	✗	✓	✗	✓	✗	✗	pipe leakage
[49]	2019	SAE	Deep MLP	Air Conditioning	✓	EMD+SVM, VMD+SVM	Data of vibration signals from an air conditioner collected by the Laser Doppler Vibrometer (LDV) LV-S01		✗	✓	✓	✓	✗	✗	Compressor lacks soundproof cotton, Fan blade cracks, fan blade weight imbalance, Compressor copper tube lacks, damping block, Fan bracket loose
[14]	2019	Deep MLP	Deep MLP	Air Handling Units	✗	Deep MLP	Simulation data of a 10-floor office building located in Taipei, Taiwan	✓	✗	✓	✗	✓	✗	✗	Abnormal chilled water flow rate, Abnormal supply air flow rate, Abnormal sensor measurement errors
[50]	2019	DBM	DBN	Heating System	✗	MLP, SVM	Dataset from the pipe network of a heating system consisted of 13 pipe sections and 10 nodes	✓	✗	✓	✗	✓	✗	✗	pipe leakage
[13]	2020	2D CNN	2D CNN	Heat Source System With Water Thermal Storage Tanks	✗	2D CNN	simulation data + data from an office building in Tokyo, Japan	✓	✓	✓	✗	✓	✗	✗	Fault 1: chiller deterioration by condenser fouling, Fault 2: improper sewage pump set value, Fault 3: heat

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method (s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[51]	2020	2D CNN	2D CNN	HVAC System	✗	MLP, PCA, SVM	simulation data of a one-floor office building in Doha city and Qatar	✓	✗	✓	✗	✗	✗	✗	exchanger fouling, Fault 4: temperature sensor error at the supply side of the heat exchanger, Fault 1 + Fault 2, Fault 1 + Fault 3, Fault 1 + Fault 4, Fault 2 + Fault 3, Fault 2 + Fault 4, Fault 1 + Fault 2 + Fault 3, Fault 1 + Fault 2 + Fault 4, Fault 2 + Fault 3 + Fault 4, Fault 1 + Fault 2 + Fault 3 + Fault 4 Zone 1 VAV box damper: stuck damper at actuator 1, Zone 2 VAV box damper: stuck damper at actuator 2, Zone 3 VAV box damper: stuck damper at actuator 3, Water valve: stuck valve at actuator 4
[27]	2020	1D CNN + GRU	1D CNN + GRU	Chiller	✗	1D CNN, GRU-LSTM, BPNN, PCA BPNN, 1D CNN LSTM	ASHRAE RP-1043	✓	✗	✓	✗	✓	✗	✗	Condenser fouling, excess oil, non-condensable in refrigerant, reduced evaporator water flow, reduced condenser water flow, refrigerant leak, refrigerant overcharge
[52]	2020	SDCNNs	Ensemble 1D CNN	Chiller	✗	1D CNN, SVM, DT, LightGBM, NB	ASHRAE RP-1043	✓	✗	✓	✗	✗	✗	✗	Condenser fouling, Excess oil, Non-condensables in

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method(s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types	
[53]	2020	2D CNN	2D CNN	District Heating	×	2D CNN, KNN, RF, MLP	simulation data of an integrated energy-based direct district system	real building data	simulation/lab experiment data	Numerical	Image	✓	✗	✓	✗	refrigerant, Reduced condenser water flow, Reduced evaporator water flow, Refrigerant leak, Refrigerant overcharge temperature sensor of the PI for electric boiler, temperature sensor of the PI for gas-fired boiler, flow rate sensor of the PI for pumps in the primary network, electric water boiler, gas-fired water boiler, pump, solar water heater, heat exchanger, pipeline faults and 21 sub-faults of three types: bias, drift and complete failure 9-level refrigerant charge faults
[54]	2020	Ensemble 1D CNN	Ensemble 1D CNN	Air Conditioning	×	1D CNN, DT, MLP, SVM, DBN	dataset from a refrigerant charge fault experiment bed of a commercial VRV system	✓	✗	✓	✗	✓	✗	✗	✗	Energy leakage
[36]	2020	2D CNN	2D CNN	District Heating	×	LR, LDA, L-SVM, GNB, KNN, DT, RF, AB	infrared images data from 12 cities in Denmark	✓	✓	✓	✗	✗	✗	✗	✗	Reduced condenser water flow, Reduced evaporator water flow, Refrigerant leak, Refrigerant Overcharge, Excess Oil,
[55]	2020	CWGAN+SVM/RF/DT/BN/KNN/LR	GAN	Chiller	×	SVM, RF, DT, BN, KNN, LR	ASHRAE RP-1043	✓	✗	✓	✗	✗	✗	✗	✗	(continued on next page)

Table 1 (continued)

Study	Year	Proposed method(s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[56]	2020	CWGAN-ELQCP + GAN SVM/RF/DT/ MLP/KNN	Air Handling Units	×	SVM, RF, DT, MLP, KNN	ASHRAE RP-1312	real building data	simulation/ lab experiment data	real building data	Numerical Image		✓	✗	✗	Condenser Fouling, Non-condensables in Refrigerant Exhausted air damper stuck, Return fan at fixed speed, Cooling coil valve control unstable, Cooling coil valve partially closed, Outdoor air damper leak, AHU duct leaking
[57]	2020	AANN	Deep MLP	HVAC System	✓	PCA	Data of a 3-zone HVAC system simulated in TRNSYS		✓	✗	✓	✗	✓	✓	Faults of three zones temperature, chilled water tank temperature, chilled water tank temperature, AHU output air temperature, cooling coil return water temperature sensors
[58]	2021	MCNNs	Ensemble 1D CNN (multiscale)	Air Handling Units	×	MLP, PCA + SVM, SVM, WT-MLP, ELM	simulation data of an AHU experiment platform in a lab		✓	✗	✓	✗	✗	✗	duct air leakage, fan efficiency decrease, cooling coil valve stuck, and outdoor damper excess
[59]	2021	LSTM+GA	LSTM	Chiller	✗	RNN, MLP, SVM	simulation data of a 90 tons chiller (316 kW)		✓	✗	✓	✗	✓	✗	reduced condenser water flow, reduced evaporator water flow, refrigerant leakage, refrigerant overcharge, condenser fouling
[22]	2021	LSTMs	LSTM	Chiller	✗	LSTMs	ASHRAE RP-1043		✓	✗	✓	✗	✗	✗	Reduced condenser water

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method(s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[23]	2021	SP-CNN	2D CNN	Chiller	✗	PNN, SLENNet	ASHRAE RP-1043	✓	✗	✓	✗	✓	✗	✗	flow, Reduced evaporator water flow, Refrigerant, Excess oil, Refrigerant overcharge, Condenser fouling, non-condensable in refrigerant
[60]	2021	RACNNs	1D CNN	Air Handling Units	✗	MLP, SVM	dataset of an AHU fabricated in the laboratory of Nanyang Technological University	✓	✗	✓	✗	✓	✗	✗	Condenser fouling, Reduced condenser water flow, Non-condensables in refrigerant, Reduced evaporator water flow, Refrigerant leak/undercharge, Refrigerant overcharge, Excess oil
[61]	2021	CVAE (only faulty data) + LightGBM	CVAE	Air Handling Units	✓	CVAE (both normal and faulty data) + LightGBM, ROS + CVAE + LightGBM, RUS + LightGBM, ROS + LightGBM, SMOTE + LightGBM, ANS + LightGBM	ASHRAE RP-1312	✓	✗	✓	✓	✓	✓	✗	Exhaust air damper stuck, outdoor air damper stuck, cooling coil valve stuck, heating coil valve leaking, return fan at fixed speed
[26]	2021	FETCN	BDLSTM + LSTM + 1D CNN	Chiller	✓	FELSTM, FESVM, FEELM	ASHRAE RP-1043	✓	✗	✓	✗	✓	✗	✗	Condenser fouling, excess oil, refrigerant leak, refrigerant

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method (s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[62]	2021	1D CNN	1D CNN	Chiller	✗	1D CNN	ASHRAE RP-1043	✓	✗	✓	✗	✗	✓	✗	overcharge, reduced condenser water flow, reduced evaporator water flow, non-condensables gas in refrigerant Condenser fouling, Excess oil, Non-condensables in refrigerant, Reduced condenser water flow, Reduced evaporator water flow, Refrigerant leak, Refrigerant overcharge refrigerant overcharge/ undercharge, indoor heat exchanger fouling, outdoor heat exchanger fouling, and sensor fault
[63]	2021	1D CNN	1D CNN	Air Conditioning	✗	MLP, SVM, DT	dataset of a VRF system uses R410A as the refrigerant in a psychrometric testing room	✓	✗	✓	✗	✗	✗	✗	refrigerant overcharge/ undercharge, indoor heat exchanger fouling, outdoor heat exchanger fouling, and sensor fault
[19]	2021	Deep RNNs	LSTM	HVAC System	✗	RF, GB	Dataset of three types of building system located at the Lawrence Berkeley National Laboratory in Berkeley, California, FLEXLAB	✓	✗	✓	✗	✗	✗	✗	Outside air (OA) damper stuck, Valve of heating coil stuck, Valve of heating coil leaking, Valve of cooling coil stuck, Valve of cooling coil leaking, Outdoor air temperature bias, Unfaulted normal operation reduced condenser water flow, non-condensable gas in refrigerant, condenser fouling, reduced evaporator water
[64]	2021	Semi-GAN	GAN	Chiller	✗	MLP, DBN, LS-SVM, semi-SVM	ASHRAE RP-1043	✓	✗	✓	✗	✗	✗	✗	(continued on next page)

Table 1 (continued)

Study	Year	Proposed method(s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[55]	2021	CWGan-GP + GANomaly + RF/SVM/KNN/DT/NB/MLP	GAN	Chiller	✗	CWGan-GP + VAE/ensemble + RF/SVM/KNN/DT/NB/MLP	ASHRAE RP-1043	✓	✗	✓	✗	✓	✗	✗	flow, excess oil, refrigerant overcharge, refrigerant leak Reduced Condenser Water Flow, Reduced Evaporator Water Flow, Refrigerant Leak, Refrigerant Overcharge, Excess Oil, Condenser Fouling, Non-condensables in Refrigerant Refrigerant leakage, Refrigerant overcharge, and Reduced evaporator water flow
[65]	2021	2D CNNs with transfer learning	2D CNN	Chiller	✗	2D CNNs without transfer learning	Simulation data from two experimental water-cooled screw chillers with capacities of 422 kW (120 tons) and 703 kW (200 tons)	✓	✗	✓	✗	✓	✗	✗	Reduced evaporator water flow, Condenser fouling, Reduced condenser water flow, Non-condensables in refrigerant, Refrigerant leak/undercharge, Refrigerant overcharge, Excess oil
[24]	2021	Deep MLP + SA	Deep MLP	Chiller	✗	Deep MLP	ASHRAE RP-1043	✓	✗	✓	✗	✗	✗	✗	Compressor discharge pressure, Throttle device inlet pressure, Throttle device outlet pressure, Compressor suction temperature, Compressor
[66]	2022	BDLSTM -based deep recurrent canonical correlation analysis (BLCCA)	BDLSTM	Chiller	✗	PCA, CCA, DCCA	Dataset of an air-cooled chiller system installed in the Honeywell Home and Building Control Laboratory in Tianjin University	✓	✗	✓	✗	✓	✗	✗	(continued on next page)

