

Novel Actuator Fault Diagnosis Framework for Multizone HVAC Systems Using 2-D Convolutional Neural Networks

Mariam Elnour and Nader Meskin^{ID}, *Senior Member, IEEE*

Abstract—Heating, ventilation, and air conditioning (HVAC) systems are used to condition the indoor environment in buildings. They can be subjected to malfunctioning since they are the most extensively operated buildings' components that account alone for almost half of the total building energy usage. Therefore, fault diagnosis (FD) of the HVAC system is important to maintain the system's reliability and efficiency and provide preventive maintenance. This article presents a supervised FD strategy for single actuator faults in HVAC systems given that actuators, such as dampers and valves, are mostly prone to faults resulting in thermal discomfort and energy inefficiency in buildings. The proposed approach is based on 2-D convolutional neural networks (CNNs) using an efficient 1-D-to-2-D data transformation performed on the time-series signals acquired from the HVAC system. The performance of the CNNs is ensured by an optimal tuning of its significant hyperparameters using the Bayesian optimization algorithm toward maximizing the classification accuracy. The proposed 1-D-to-2-D data transformation approach is computationally efficient and eliminates the use of advanced signals preprocessing. It is performed in two schemes: the static and dynamic schemes to analyze the correlation between the system's variables and consider the temporal effects of the time-series signals without compromising the detection time. The proposed approach is developed and validated using simulation data collected from a three-zone HVAC system simulator using Transient System Simulation Tool (TRNSYS). It demonstrates improved performance compared to the 1-D CNN-based approach and the other standard data-driven approaches for actuator FD in HVAC systems.

Note to Practitioners—This article was motivated by the problem of actuator faults, such as dampers and valves in heating, ventilation, and air conditioning (HVAC) systems. They are at risk of malfunctioning or failure resulting in reduced system efficiency, potential interference with supervision schemes execution, and equipment deterioration. This article presents a novel approach for single actuator faults in HVAC systems using the recently evolving topology of convolutional neural networks that are known for their high accuracy and reliable performance compared to the standard approaches. It utilizes the building historical data that can be obtained from the building management systems and without the need for either expert knowledge or details about the HVAC system. In future research, we plan to address the problem of multiple actuator faults, as well as simultaneous sensor and actuator faults.

Manuscript received January 25, 2021; revised March 9, 2021; accepted March 18, 2021. This article was recommended for publication by Editor Q. Zhao upon evaluation of the reviewers' comments. (*Corresponding author: Nader Meskin*)

The authors are with the Department of Electrical Engineering, Qatar University, Doha, Qatar (e-mail: nader.meskin@qu.edu.qa; me1003659@qu.edu.qa).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TASE.2021.3067866>.

Digital Object Identifier 10.1109/TASE.2021.3067866

Index Terms—Actuator fault diagnosis (FD), convolutional neural network (CNN), heating, ventilation, and air conditioning (HVAC) system.

I. INTRODUCTION

GLOBALLY, the building sector accounts for 30% of the energy consumption and more than 55% of the electricity demand. Specifically, the heating, ventilation, and air conditioning (HVAC) system is the most extensively operated component, and it is responsible alone for 40% of the final building energy consumption [1]. It is used to provide healthy and comfortable indoor conditions for occupants with minimum energy usage. The development toward sustainable buildings is progressing, but improvements are still not keeping up with the growing building sector and the increasing global demand for energy. HVAC systems include a considerable number of sensors, controlled actuators, and other components, such as filters and dampers. They are at risk of malfunctioning or failure resulting in reduced efficiency, potential interference with supervision schemes execution, and equipment deterioration. Hence, fault diagnosis (FD) of HVAC systems is essential to improve their reliability, efficiency, and performance and provide preventive maintenance.

There are several research works investigating and proposing various solutions for HVAC system actuator FD, given that actuators—such as damper and valves—are mostly prone to faults resulting in thermal discomfort and energy inefficiency in buildings. A summary of the proposed approaches found in the literature for actuator FD of HVAC systems is presented in Table I, which are model-based methods [2]–[4], data-driven methods [5]–[10], rule-based approaches [11], and hybrid approaches [12]–[16] using a combination of the previous approaches. Model-based FD methods require the mathematical model of the HVAC system. They rely on monitoring the consistency between the measured outputs of the practical systems and the model-predicted outputs to identify a fault incident as in [2] utilizing total variation (TV) filter with an enhanced particle filter (PF), in [3] based on unscented Kalman filtering (UKF), and in [4] using state observers.