Table 1 (continued)

Study	Year	Proposed method (s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method real building data	Tested using real simulation/lab experiment data	Input data format Numerical building data	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[25]	2022	S-DAGMM-DNN	Ensemble deep MLP	Chiller	✓	SAE-DNN, DNN, KNN, SVM, RF, AdaBoost	ASHRAE RP-1043	✓	✗	✓	✗	✓	✗	✗	discharge temperature, Air temperature of condenser outlet, Refrigerant temperature before throttling, Temperature of supply chilled water, Temperature of return chilled water sensor faults Reduced Condenser Water Flow, Reduced Evaporator Water Flow, Refrigerant Leak, Refrigerant Overcharge, Excess Oil, Condenser Fouling, Non-condensables in Refrigerant MZVAV: 10 faults of outdoor air damper, valves of heating and cooling coils. SZCAV: 14 faults of outdoor air damper, heating and cooling coil valves. SZVAV: 7 faults of outdoor air damper, heating, and cooling coil valves. ASHRAE RP-1312: exhaust air damper stuck, outdoor air damper stuck, cooling coil valve stuck, heating coil valve
[67]	2022	2D CNNs with transfer learning	2D CNN	Air Handling Units	✗	2D CNNs without transfer learning, MLP	ASHRAE PR-1312 + three datasets (MZVAV, SZCAV, SZVAV) from the automated fault detection and diagnosis project conducted by Lawrence Berkeley National Laboratory	✓	✗	✓	✗	✓	✗	✗	

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method(s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non-steady-state process	FDD of unknown/new fault	Fault types
[68]	2022	1D CNN + LSTM + Attention with dynamic cross-entropy and Swish activation	1D CNN + LSTM + Attention	Chiller	✓	1D CNN + LSTM + Attention with enhanced cross-entropy and Swish activation, 1D CNN + LSTM + Attention with soft cross-entropy and Swish activation, 1D CNN + LSTM + Attention with focal loss and Swish activation, 1D CNN + LSTM with soft cross-entropy and SeLU activation, 1D CNN + LSTM with soft cross-entropy and Swish activation, 1D CNN + LSTM with focal loss and Swish activation, 1D CNN + LSTM with enhanced cross-entropy and SeLU activation	ASHRAE RP-1043	✓	✗	✓	✗	✓	✗	✗	leaking, and fixed return fan speed faults of three severities. Reduced condenser water flow, reduced evaporator water flow, refrigerant leak, refrigerant overcharge, excess oil, condenser fouling, non-condensable in refrigerant
[69]	2022	MLP + Attention (modified Transformer) + 2D CNN	Deep MLP + Attention (modified Transformer) + 2D CNN	Heat Pump	✓	Convolution-sequence (C-S model) + 2D CNN, RNN + 2D CNN, MLP + Attention without initial state	ASHRAE RP-1043	✓	✗	✓	✗	✗	✗	Condenser fouling, reduced condenser water flow, reduced evaporator water flow, non-condensables in refrigerant, refrigerant leak,	

(continued on next page)

Table 1 (continued)

Study	Year	Proposed method (s)	Main building block(s)	System type	Encoder-decoder structure	Benchmarking methods	Dataset	Faults generation method	Tested using real building data	Input data format	Diagnosis of simultaneous faults	FDD of different severity levels	FDD of non steady-state process	FDD of unknown/new fault	Fault types
[70]	2022	Modified Transformer	Deep MLP + Attention (modified Transformer)	Air Handling Units	✓	recognition + 2D CNN GRU, classifier-chains integrated random forest	dataset of an on-site AHU	✓	✗	✓	✗	✓	✗	✗	refrigerant overcharge Fault 1: Air leak, Fault 2: Filter fouling, Fault 3: Outdoor air damper stuck, Fault 4: Return air damper stuck, Fault 5: Supply air damper stuck, Fault 6: Reduced chilled water, Fault 1 + Fault 6, Fault 2 + Fault 6, Fault 3 + Fault 4, Fault1 + Fault 4 + Fault 6, Fault 1 + Fault 5 + Fault 6, Fault 1 + Fault 3 + Fault 5 + Fault 6, Fault 2 + Fault 3 + Fault 5 + Fault 6, Fault 2 + Fault 3 + Fault 5 + Fault 6, Fault 5 + Fault 6

worse than that of a shallow MLP in the early years [28]. From the reviewed studies, two approaches are used mitigate these drawbacks. The first approach is using an encoder-decoder structured model with an unsupervised layer-wise training paradigm [49,57,47] and initialize the classifier with the pre-trained weights and fine-tune the whole model for FDD. Because training multiple layers from scratch in one shot is not involved in this approach, the vanishing gradient problem is alleviated. In addition, the distribution from unlabelled data can be learned from the pre-training process, which prevents the model from being stuck in the local optima. This approach is particularly useful if labelled training samples are scarce, while there are abundant unlabelled training samples, which aligns with the data availability of HVAC systems in practice. The second approach is using another optimization method such as simulation annealing algorithm to optimize the parameters [24]. Compared to the first approach, the second approach is less used, because it only mitigates the local minima problem, while other issues discussed above cannot be effectively addressed using the second approach.

In general, components within an HVAC system are strongly coupled, and the output of each component is time-dependant. However, temporal information cannot be effectively utilized by a vanilla deep MLP model by design. Therefore, other DL methods such as recurrent neural network (RNNs), which are more effective in time series modelling, are considered natural for HVAC FDD compared to a deep MLP model.

4.2. CNN based methods

4.2.1. 1D CNN based methods

Compared to MLPs, 1D CNNs show superior performance in extracting local temporal features from sequential data in terms of FDD [72]. Therefore, 1D CNNs can be either used as a feature extractor connecting to another downstream classifier for FDD [42] or used as an end-to-end classifier [60,62,63], which offers greater design flexibility. Although 2D CNN-based methods can also be adapted for sequence data analysis, extra data conversion steps are needed, which costs extra computation and involves potential information. A 1D CNN processes building operational data directly by design without extra data conversion steps. In addition, the number of trainable parameters of a 1D CNN is less than that of a 2D CNN in general due to their designed kernel shapes, allowing a 1D CNN to use larger kernel sizes and add more layers while maintaining the parameter size and computation cost. Such high efficiency enables a 1D CNN to be applied for online FDD or real-time condition monitoring of an HVAC system.

4.2.2. 2D CNN based methods

Recent studies show that utilizing multiple sensors with sensor fusion techniques leads to higher accuracy and more robust FDD results [73, 74]. A 2D CNN is designed to process multidimensional matrices, and such desired property enables it to utilize sensor fusion techniques to improve the FDD performance. Commonly, multiple sensors are installed at various locations of the HVAC systems, and the corresponding metre measurements can be consolidated together with the temporal information to form the input for 2D CNN. By doing so, the superior capabilities of image analysis and automatic feature extraction by a 2D CNN can be fully utilized.

Moreover, formulating the data for a 2D CNN unleashes the feasibility of applying transfer learning approaches for HVAC FDD, which is extremely useful to tackle the lack of labelled building operational data issue. In addition, the transfer learning approach can leverage the knowledge learned from one HVAC system to facilitate the FDD for HVAC systems with different system specifications [65,67], which is another considerable advantage compared to the conventional approaches that are system specific. On the other hand, except 2d CNN, other DL methods included in this review have not been applied for HVAC FDD using the transfer learning paradigm yet.

Methods used to formulate the input data for a 2D CNN include

converting numeric features of the original data into the corresponding images with the y-axis representing the time and the x-axis representing the features [13,16,39,40] or with the y-axis representing the features and the x-axis representing the time [53], placing each measured feature value row-wise and column-wise in a matrix at each timestep [51,23]. However, the data conversion process causes extra computational cost, and information embedded in the original 1D data can be distorted during the converting procedure.

Another scenario that makes a 2D CNN nature choice over other methods is when the collected data for analysis are images, which can be fed into a 2D CNN without data conversion. Although using image data directly eliminates the extra data conversion steps and potential information loss involved in this procedure, the model performance is influenced by the quality of the images depending on various factors such as the type of cameras, weather conditions, etc. [36].

4.3. RNN based methods

4.3.1. LSTM

RNNs are the most commonly used neural network (NN) architecture for time series data modelling in general. The feedback loops of the recurrent cells inherently capture the temporal patterns of the sequences [23]. Owing to this property, RNNs are effective in modelling highly time-dependant and strongly coupled events, and such characteristics are common within an HVAC system.

Similarly, due to the recurrent structures by design, RNN-based methods are more robust to the changes in system dynamics when performing FDD for a thermodynamics-changing HVAC system. From the literature, while other methods failed to detect the fault when the AHU system changes from occupied to unoccupied, RNN-based methods successfully detect the faults throughout the whole period regardless of such changes [11].

However, the main drawback of a conventional RNN is its limitation of modelling the long-term temporal patterns because of gradient exploding and vanishing [75]. Therefore, a vanilla RNN is barely used for modelling complex HVAC systems, while its variation, LSTM is more used in the most recent HVAC FDD literature. The design gating mechanisms and cell states enable the LSTM to perform better in modelling long-term temporal dependencies embedded inside the sequence data [76]. To improve the performance of a vanilla LSTM for HVAC FDD [22, 43–45], optimization methods such as generic algorithm is used in [59] to optimize learning rate, batch size, etc. As opposed to the optimization, sets of hyperparameters and network structure configurations are hand-crafted and evaluated in [19] to search for the optimal setting, it is also concluded from this study that adding more LSTM neurons or layers potentially causes low convergence speed and overfitting problems.

4.3.2. GRU

Gated recurrent unit (GRU) is another variant of RNN, and compared to LSTM, the structure of a GRU is relatively simple. Therefore, using a GRU is beneficial when the computational cost is a bottleneck. In terms of HVAC FDD, such characteristics of a GRU are highly favoured by an online HVAC health state monitoring system as potential faults can be detected earlier if the model is more efficient. In addition, from the maintenance and cost perspectives, fixing a fault at the less severe stages is much easier for a technician and cost-effective [77]. On the other hand, the less complex structure can potentially limit the capability of a GRU compared to that of an LSTM when modelling a complex HVAC system. Although the systematic evaluation of GRU versus LSTM in terms of HVAC FDD performance is not found in the literature, studies showed that GRU achieved the same performance or even better performance for industrial machinery and chemical processes FDD [78,79].

4.3.3. BDLSTM/BGGRU

Other useful variations of RNNs are bidirectional LSTM/GRU, useful information from another direction that a unidirectional LSTM/GRU

misses out can be captured using these variations. In terms of HVAC FDD, minor early-stage faults are difficult to detect due to the almost non-observable deviations without discriminative features. Such faults can be more easily identified if the deviation can also be analysed in reverse chronological order. Therefore, using a BDLSTM/GRU can lead to better performance for detecting minor drifting faults than using a conventional LSTM/GRU [80]. In addition, HVAC systems present coupling and time-varying dynamics in general. For instance, because the circulation of refrigeration is a closed loop, the output value of the component involved in the circle at the current time step is not only influenced by its previous state but also corresponds to the subsequent state. Such characteristics align with the bidirectional design and make BDLSTM/GRU useful for HVAC FDD [41,66], although the number of parameters of a bidirectional architecture is doubled compared to the conventional LSTM/GRU and thus results in higher computational cost.

4.4. Generative models

4.4.1. GAN and VAE

Most data-driven methods in the literature for HVAC FDD are supervised learning based [20]. However, faults are rare in real HVAC systems; therefore, most methods are tested using either simulation or experimental datasets, which can deviate from reality. Two types of DL-based generative models, i.e. VAE [128], GAN [129] and their variations [20,55,56,81,82], have been used to improve the feasibility of applying such methods for the real-world HVAC FDD. GANs have been widely used to generate synthetic images close to real-world samples for computer vision tasks [83]. However, the conventional GAN needs to be modified to be adapted to the HVAC data. For example, the up-sampling structure of the corresponding generator needs to be replaced by a down sampling structure corresponding to the dimension of the building operational data, as the feature dimension of HVAC data is much lower than that of image data in general. GANs can generate samples that fit the real distribution without explicitly assuming a certain form of a probability distribution. However, training a GAN is difficult due to the problems of mode collapse, non-convergence, and instability [84]. On the contrary, training a VAE is relatively easy as gradient descent algorithms can be directly applied to minimize the reconstruction loss and the Kulback-Leibler divergence loss [85]. Although from the literature, data generated by a VAE are less variant, especially for high dimensional data such as text and images [86]. This is not a problem for the data of HVAC systems in general. Because the dimensionality of building operational data is much lower than that of images and text data, VAEs can be more effective for HVAC faults generation than GANs.

However, none of the HVAC FDD approach based on GANs and VAEs is end-to-end, the output of the generative models needs to be fed to a downstream classifier for FDD, and therefore, the FDD result can be impacted negatively if the quality of generated data from the upstream generative model is poor.

4.4.2. Deep belief networks

DBN is a probabilistic generative model that stacks multiple layers of Restricted Boltzmann Machines (RBMs) [87]. In general, Deep learning methods require a large amount of labelled data for training, while the lack of labelled building operational data makes applications of DL methods for HVAC FDD using real building data challenging. GANs and VAEs address the data challenging by synthetic data generation, while DBM tackles this problem from a different angle, the training process of a DBN consists of unsupervised layer-by-layer pre-training and supervised fine-tuning stages. Therefore, the unlabelled data can be fully utilized by the unsupervised learning process of the DBN [88].

Moreover, the pre-trained weights can be used to initialize the following layers within the network, improve the model generalization ability, and reduce the risk of overfitting [89], which is critical to HVAC FDD given the limited amount of faulty data available from a real-life building system. Another advantage of DBN is that using the staged

training strategy, the gradient vanishing problem of training a deep structured neural network is alleviated, which enables the feasibility of adding more hidden layers to model complex nonlinearities of an HVAC system and extract representative features of faults from different abstraction levels [90]. In addition, as opposed to GANs and VAEs, a separate classifier is not need in the studies [9,48,50] using DBN approaches, which mitigates the non-end-to-end problem found in GANs and VAEs.

However, DBN has been less used in recent years, because the same problems can be solved by a stacked autoencoder, which shares the same advantages of unsupervised, layer-wise, multi-stage training, weights initialization using the pre-trained model, etc. with the DBN. In addition, training a DBN is difficult due to its complex data models [91], while a stacked AE can be trained easily using the ordinary back-propagation method. Therefore, it can be found from the literature that encoder-decoder-based approaches are more used by researchers for HVAC FDD.

4.5. DL based hybrid models

A hybrid DL approach refers to ensembling multiple DL models into one hybrid model or stacking different types of DL building blocks to form one hybrid structured model. Hybrid models are more robust as they often complement the advantages of the individual techniques involved and improve the overall performance [92].

The hybrid models used in the reviewed studies can be categorized into three types.

The first type involves combining 1D CNN with LSTM/GRU/BDLSTM [15,23,26], which utilizes both local and global temporal features extracted by 1D CNN and LSTM/GRU, respectively. Results of these studies show that this hybridizing strategy works well for diagnose gradual and minor faults such as fouling components at the early stage, which is generally challenging [22]. This hybridizing approach can be further enhanced by attention algorithms in terms of the feature extraction process. A fused attention mechanism that consisted of self-attention (SA) [93] and external-attention (EA) [94] is used in [68]. The SA was used to weigh the features based on their importance, while EA was used to discover the correlation between different features. As a result, SA enabled the model to focus more on the relatively important features, while EA allowed the model to learn the most discriminative features across the whole dataset. Results show that the proposed method worked well for imbalanced dataset with a small amount of faulty data, which addresses the data scarcity issue in this problem domain.

The second type involves combining a deep AE with 2D CNN [95], it takes advantage of using deep AE to generate high quality multi-dimensional residual that is the thermodynamic deviation and critical feature for HVAC FDD, and by reshaping the residual data into matrix, it can be effectively processed by 2D CNN. The third type ensembles multiple DL building blocks, e.g., ensemble deep MLPs [25] and 1D CNNs [52]. Because an ensemble model is more robust to variance, it can be used to remove noise data [25] before FDD. More importantly, although a vanilla 1D CNN is effective in modelling the fast signals in the short fixed-length segments, its performance degrades when extracting discriminative features of multiscale monitoring signals, which is common in HCAC systems. For example, within an AHU system, signals such as temperature and humidity change slowly, while airflow rate fluctuates rapidly.

The third type of hybrid approach mitigates this issue by using multiple kernels of different sizes to capture the temporal features of different time scales, i.e., a larger sized kernel can be used to capture the overall tendency of slow-changing signals, while a smaller kernel can be applied to capture the sudden changes of rapid-change signals [93]. Due to the efficiency of 1D CNN, the computation cost of such an approach is still reasonable. Based on the discussions above, advantages and disadvantages of each method briefly are summarized in Table 2.

Table 2
Advantages and disadvantages of each method.

DL methods	Advantages	Disadvantages
MLP	simple design	states of time steps are not preserved, limited ability of modelling the time-dependence of an HVAC system
1D CNN	high efficiency, suitable for online FDD or real-time condition monitoring of an HVAC system	local temporal feature extractor, limited for modelling the long-term temporal dependency comparing to LSTM based methods
2D CNN	nature choice for image data, feasibility for transfer learning, effective for processing multidimensional metrics that enables the use of sensor fusion techniques to improve the FDD performance	data conversion steps for non-image data that potentially causes information loss and extra computational cost, quality of collected image data heavily relies on type of cameras, weather conditions, etc., large amount of training parameters
LSTM	effective for modelling thermodynamics-changing characteristic and long-term temporal patterns of the system	complex gating mechanisms, potential risk of low convergence speed and overfitting
GRU	less complex structure than LSTM, less computational cost, favoured by an online HVAC health state monitoring system	the simplified structure can potentially limit the capability of a GRU when modelling a complex HVAC system
BDSM/BDGRU	capable of modelling temporal dependencies from both forward and reverse directions, effective for detecting minor drifting faults	complex structures and more parameters
GAN	generate samples that fit the real distribution without explicitly assuming a certain form of a probability distribution, useful for faults generation and mitigating the data scarcity problem of FDD	difficult to train due to mode collapse, non-convergence, and instability
VAE	fast and easy to train comparing to GAN	data generated by a VAE are less variant
DBN	works for a small amount of faults data	difficult due to its complex data models
DL based hybrid models	more robust and complement the advantages of individual models involved	increased complexity in design, structure, more computational cost, not ideal for real-time FDD

5. Meta analysis

To answer the question whether DL methods are effective for HVAC FDD quantitatively, meta analysis is conducted and the results are discussed below.

5.1. Effectiveness analysis of fault detection

In terms of the quantitative results of fault detection, Fig. 3 shows the pooled sensitivity of DL for HVAC fault detection. From Fig. 3, the pooled result shows that there are 51,003 fault cases in total, amongst which 48,221 fault cases are correctly identified. The sensitivity values range from 0.855 to 0.996, the lowest value reported by Tra. On the other hand, the highest sensitivity values are reported by Wang and Han; although the upper bound value of 95% confidence interval (CI) is the same in both studies, the 95% CI of Wang's study is smaller due to the larger sample size used in this study, which indicates Wang's results are relatively reliable. The pooled sensitivity was 0.985, which means 98.5% of the fault cases detected by the DL methods are true faults. The 95% confidence interval is 0.958–0.995, suggesting the accuracy of a fault case classification is between 95.8% and 99.5%.

The pooled specificity of DL for HVAC fault detection is depicted in Fig. 4. The number of true negative cases is 6620, and the total number of negative cases is 7449. The specificity values range from 0.697 to 0.997. The lowest specificity value is presented by Tra, while the highest specificity is reported in Wang and Li's studies, while the 95% CI of Wang's study is smaller due to the large sample size and therefore suggests a more reliable result compared to Li's study. The pooled specificity is 0.966, meaning 96.6% of the normal cases classified by the DL methods are fault-free. The pooled 95% CI is 0.894–0.990, indicating the accuracy of a normal case classification is between 89.4% and 99.0%.

It is also observed from the forest plots that Tra's results deviate from the results of the rest five studies. The main reason is that Tra's study uses only 1% of labelled training data. Tra's experiment design aims to simulate the real-world faulty data scarcity scenario. However, it can cause the problem that the discriminative fault information within these small, labelled data is either not representative or insufficient to train the model as DL models require a large amount of training data to achieve good performance in general.

The conclusion of a single study may fail to be generalized to other studies because the DL methods, sample sizes, and objectives are adopted in different studies, while meta-analysis combines, summarizes, and interprets the results of the primary studies to derive the unknown generalized conclusion to a certain extent. Results show both pooled sensitivity and specificity values are greater than 95%, which suggests using DL methods can lead to high sensitivity and specificity for HVAC fault detection in general.