Several data-driven approaches using various machine learning algorithms were developed for HVAC actuator FD, such as support vector machine (SVM) [5], principal component analysis (PCA) [6], standard neural networks (NNs) [7], [8], and Bayesian networks [9], [10]. In addition, hybrid methods that combine expert knowledge and data-driven approaches were presented in [12]–[15] and using model-based and

TABLE I
SUMMARY OF HVAC SYSTEM ACTUATOR FD METHODS IN THE LITERATURE

[Ref.]	Method	Model-based	Method type Data-based	Rule-based	Data source	Fault type
[2]	TVF and KF	✓			-	Stuck valve
[3]	UKF	✓			-	Stuck valve
[4]	Observer	✓			-	Stuck valve
[5]	SVM (Supervised learning)	✓			Simulation (MATLAB)	Stuck damper/valve
[6]	PCA and pattern matching (Unsupervised learning)	✓			Experimental	Stuck damper
[7]	Neural networks (Supervised learning)	✓			Simulation (TRNSYS)	Stuck valve
[8]	Neural networks (Supervised learning)	✓			Simulation (EnergyPlus)	Stuck damper/valve
[9, 10]	Bayesian network (Supervised learning)	✓			Experimental	Stuck damper/valve
[11]	Control chart and rule-based			✓	Operational	Stuck valve
[12]	ANN and fuzzy logic (Supervised learning)	✓	✓		Simulation (MATLAB)	Stuck valve
[13]	Diagnostic Bayesian networks	✓	✓		Experimental	Stuck damper
[14]	Diagnostic Bayesian networks	✓	✓		Simulation	Stuck valve
[15]	PCA and rule-based	✓	✓	✓	Experimental	Stuck damper/valve
[16]	KF/UPF and HMM	✓	✓		Simulation and Experimental	Stuck damper/valve
-	Convolutional neural networks (proposed method)		✓		Simulation (TRNSYS)	Stuck damper/valve

data-driven approaches as in [16], which utilizes Kalman Filter (KF), unscented PF (UPF), and hidden Markov model (HMM). The development in this field is still progressing, but it is still not keeping up with the growing buildings sector and the increasing global demand for energy. For instance, some developed approaches have limited performance capability for large-scale buildings [5], and others do not address the fault isolation matter [6]. The limitation of the proposed approaches in [11]–[15] is their dependence on a set of expert rules, while the other methods [7]–[9] suffer from limited diagnosis accuracy due to the relatively high rate of false alarms (or missed alarms).

In this article, we follow up the preliminary findings in [17] by further investigating and demonstrating the deployment of the recently evolving topology of the convolutional neural networks (CNNs) with Bayesian optimization for actuator FD in HVAC systems in its different configurations, which are 1-D CNN, static 2-D CNN, and dynamic 2-D CNN. That is, CNNs are characterized by their high-performance accuracy and extreme capability in learning and realizing complex functions and interdependency from any given data in a form of a network model. They are mainly used for computer vision and image processing applications and have been recently used with 1-D data for condition monitoring and FD [18]–[29].

1-D CNNs were used for bearing fault detection in [18], structural damage detection in [19], and modular multilevel converter (MMC) circuits FD in [20]. They are effective when dealing with relatively fast signals, in which particular features or patterns can be captured in short fixed-length segments. In addition, 2-D CNNs were used with 1-D signals as well where the raw data were converted to a 2-D configuration using various conversion approaches. For instance, a 2-D CNN-based method was proposed in [21] for bearing FD using vibration images produced by signal amplitude to pixel intensity mapping, while [22] used the actual images of the vibration signals segments as the CNN inputs. A process FD framework was developed in [23], in which the time-series process variables were configured into a matrix, where the x -axis represented time and the y -axis represented the process variables.

In other works, signals processing techniques were applied to the raw signals to obtain their 2-D representations as

in [24] using the discrete Fourier transform (DFT), in [25] using time-frequency representation, in [26] using continuous wavelet transform scalogram (CWTS), and in [27] using spectral energy maps (SEMs) for bearing and rotary machinery FD. In [28], the planetary gearbox FD method was presented using discrete wavelet transform (DWT) on vibration signals, and in [29], a CNN-based stability monitoring system for power systems was developed using the heatmap representation of the measurements.

The limitation of the aforementioned CNN-based algorithms found in the literature is that they require advanced data preprocessing and/or an indicative sequence of input data to make the diagnosis decision. For slow signals such as temperature, the time frame window should be large enough for the network to learn to capture the fault occurrence, which consequently leads to longer detection time and increased computational requirements.

The contribution of the work presented in this article can be summarized as follows.

- 1) A supervised data-driven FD approach using 2-D CNNs is proposed for single actuator faults in the HVAC system utilizing the system's historical data and without the need for expert knowledge or the mathematical models of the HVAC system.
- 2) A novel 1-D-to-2-D data transformation approach is proposed by configuring the time-series signals into a 3-D matrix from each data sample that is done in two schemes, which are the static configuration scheme, in which each data sample represents a single instant of time, and the dynamic configuration scheme using a short sequence of time instants in order to account for the temporal features in the time-series signals. It aims to tackle the limitations of the state-of-the-art CNN-based FD approaches in terms of the data preprocessing requirements, computational complexity, and detection time.
- 3) The proposed approach demonstrates a notable improvement compared to the 1-D CNN-based FD method and the standard data-driven approaches in the previous works.