Fig. 5 shows the SROC curve of HVAC fault detection using DL methods with the pooled log DOR estimated. The summary estimate represents the pooled result of sensitivity and specificity from the individual primary studies, and the surrounding dashed grey line represents its corresponding 95%CI. The pooled log DOR is 7.578 with a 5.933–9.222 95% CI, which means HVAC systems with faults are approximately 7.6 (after log transformation) times more likely to be detected by the DL methods than those without any faults. A higher DOR value suggests the discrimination ability of the DL methods is better in

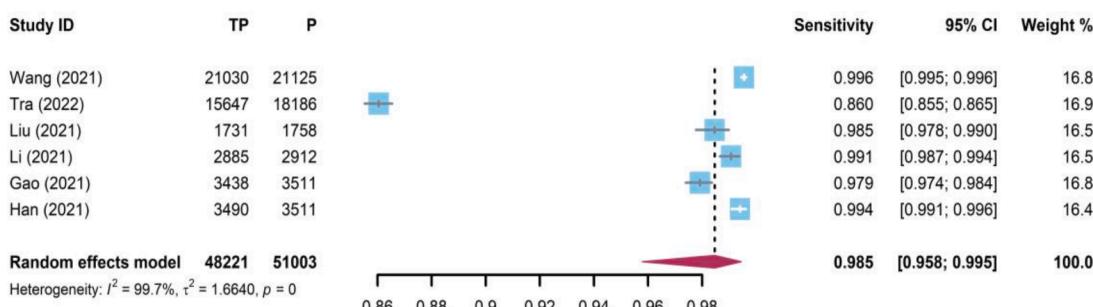


Fig. 3. Pooled Sensitivity of DL for HVAC fault detection.

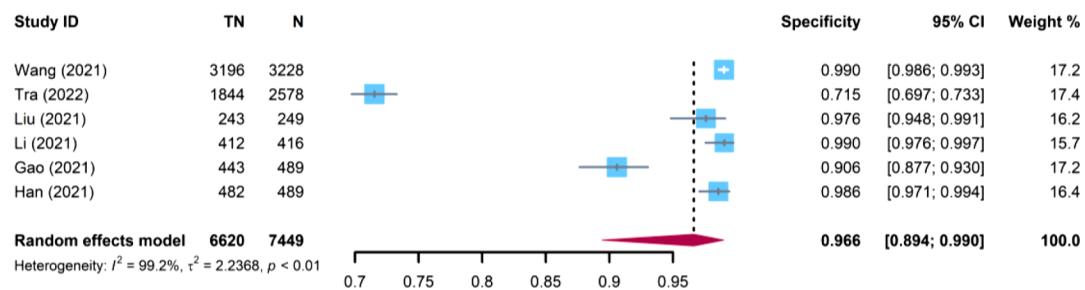


Fig. 4. Pooled Specificity of DL for HVAC fault detection.

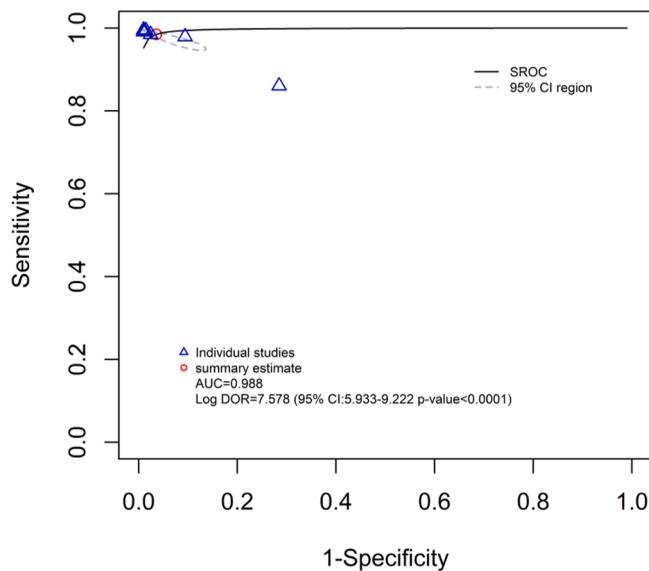


Fig. 5. SROC curve and the pooled log DOR of DL for HVAC fault detection.

terms of HVAC fault detection. The p-value was lower than 0.0001, a p value is the evidence against a null hypothesis, the null hypothesis in this context is HVAC FDD using DL methods is not effective. A p value less than 0.05 (typically ≤ 0.05) is statistically significant. It indicates strong evidence against the null hypothesis. In addition, the area under the curve (AUC) equals 0.988. AUC is the measure of the ability of a classifier to distinguish between classes, and an AUC value greater than 0.9 suggests the discrimination ability of DL methods is outstanding [96].

To evaluate the validity and robustness of the pooled summary, the sensitivity analysis is performed by iteratively removing one study at a time and performing the pooling using the remaining studies. The sensitivity analysis result is shown in Table 3. the log DOR value ranges from 7.313 to 8.409 under the influence of excluding one primary study, amongst which the most influential study is Tra's, this result is aligned with the observations from Figs. 3, 4, and the cause is discussed earlier.

Another problem that potentially distorts the estimated effect under investigation is publication bias, which refers to an editorial predilection

for publishing particular findings, e.g., positive results, which leads to authors' failure to submit negative findings for publication. In addition, large studies are more likely to be published because authors may be strongly tempted to dredge through the data from an essentially negative study to find positive results and publish only those [97]. The funnel plot is shown in Fig. 6(a) to investigate the publication bias. Each dot represents an individual study, the dashed middle line represents the overall effect, and the two dashed lines on the left and right represent the corresponding 95% confidence intervals. Ideally, the included studies should scatter symmetrically around the overall effect line. However, an asymmetrical pattern is observed in Fig. 6(a), especially for Tra's study, which is deviated from the estimated true effect. The result suggests the presence of publication bias. The reason is similar to that was discussed, i.e., the authors of this study deliberately reduced the number of labelled data for training.

To further confirm the assumption, a basic outlier detection for meta-analysis is performed using the brute force approach proposed by Harrer et al. [98]. Studies are defined as outliers if their 95% confidence interval lies outside the 95% confidence interval of the pooled effect. Tra's study is confirmed as an outlier study according to this method. To investigate the influence of the outlier study on the publication bias, Tra's study is removed, and the results of the remaining five studies are used to generate the funnel plot. As shown in Fig. 6(b), the funnel plot's observed pattern is more symmetrical than that in Fig. 6(a).

5.2. Effectiveness analysis of HVAC fault diagnosis

Since various types of faults exist in the systems, it is challenging to enumerate the results of each fault quantitatively. According to the survey [99], condenser fouling is ranked amongst the faults of chillers as the most critical fault in terms of repairing cost and occurrence frequency. Therefore, condenser fouling fault diagnosis results are selected for analysis, and the same meta-analysis process is used. Figs. 7 and 8 show the pooled sensitivity and specificity of diagnosing condenser fouling fault, respectively. From Fig. 7, there are 7304 samples of condenser fouling fault, amongst which 7234 are correctly diagnosed. The pooled sensitivity value is 0.995 with a 95% CI of 0.983–0.998. The lowest sensitivity value is 0.976, though it is still very close to 1. Two studies deviate from the rest, the reason for Tra's study is discussed in Section 5.1, while the sample size of Liu's study is the smallest amongst all included studies and thus results in a relatively wide CI. Apart from the small sample size, another reason is that all included studies reported high sensitivity values and the lower bounds of CI are high, therefore, the CI width of each study is very narrow. Therefore, although the CI width of Liu's study is just 0.043, it is still considered relatively wide compared to the rest of the studies with extremely narrow CI ranges.

In terms of the pooled specificity, there are 51,148 samples of non-condenser fouling fault, and 47,607 of them are correctly classified as non-condenser fouling fault free. The pooled specificity value is 0.981 with a 95% CI of 0.946 – 0.993. The lowest reported specificity is 0.823 from Tra's study, the specificity values reported in the rest studies reported are all greater than 0.96 with a narrow CI.

Table 3
Sensitivity analysis of Log DOR for HVAC fault detection using DL.

Study excluded	Log DOR
Wang (2021)	7.407
Tra (2022)	8.409
Liu (2021)	7.446
Li (2021)	7.381
Gao (2021)	7.519
Han (2021)	7.313

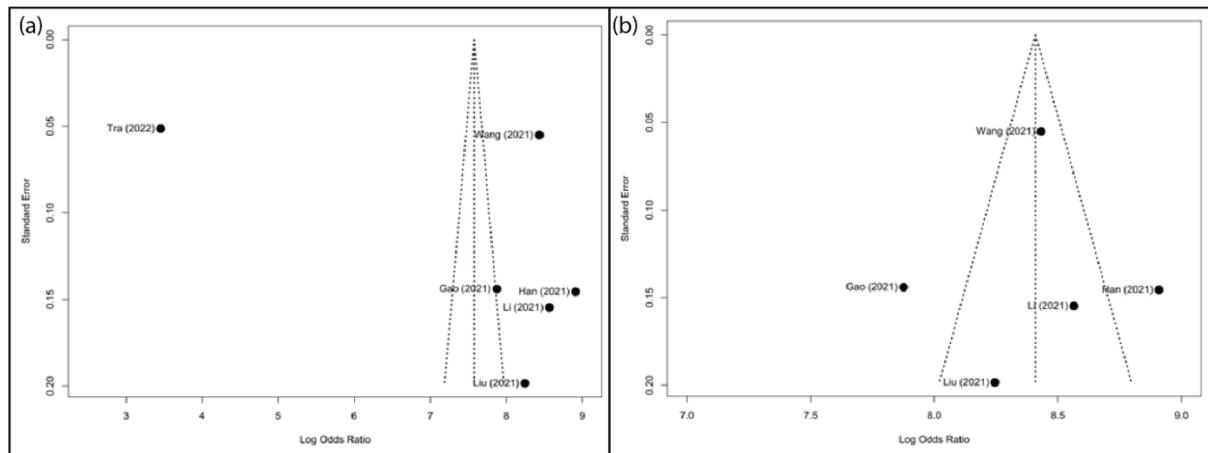


Fig. 6. . (a) The funnel plot of study results, (b) The funnel plot of study results without the outlier study.

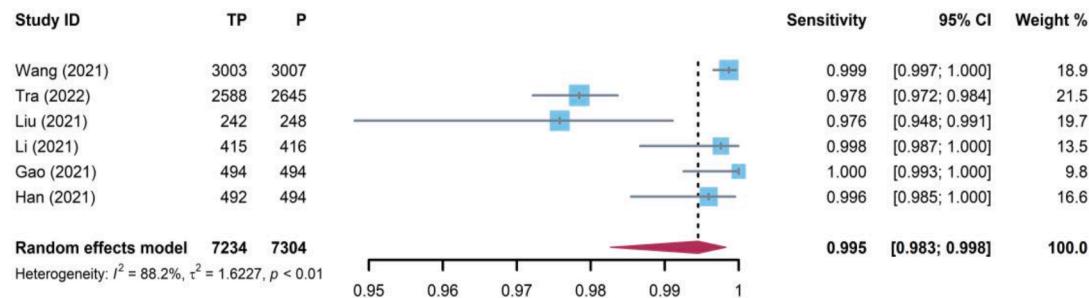


Fig. 7. Pooled Sensitivity of DL for HVAC fault diagnosis.

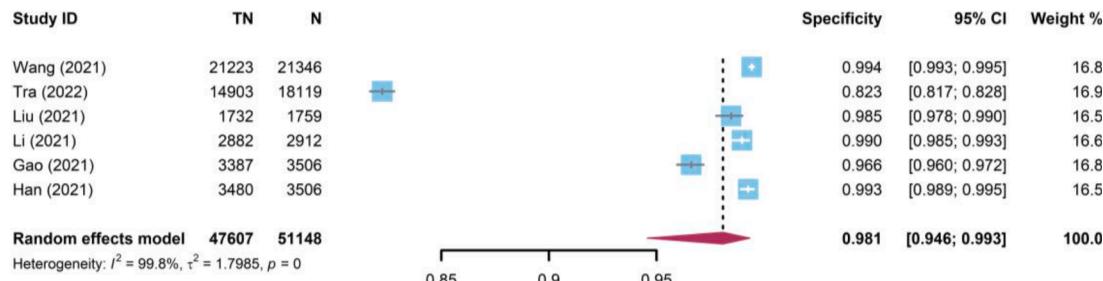


Fig. 8. Pooled Specificity of DL for HVAC fault diagnosis.

From Fig. 9, the AUC is greater than 0.9, which suggests that DL methods discriminate condenser fouling faults from normal cases effectively. Besides, the pooled log DOR is 7.639 with a 95% CI of 6.140 – 9.138, and the p-value was lower than 0.0001, which means the condenser fouling faults are approximately 7.7 (after log transformation) times more likely to be diagnosed by the DL methods than those without condenser fouling faults.

The sensitivity analysis result using the leave one out approach is shown in Table 4 tab. Due to the same reason, Tra's study is the most influential study; by excluding Tra's study the pooled log DOR increases to 8.377 from 7.639.

Publication bias in terms of condenser fouling fault diagnosis is investigated. Meanwhile, Tra's study is detected as an outlier using Harrer's method. Fig. 10(a,b) shows the funnel plots before and after removing Tra's study.

The distribution of individual studies changes after removing Tra's study, Liu and Li's studies are moved into the ideal funnel, while Gao's study deviates from the lower bound of the pooled 95% CI represented

by the left side dash line. Because the number of included studies is small, changing one influential study can result in a relatively significant change in the overall distribution. In addition, the results show asymmetrical patterns in both funnel plots, either due to publication bias or the small number of included studies.

6. The current status and knowledge gaps

The distribution of different DL methods for HVAC FDD is shown in Fig. 11. The most commonly used DL methods are 2D CNN and hybrid DL models, with 11 and 9 studies that use 2D CNN and hybrid DL approaches, respectively. It shows that researchers have favoured 2D CNN, LSTM, 1D CNN, and hybrid DL methods in the past three years. More diverse data formats such as images are collected and explored by researchers, which enables the application of 2D CNNs, and unleashes the potential of transfer learning to solve the data scarcity. Another main reason for transfer learning being favored by researchers is that it enables knowledge learned from one system to be generalized to other

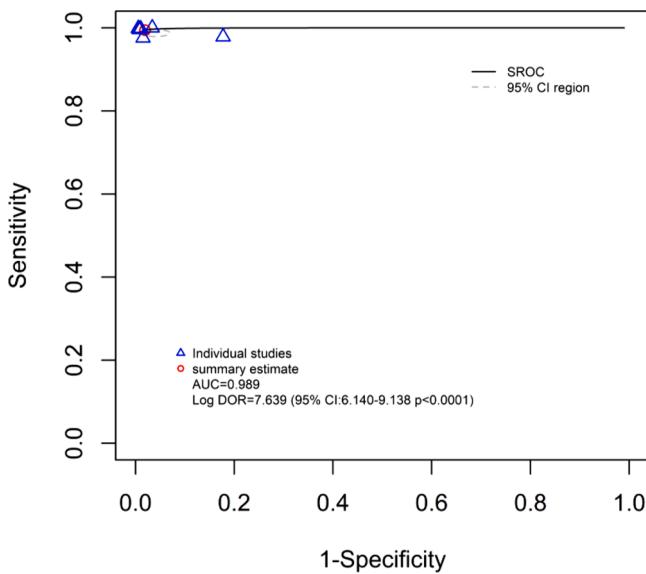


Fig. 9. SROC curve and the pooled log DOR of DL for HVAC fault diagnosis.

Table 4
Sensitivity analysis of Log DOR for condenser fouling fault diagnosis using DL.

Study excluded	Log DOR
Wang (2021)	7.388
Tra (2022)	8.377
Liu (2021)	7.511
Li (2021)	7.461
Gao (2021)	7.679
Han (2021)	7.411

systems with different operating characteristics. The improved generalization ability is a huge advantage over other DL methods, which are typically customized for a specific building energy system, and hence are limited by insufficient extrapolation capabilities [8].

On the other hand, LSTM and 1D CNN are natural choices for sequential HVAC data analysis without the need for data format conversion. While FDD efficiency is critical, 1D CNN can be a more suitable option. The reason hybrid DL methods are popular is that they effectively model complex systems that cannot be solved by a single DL model, e.g., using an ensemble multi-scale DL model to capture temporal features with different time scales of an HVAC system.

Although the reviewed studies show promising results, most are not tested against a real-world scenario using an imbalanced dataset. The

overall imbalance ratio (IR) distributions of training, validation, and testing datasets of the reviewed studies are shown in Fig. 12. The imbalanced ratio is calculated using Eq. (4).

$$IR = N_{maj}/N_{min} \quad (4)$$

Where N_{maj} is the sample size of the majority class and N_{min} is the sample size of the minority class. The dataset is balanced if the IR equals one, while it is an imbalanced dataset if the IR is greater than one. From Fig. 12, most studies do not specify the proportion of faults and normal conditions. To be more specific, one study uses an imbalanced training dataset and three use both imbalanced and balanced datasets. Regarding the testing data, two studies evaluate both imbalanced and balanced datasets, while the validation datasets used in the reviewed studies are either balanced or without IR being reported.

Therefore, the satisfying results reported in the studies remain to be validated using real-world data. Fig. 13 shows the learning paradigms of DL methods adopted by researchers. Although supervised learning is the most used learning paradigm in the past five years, other learning paradigms have been actively explored by researchers. Because of the data scarcity problem, the vanilla supervised learning paradigm is difficult to be applied to real-world HVAC FDD applications in practice. Transfer learning and semi-supervised learning are two types of new learning paradigms in 2021 and 2022 that have shown promising results for solving the data scarcity problem. However, there are only one semi-supervised learning and two transfer learning applications so far; thus, research concerning these two learning paradigms remains a research gap.

Apart from the learning paradigms, generative models can be another research direction to solve the same data scarcity problem. Amongst the 47 reviewed studies, generative models are used in only six studies. Therefore, there is a research space for generative models to improve the feasibility of applying DL methods for real-world HVAC systems FDD. Especially for VAE-based methods, compared to GANs, they are easier to train. Besides, since the dimensionality of HVAC operational data is relatively low comparing to image data, the quality of the generated results does not differ significantly. However, GANs are used in five studies, while only one study utilizes a VAE-based method for HVAC operational data generation.

Method-wise, other knowledge gaps identified from reviewed studies include FDD of simultaneous faults and FDD of unknown/new faults. Only six studies are tested with FDD of simultaneous faults and only one study investigates FDD of unknown/new faults. In addition, FDD of non-steady processes is another aspect that has not been well tested. Taking the reverse cycle defrosting process of a VRV system as an example, when the initiation of defrosting control is started, the compressor speed decreases to prepare the change in direction of the four-way valve. Suction and discharge pressure change rapidly in reverse. As a result, the discharge pressure diminishes, and the suction pressure increases

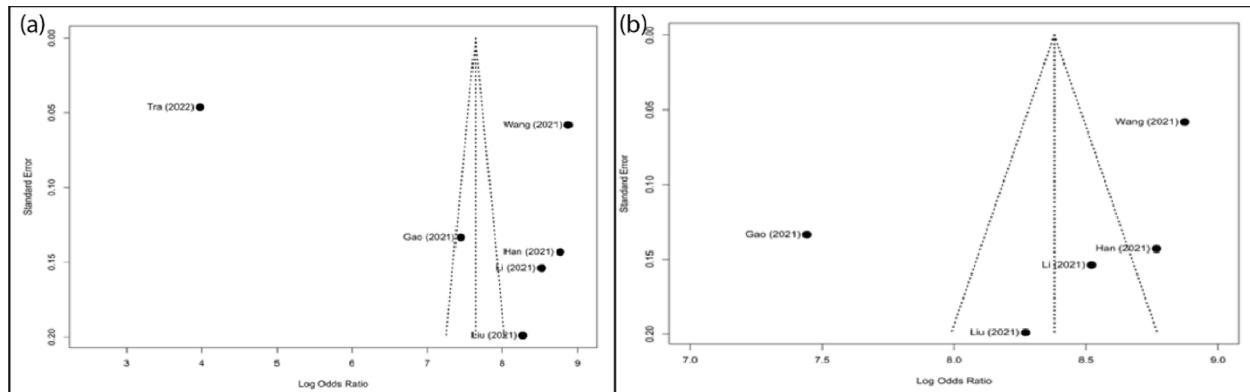


Fig. 10. (a) The funnel plot of study results (b) The funnel plot of study results without the outlier study.

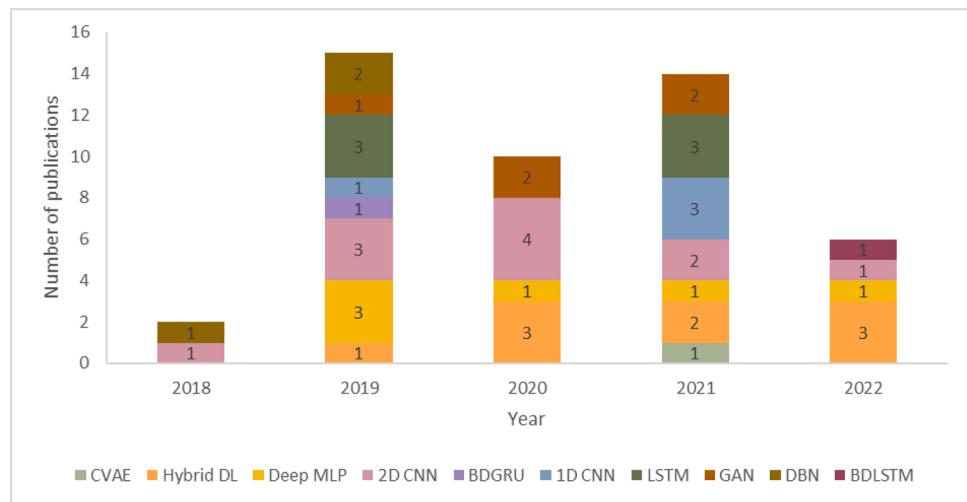


Fig. 11. Summary of FDD applications using different DL methods.

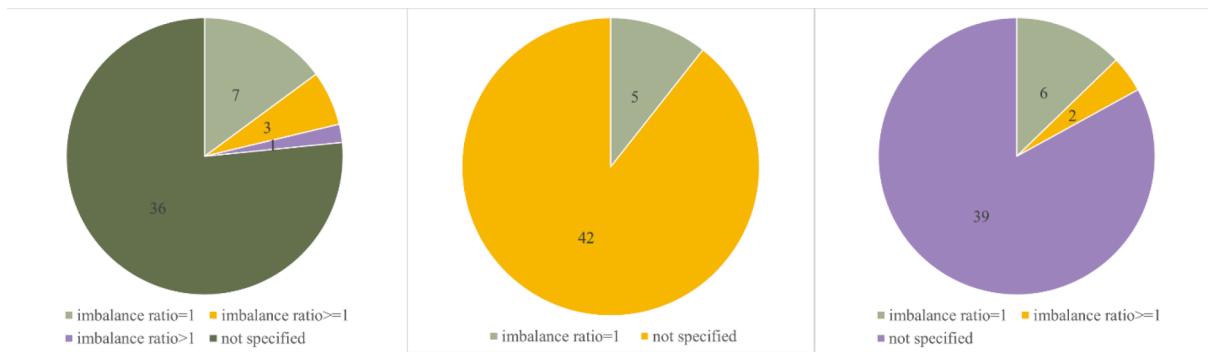


Fig. 12. Distribution of data imbalance ratio (training, validation, and testing data from left to right).

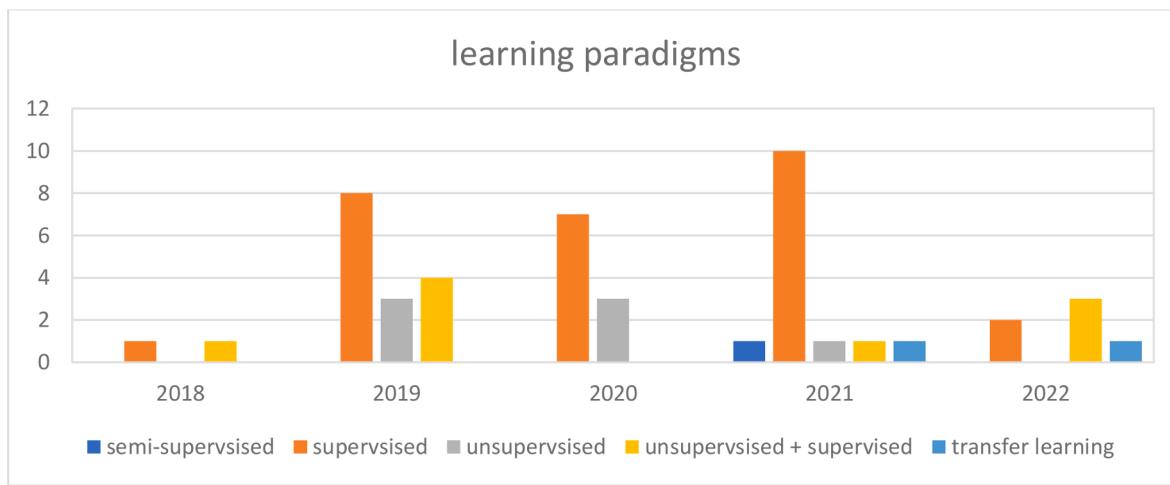


Fig. 13. Summary of learning paradigms.

sharply to achieve a balance. The series of drastic changes involved in such a transient can cause various faults, such as valve leakage, valve stuck, etc. FDD of unsteady processes are tested in only four studies.

These characteristics are commonly found in real-world HVAC systems that increase the complexity of FDD and have not been thoroughly studied yet.

Data-wise, the lack of public benchmarking datasets of real building systems limits the development of DL approaches for real-world HVACs

FDD. Fault data used in 44 out of 47 reviewed studies are either lab experiments or simulated data, including the two most used public datasets from ASHRAE. However, real-world complexities are generally not fully covered in the lab experiment or simulated data, e.g., neither simultaneous faults nor unknown/new faults scenarios are incorporated by ASHRAE datasets. Composing a public dataset covering real-world complexities can be highly beneficial for future researchers in this area to develop the approaches that can be applied for real building

systems FDD. Last but not least, current studies utilized only two types of data format. The 38 studies use numerical data input, and nine studies analyse image data for FDD. However, with the development of DL methods, state-of-the-art results have been achieved by DL methods for various multimodality data analysis tasks. Therefore, collecting and exploring multiple forms of data, such as documented issue logs, recorded/monitored audio, and video clips of the HVAC systems, can be beneficial and considered in future studies.

7. Conclusions

This study performs a systematic review with meta analysis of DL methods for HVAC FDD. 47 studies are included in the qualitative analysis, and six studies are included in the quantitative analysis. In terms of the fault detection, the pooled sensitivity value is 0.985 (95% CI: 0.958 – 0.995), The pooled specificity value is 0.966 (95% CI: 0.894 – 0.990), and the pooled log DOR is 7.578 (95% CI: 5.933 – 9.222), and the pooled AUC is 0.988. Regarding fault diagnosis, one of the most critical faults from the literature, i.e., condenser fouling fault, is selected for meta analysis. The pooled sensitivity of diagnosing is 0.995 (95% CI: 0.983 – 0.998), the pooled specificity value is 0.981 (95% CI: 0.946 – 0.993), the pooled log DOR is 7.693 (95% CI: 6.140–9.138), and the pooled AUC is 0.989. The results suggest DL methods can be effectively used for HVAC FDD. Although, the published results are satisfying, the existence of potential publication bias indicates that there exist studies

without significant results that are not published.

It is also concluded that the most used DL method is 2D CNN due to the development of transfer learning and data format diversity. In addition, most of the existing studies use simulation/lab experiment data due to the challenge of collecting fault data of real building systems. Challenges in real-word HVAC systems including FDD of non-steady processes, simultaneous faults, and unknown/new faults are research areas to be explored in the future.

Data availability

No data was used for the research described in the article.

Acknowledgments

We thank Dr. Chris Bales' kind help in providing domain expert knowledge and manuscript revising comments and Dr. Johan Håkansson's help in giving manuscript revising comments. In addition, we acknowledge that Dr. Fan Zhang and Dr. Paria Sadeghian made equal contributions to the corresponding author role of this study. Dr. Fan Zhang managed the conceptualization, methodology, and entire research life cycle of this study, while Dr. Paria Sadeghian managed the entire manuscript submission and publication process and coordinated the communications between the journal and the authors.

Appendix A

Nomenclature for common algorithms	
Conditional Wasserstein generative adversarial network with gradient penalty	CWGAN-GP
Conditional Wasserstein generative adversarial network	CWGAN
Auto-Associative Neural Network	AANN
Supervised multiclass deep autoencoding Gaussian mixture model	S-DAGMM
Stacked autoencoder	SAE
Deep bidirectional long short-term memory recurrent neural network	BDLSTM
Conditional Wasserstein generative adversarial network with ensemble learning quality control protocol	CWGAN-ELQCP
Multiscale convolutional neural networks	MCNNs
Rule and convolutional neural networks	RACNNs
Shallow and deep CNNs	SDCNNs
Data-temporal attention network	DAN
Data self-production and deep convolutional neural network	SP-CNN
Feature-enhanced temporal convolutional network	FETCN
Support vector data description	SVDD
Linear Support Vector Machine	L-SVM
Feature-enhanced SVM	FESVM
Feature-enhanced LSTM	FELSTM
Feature-enhanced ELM	FEELM
Linear Discriminant Analysis	LDA
Gaussian Naïve Bayes	GNB
Adaboost	AB
Probabilistic neural network	PNN
Sparsely local embedding deep convolutional neural	SLENNet
Encoder-decoder network	EDN
Encoder-decoder network with a fusion module	EDN-F
Temporal attention network	TAN
Auxiliary neural network	AUX-NN
Combined neural network	CB-NN
Conditional Wasserstein generative adversarial network with gradient penalty	CWGAN-GP

References

- [1] González-Torres M, Pérez-Lombard L, Coronel JF, Maestre IR, Yan D. A review on buildings energy information: trends, end-uses, fuels and drivers. Energy Rep Nov. 2022;8:626–37. <https://doi.org/10.1016/j.egyr.2021.11.280>.
- [2] Madani H, Roccatello E. A comprehensive study on the important faults in heat pump system during the warranty period. Int J Refrig 2014;48:19–25. <https://doi.org/10.1016/j.ijrefrig.2014.08.007>.
- [3] Ginestet S, Marchio D, Morisot O. Evaluation of faults impacts on energy consumption and indoor air quality on an air handling unit. Energy Build 2008;40(1):51–7. <https://doi.org/10.1016/j.enbuild.2007.01.012>.
- [4] Borgstein EH, Lamberts R, Hensen JLM. Mapping failures in energy and environmental performance of buildings. Energy Build 2018;158:476–85. <https://doi.org/10.1016/j.enbuild.2017.10.038>.
- [5] Li Y, O'Neill Z. A critical review of fault modeling of HVAC systems in buildings. Build Simul Oct. 2018;11(5):953–75. <https://doi.org/10.1007/s12273-018-0458-4>.

- [6] Kim W, Katipamula S. A review of fault detection and diagnostics methods for building systems A review of fault detection and diagnostics methods for building systems. *Sci Technol Built Environ* 2018;24(1):3–21. <https://doi.org/10.1080/23744731.2017.1318008>.
- [7] Shi Z, O'Brien W. Development and implementation of automated fault detection and diagnostics for building systems: a review. *Autom Constr* 2019;104:215–29. <https://doi.org/10.1016/j.autcon.2019.04.002>.
- [8] Zhao Y, Li T, Zhang X, Zhang C. Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future. *Renew Sustain Energy Rev* 2019;109(November 2018):85–101. <https://doi.org/10.1016/j.rser.2019.04.021>.
- [9] Guo Y, et al. Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving. *Appl Energy* 2018;225:732–45. <https://doi.org/10.1016/j.apenergy.2018.05.075>.
- [10] Mirnaghhi MS, Haghhighat F. Fault detection and diagnosis of large-scale HVAC systems in buildings using data-driven methods: a comprehensive review. *Energy Build* 2020;229:110492. <https://doi.org/10.1016/j.enbuild.2020.110492>.
- [11] Shahnazari H, Mhaskar P, House JM, Salsbury TI. Modeling and fault diagnosis design for HVAC systems using recurrent neural networks. *Comput Chem Eng* Jul. 2019;126:189–203. <https://doi.org/10.1016/j.compchemeng.2019.04.011>.
- [12] Li G, Hu Y, Liu J, Fang XL, Kang J. Review on fault detection and diagnosis feature engineering in building heating, ventilation, air conditioning and refrigeration systems. *IEEE Access* 2021;9:2153–87. <https://doi.org/10.1109/ACCESS.2020.3040980>.
- [13] Miyata S, Lim J, Akashi Y, Kuwahara Y, Tanaka K. Fault detection and diagnosis for heat source system using convolutional neural network with imaged faulty behavior data. *Sci Technol Built Environ* Jan. 2020;26(1):52–60. <https://doi.org/10.1080/23744731.2019.1651619>.
- [14] Lee K-PP, Wu B-HH, Peng S-LL. Deep-learning-based fault detection and diagnosis of air-handling units. *Build Environ* Jun. 2019;157(February):24–33. <https://doi.org/10.1016/j.enbuild.2019.04.029>.
- [15] Sun Z, Jin H, Gu J, Huang Y, Wang X, Shen X. Gradual fault early stage diagnosis for air source heat pump system using deep learning techniques. *Int J Refrig* Nov. 2019;107:63–72. <https://doi.org/10.1016/j.ijrefrig.2019.07.020>.
- [16] Eom YH, Yoo JW, Hong SBin, Kim MS. Refrigerant charge fault detection method of air source heat pump system using convolutional neural network for energy saving. *Energy* 2019;187:115877. <https://doi.org/10.1016/j.energy.2019.115877>.
- [17] Buffa S, et al. Advanced control and fault detection strategies for district heating and cooling systems—A review. *Appl Sci* 2021;11(1):1–32. <https://doi.org/10.3390/app11010455>.
- [18] Sun Z, et al. Studies on the online intelligent diagnosis method of undercharging sub-health air source heat pump water heater. *Appl Therm Eng* Mar. 2020;169:114957. <https://doi.org/10.1016/j.applthermaleng.2020.114957>.
- [19] Taheri S, Ahmadi A, Mohammadi-Ivatloo B, Asadi S. Fault detection diagnostic for HVAC systems via deep learning algorithms. *Energy Build* Nov. 2021;250:111275. <https://doi.org/10.1016/j.enbuild.2021.111275>.
- [20] Yan K, Chong A, Mo Y. Generative adversarial network for fault detection diagnosis of chillers. *Build Environ* Apr. 2020;172(January):106698. <https://doi.org/10.1016/j.enbuild.2020.106698>.
- [21] Moher D, Liberati A, Tetzlaff J, Altman DG, Group TP. Preferred Reporting Items for Systematic Reviews and Meta-Analyses : the PRISMA Statement. *Public Libr Sci* Aug. 2009;6(4):264–9. <https://doi.org/10.1371/journal.pmed.1000097>.
- [22] Liu Y, Wang X, Fan C, Huang B, Wang J. A deep recurrent neural network-based method for automated building system fault diagnosis. In: Proceedings of the 25th International Symposium on Advancement of Construction Management and Real Estate; 2021. p. 613–24. <https://doi.org/10.1007/978-981-16-3587-8>.
- [23] Gao J, Han H, Ren Z, Fan Y. Fault diagnosis for building chillers based on data self-production and deep convolutional neural network. *J Build Eng* 2021;34:102043. <https://doi.org/10.1016/j.jobe.2020.102043>.
- [24] Han H, Xu L, Cui X, Fan Y. Novel chiller fault diagnosis using deep neural network (DNN) with simulated annealing (SA). *Int J Refrig* 2021;121:269–78. <https://doi.org/10.1016/j.ijrefrig.2020.10.023>.
- [25] Tra V, Amayri M, Bouguila N. Outlier detection via multiclass deep autoencoding Gaussian mixture model for building chiller diagnosis. *Energy Build* Mar. 2022;259:111893. <https://doi.org/10.1016/j.enbuild.2022.111893>.
- [26] Li C, Shen C, Zhang H, Sun H, Meng S. A novel temporal convolutional network via enhancing feature extraction for the chiller fault diagnosis. *J Build Eng* 2021;42:103014. <https://doi.org/10.1016/j.jobe.2021.103014>.
- [27] Wang Z, Dong Y, Liu W, Ma Z. A novel fault diagnosis approach for chillers based on 1-D convolutional neural network and gated recurrent unit. *Sensors* Apr. 2020;20(9):2458. <https://doi.org/10.3390/s20092458>.
- [28] Feng S, Zhou H, Dong H. Using deep neural network with small dataset to predict material defects. *Mater Des* 2019;162:300–10. <https://doi.org/10.1016/j.matdes.2018.11.060>.
- [29] Xiao W, Huang J, Wang B, Ji H. A systematic review of artificial intelligence in the detection of cutting tool breakage in machining operations. *Meas J Int Meas Confed* 2022;190:110748. <https://doi.org/10.1016/j.measurement.2022.110748>.
- [30] Bossuyt P SR, Davenport C, Deeks J, Hyde C, Leeflang M. Chapter 11: interpreting results and drawing conclusions. In: *Cochrane handbook for systematic reviews of diagnostic test accuracy version 0.9*. The Cochrane Collaboration; 2013. editors. C. Deeks JJ, Bossuyt PM, Gatsonis C, editor.
- [31] Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;56(11):1129–35. [https://doi.org/10.1016/S0895-4356\(03\)00177-X](https://doi.org/10.1016/S0895-4356(03)00177-X).
- [32] Kim KW, Lee J, Choi SH, Huh J, Park SH. A practical review for clinical researchers-part II. statistical methods of meta-analysis. *Korean J Radiol* 2015;16(6):1175–87.
- [33] Higgins JPT GS. Summarising effects across studies. In: Wiley C, editor. *Cochrane handbook for systematic reviews of interventions*. Chichester: The Cochrane Library; 2006. Issue 4editor.
- [34] Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J. R. Stat. Soc. Ser. A (Statistics Soc.)* 2009;172(1):137–59.
- [35] Jackson D, Law M, Stijnen T, Viechtbauer W, White IR. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Stat Med* 2018;37(7):1059–85. <https://doi.org/10.1002/sim.7588>.
- [36] Hossain K, Villebro F, Forchhammer S. UAV image analysis for leakage detection in district heating systems using machine learning. *Pattern Recognit Lett* 2020;140:158–64. <https://doi.org/10.1016/j.patrec.2020.05.024>.
- [37] Egger M MC, Davey Smith G, Schneider M. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629–34.
- [38] Liu X, Li Y, Liu X, Shen J. Fault diagnosis of chillers using very deep convolutional network. In: 2018 Chinese Automation Congress (CAC); Nov. 2018. p. 1274–9. <https://doi.org/10.1109/CAC.2018.8623749>.
- [39] Miyata S, Akashi Y, Lim J, Tanaka K, Tanaka S, Kuwahara Y. Model-based fault detection and diagnosis using convolutional neural networks in heat source system with water thermal storage tanks. *IOP Conf Ser Earth Environ Sci* 2019;238(1):12070. <https://doi.org/10.1088/1755-1315/238/1/012070>.
- [40] Miyata S, Akashi Y, Lim J, Kuwahara Y, Tanaka K. Model-based fault detection and diagnosis for HVAC systems using convolutional neural network. *Build Simul Conf Proc* 2019;2(January):853–60. <https://doi.org/10.26868/25222708.2019.210311>.
- [41] Li DDDDD, Li DDDDD, Li C, Li L, Gao L. A novel data-temporal attention network based strategy for fault diagnosis of chiller sensors. *Energy Build* Sep. 2019;198:377–94. <https://doi.org/10.1016/J.ENBUILD.2019.06.034>.
- [42] Liu J, Zhang M, Wang H, Zhao W, Liu Y. Sensor fault detection and diagnosis method for AHU Using 1-D CNN and clustering analysis. *Comput Intell Neurosci* 2019;2019:1–20. <https://doi.org/10.1155/2019/5367217>.
- [43] Behravan A, Abboush M, Obermaisser R. Deep learning application in mechatronics systems' fault diagnosis, a case study of the demand-controlled ventilation and heating system. In: 2019 Adv Sci Eng Technol Int Conf ASET 2019; Mar. 2019. p. 1–6. <https://doi.org/10.1109/ICASET.2019.8714453>.
- [44] Yan K, Hua J. Deep learning technology for chiller faults diagnosis. In: 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech); Aug. 2019. p. 72–9. <https://doi.org/10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.900027>.
- [45] Gao L, Li DDDDD, Li DDDDD, Yao L, Liang L, Gao Y. A novel chiller sensors fault diagnosis method based on virtual sensors. *Sensors* Jul. 2019;19(13):3013. <https://doi.org/10.3390/s19133013>.
- [46] Zhong C, Yan K, Dai Y, Jin N, Lou B. Energy efficiency solutions for buildings: automated fault diagnosis of air handling units using generative adversarial networks. *Energies* 2019;12(3):1–11. <https://doi.org/10.3390/en12030527>.
- [47] Elnour M, Meskin N, Al-Naeemi M. Sensor fault diagnosis of multi-zone HVAC systems using auto-associative neural network. In: CCTA 2019 - 3rd IEEE Conf Control Technol Appl; 2019. p. 118–23. <https://doi.org/10.1109/CCTA.2019.8920554>.
- [48] Fan Q, Guo Y, Wu S, Liu X. Two-level diagnosis of heating pipe network leakage based on deep belief network. *IEEE Access* 2019;7:182983–92. <https://doi.org/10.1109/ACCESS.2019.2960107>.
- [49] Liu Z, Wang Y, Zhou Y. Research on fault diagnosis of air conditioner based on deep learning. *Int J Robot Control* 2019;2(1). <https://doi.org/10.5430/ijrc.v2n1p18>.
- [50] Guo Y, Fan Q, Liu X. Research on leakage diagnosis of heating pipe network based on deep belief network. In: *Proc. - IEEE Int Conf Ind Internet Cloud, ICII* 2019, no. ICII; Nov. 2019. p. 303–4. <https://doi.org/10.1109/ICII.2019.900059>.
- [51] Elnour M, Meskin N. Actuator fault diagnosis in multi-zone HVAC systems using 2D convolutional neural networks. In: 2020 IEEE Int Conf Informatics, IoT, Enabling Technol, ICIoT 2020; Feb. 2020. p. 404–9. <https://doi.org/10.1109/ICIoT4896.2020.9089508>.
- [52] Sun X, Yan K, Zhou X. Fault Detection and Diagnosis of Chillers with SD Convolutional Neural Network. In: *Proc. - IEEE Congr. Cybernetics 2020 IEEE Int. Conf. Internet Things, iThings 2020*, IEEE Green Comput. Commun. GreenCom 2020, IEEE Cyber, Phys. Soc. Comput. CPSCom 2020 IEEE Smart Data, SmartD; Nov. 2020. p. 829–36. <https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybernetics50389.2020.000141>.
- [53] Li M, Deng W, Xiahou K, Ji T, Wu Q. A data-driven method for fault detection and isolation of the integrated energy-based district heating system. *IEEE Access* 2020;8:23787–801. <https://doi.org/10.1109/ACCESS.2020.2970273>.
- [54] Cheng H, et al. Ensemble 1-D CNN diagnosis model for VRF system refrigerant charge faults under heating condition. *Energy Build* Oct. 2020;224:110256. <https://doi.org/10.1016/j.enbuild.2020.110256>.
- [55] Yan K. Chiller fault detection and diagnosis with anomaly detective generative adversarial network. *Build Environ* 2021;201:107982. <https://doi.org/10.1016/j.enbuild.2021.107982>.
- [56] Yan K, Huang J, Shen W, Ji Z. Unsupervised learning for fault detection and diagnosis of air handling units. *Energy Build* Mar. 2020;210:109689. <https://doi.org/10.1016/j.enbuild.2019.109689>.

- [57] Elnour M, Meskin N, Al-Naemi M. Sensor data validation and fault diagnosis using Auto-Associative Neural Network for HVAC systems. *J Build Eng* 2020;27(April 2019):100935. <https://doi.org/10.1016/j.jobe.2019.100935>.
- [58] Cheng F, Cai W, Zhang X, Liao H, Cui C. Fault detection and diagnosis for Air Handling Unit based on multiscale convolutional neural networks. *Energy Build* 2021;236:110795. <https://doi.org/10.1016/j.enbuild.2021.110795>.
- [59] Tian C, Wang Y, Ma X, Chen Z, Xue H. Chiller fault diagnosis based on automatic machine learning. *Front Energy Res Oct* 2021;9:1–17. <https://doi.org/10.3389/fenrg.2021.753732>.
- [60] Liao H, Cai W, Cheng F, Dubey S, Rajesh PB. An online data-driven fault diagnosis method for air handling units by rule and convolutional neural networks. *Sensors* 2021;21(13). <https://doi.org/10.3390/s21134358>.
- [61] Fan C, Li X, Zhao Y, Wang J. Quantitative assessments on advanced data synthesis strategies for enhancing imbalanced AHU fault diagnosis performance. *Energy Build* 2021;252:111423. <https://doi.org/10.1016/j.enbuild.2021.111423>.
- [62] Li G, et al. An explainable one-dimensional convolutional neural networks based fault diagnosis method for building heating, ventilation and air conditioning systems. *Build Environ* 2021;203(March):108057. <https://doi.org/10.1016/j.buildenv.2021.108057>.
- [63] Zhou Z, Li G, Chen H, Zhong H. Fault diagnosis method for building VRF system based on convolutional neural network: considering system defrosting process and sensor fault coupling. *Build Environ* 2021;195:107775. <https://doi.org/10.1016/j.buildenv.2021.107775>.
- [64] Li B, Cheng F, Zhang X, Cui C, Cai W. A novel semi-supervised data-driven method for chiller fault diagnosis with unlabeled data. *Appl Energy* 2021;285(December 2020):116459. <https://doi.org/10.1016/j.apenergy.2021.116459>.
- [65] Liu J, et al. Transfer learning-based strategies for fault diagnosis in building energy systems. *Energy Build* 2021;250:111256. <https://doi.org/10.1016/j.enbuild.2021.111256>.
- [66] Gao L, Li D, Yao L, Gao Y. Sensor drift fault diagnosis for chiller system using deep recurrent canonical correlation analysis and k-nearest neighbor classifier. *ISA Trans Mar* 2022;122:232–46. <https://doi.org/10.1016/J.ISATRA.2021.04.037>.
- [67] Fan C, He W, Liu Y, Xue P, Zhao Y. A novel image-based transfer learning framework for cross-domain HVAC fault diagnosis: from multi-source data integration to knowledge sharing strategies. *Energy Build* 2022;262:111995. <https://doi.org/10.1016/j.enbuild.2022.111995>.
- [68] Han S, Shao H, Huo Z, Yang X, Cheng J. End-to-end chiller fault diagnosis using fused attention mechanism and dynamic cross-entropy under imbalanced datasets. *Build Environ* Mar. 2022;212:108821. <https://doi.org/10.1016/J.BUILDENV.2022.108821>.
- [69] Sun Z, et al. Severity-insensitive fault diagnosis method for heat pump systems based on improved benchmark model and data scaling strategy. *Energy Build* 2022;256:111733. <https://doi.org/10.1016/j.enbuild.2021.111733>.
- [70] Wu B, Cai W, Cheng F, Chen H. Simultaneous-fault diagnosis considering time series with a deep learning transformer architecture for air handling units. *Energy Build* Feb. 2022;257:111608. <https://doi.org/10.1016/j.enbuild.2021.111608>.
- [71] Pasini A. Artificial neural networks for small dataset analysis. *J Thorac Dis* 2015;7(5):953–60. <https://doi.org/10.3978/j.issn.2072-1439.2015.04.61>.
- [72] Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabouj M, Inman DJ. 1D convolutional neural networks and applications: a survey. *Mech Syst Signal Process* Apr. 2021; 151:107398. <https://doi.org/10.1016/j.ymssp.2020.107398>.
- [73] Park J-WW, Sim S-HH, Jung H-JJ. Displacement estimation using multimetric data fusion. *IEEE/ASME Trans Mechatr* Dec. 2013;18(6):1675–82. <https://doi.org/10.1109/TMECH.2013.2275187>.
- [74] Li H, Huang J, Yang X, Luo J, Zhang L, Pang Y. Fault diagnosis for rotating machinery using multiscale permutation entropy and convolutional neural networks. *Entropy* Jul. 2020;22(8):101–10. <https://doi.org/10.3390/E22080851>.
- [75] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Networks* Mar. 1994;5(2):157–66. <https://doi.org/10.1109/72.279181>.
- [76] Cabrera D, et al. Bayesian approach and time series dimensionality reduction to LSTM-based model-building for fault diagnosis of a reciprocating compressor. *Neurocomputing* Mar. 2020;380:51–66. <https://doi.org/10.1016/j.neucom.2019.11.006>.
- [77] Alsaleem FM, Abiprojo R, Arensmeier J, Hemmelgarn G, Louis S. HVAC system cloud based diagnostics model. In: International Refrigeration and Air Conditioning Conference; 2014. p. 1–9. 2002[Online]. Available: <http://docs.lib.psu.edu/iracc>.
- [78] X. Zhang, Y. Cong, Z. Yuan, T. Zhang, and X. Bai, “Early fault detection method of rolling bearing based on MCNN and GRU network with an attention mechanism,” *Shock Vib.*, pp. 1–13, Mar. 2021, doi: 10.1155/2021/6660243.
- [79] Mirzaei S, Kang J-L, Chu K-Y. A comparative study on long short-term memory and gated recurrent unit neural networks in fault diagnosis for chemical processes using visualization. *J Taiwan Inst Chem Eng Jan.* 2022;130:104028. <https://doi.org/10.1016/j.jtice.2021.08.016>.
- [80] Yan X, Guan T, Fan K, Sun Q. Novel double layer BiLSTM minor soft fault detection for sensors in air-conditioning system with KPCA reducing dimensions. *J. Build. Eng.* Dec. 2021;44(June):102950. <https://doi.org/10.1016/j.jobe.2021.102950>.
- [81] Klys J, Snell J, Zemel R. Learning latent subspaces in variational autoencoders. *Adv Neural Inf Process Syst*, 2018-Decem.; Dec. 2018. p. 6444–54 [Online]. Available: <http://arxiv.org/abs/1812.06190>.
- [82] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial network. In: *Proceedings of the 34th International Conference on Machine Learning*; 2017. p. 214–23.
- [83] Xiaolong W, Gupta A, Wang X, Gupta A. Generative image modeling using style and structure adversarial networks BT - Computer Vision – ECCV 2016. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)* 2016;9909 LNCS:V. <https://doi.org/10.1007/978-3-319-46493-0>.
- [84] Chen H. Challenges and corresponding solutions of generative adversarial networks (GANs): a survey study. *J Phys Conf Ser* 2021;1827(1):12066. <https://doi.org/10.1088/1742-6596/1827/1/012066>.
- [85] Van Erven T, Harremoës P. Rényi Divergence and Kullback – Leibler Divergence. *IEEE Trans Inf Theory* 2014;60(7):3797–820.
- [86] Sampath V, Maurtua I, Aguilera Martín JJ, Gutierrez A. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J Big Data* 2021;8(1):27. <https://doi.org/10.1186/s40537-021-00414-0>.
- [87] Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput* Jul. 2006;18(18):1527–54. <https://doi.org/10.1109/TNN.2006.880582>.
- [88] Peng B-Ssen, Xia H, Liu Y-KK, Yang B, Guo D, Zhu S-MM. Research on intelligent fault diagnosis method for nuclear power plant based on correlation analysis and deep belief network. *Prog Nucl Energy* 2018;108(July):419–27. <https://doi.org/10.1016/j.pnucene.2018.06.003>.
- [89] Pinaya WHLL, et al. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. *Sci Rep* 2016;6(December):1–9. <https://doi.org/10.1038/srep38897>.
- [90] Tamilselvan P, Wang P. Failure diagnosis using deep belief learning based health state classification. *Reliab Eng Syst Saf* 2013;115:124–35. <https://doi.org/10.1016/j.ress.2013.02.022>.
- [91] Mohamed AR, Yu D, Deng L. Investigation of full-sequence training of deep belief networks for speech recognition. In: Proc. 11th Annu Conf Int Speech Commun Assoc INTERSPEECH 2010; 2010. p. 2846–9. <https://doi.org/10.21437/interspeech.2010-304>. October 2014.
- [92] Deb C, Zhang F, Yang J, Lee SE, Shah KW. A review on time series forecasting techniques for building energy consumption. *Renew Sustain Energy Rev* Jul. 2017; 74:902–24. <https://doi.org/10.1016/j.rser.2017.02.085>.
- [93] Ding Y, Jia M, Miao Q, Cao Y. A novel time-frequency Transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings. *Mech Syst Signal Process* 2022;168(December 2021):108616. <https://doi.org/10.1016/j.ymssp.2021.108616>.
- [94] M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu, “Beyond self-attention: external attention using two linear layers for visual tasks,” *arXiv Prepr. arXiv:2105.02358*, vol. 14, no. 8, pp. 1–11, 2021, doi: 10.48550/arXiv.2105.02358.
- [95] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30.
- [96] Hosmer Jr DW, et al. *Applied logistic regression*. Third Edit 2013;398.
- [97] Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. *J Clin Epidemiol* 2000;53(2):207–16. [https://doi.org/10.1016/S0895-4356\(99\)00161-4](https://doi.org/10.1016/S0895-4356(99)00161-4).
- [98] Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing meta-analysis with r a hands-on guide. Chapman and Hall/CRC; 2021.
- [99] Agami Reddy T. Development and evaluation of a simple model-based automated fault detection and diagnosis (FDD) method suitable for process faults of large chillers. *ASHRAE Trans* 2007;113:27–39.