This article is organized as follows. In Section II, the description of the HVAC system under study is presented.

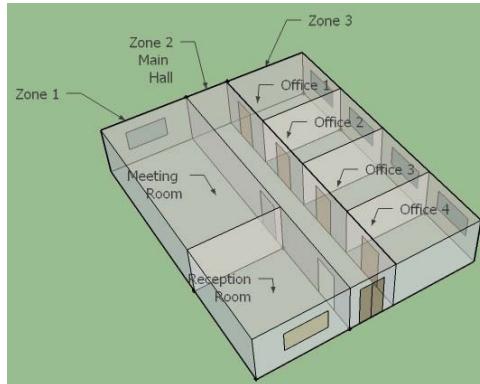


Fig. 1. Sketch of the simulated three-zone building.

The main details about the CNN are presented in Section III. The descriptions of the proposed CNN-based HVAC actuator FD framework along with the development details are provided in Section IV. The details of the validation phase of the proposed solution in terms of the evaluation metrics used and the results obtained are presented and demonstrated in Section V. Finally, conclusions and future work are summarized in Section VI.

II. DESCRIPTION OF THE HVAC SYSTEM

The building under this study is a one-floor office building presented in [30]. It is assumed to be located in Doha city in Qatar and operating from 6 A.M. to 8 P.M. during weekdays (Sunday to Thursday). It consists of three zones with a total floor area of 200 m^2 , as shown in Fig. 1. Zones 1 and 3 are of a volume of 240 m^3 and floor area of 80 m^2 each, and the first one includes a meeting room and a reception room, while the latter has four identical office rooms. The main hall is considered as a separate zone with a volume of 120 m^3 and a floor area of 40 m^2 .

The system is simulated using the Transient System Simulation Tool (TRNSYS), which is a graphical software environment that allows the simulation of transient systems behavior through energy and mass balance equations [31]. The simulation tool is used to simulate the actual dynamics of the HVAC system and then to generate data for both normal and faulty operations of the system. The building geometry and structure, and the outside weather conditions are taken into consideration in the simulator design, in addition to the occupancy, and the internal loads, such as equipment and lighting, as presented in Table II. TRNSYS has been widely used for HVAC systems simulation for research and development purposes as in [7] and [32]–[36], and it is found to be a reliable tool to simulate the system operation.

The building is equipped with a simple HVAC system for cooling application shown in Fig. 2, in which only the temperature at each zone is controlled using proportional integral derivative (PID) controllers. Shut-off variable air volume (VAV) systems are used since no heating is required as they are known to be used for cooling purposes in applications having a year-round cooling load. The supplied airflow to the zones is reduced as the cooling load decreases, and the VAV box is allowed to reduce the airflow to zero during

TABLE II
INTERNAL HEAT GAIN SOURCES IN THE SIMULATED
THREE-ZONE BUILDING

Room	Details
Meeting room	Occupation: 4 persons (8AM-9AM) Person: Seated, very light work, 120 W Lights: 5 W/m^2
Reception room	Occupation: 1 person (6 AM-8 PM) Person: Seated, light work, typing, 150 W Computer: 140 W Lights: 5 W/m^2
Main hall	Occupation: 4 persons (6 AM-8 PM) Person: Standing, light work, 120 W Lights: 5 W/m^2
Office room	Occupation: 1 person (6 AM-8 PM) Person: Seated, light work, typing, 150 W Computer: 140 W Lights: 5 W/m^2

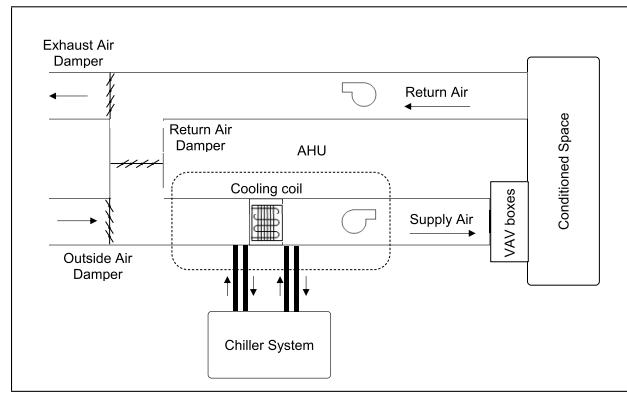


Fig. 2. Diagram of a typical HVAC system for cooling application using the VAV system.

periods of no cooling load. The HVAC system is simulated using TRNSYS, in which the cold output air from the Air Handling Unit (AHU) is supplied through supply air ducts and a supply air fan to the zones ducts with the VAV boxes terminals. The zones temperature controllers modulate the position of the air dampers according to the thermal load of the zones and the variation in weather conditions to achieve the desired setpoints. The zones return air (RA) is fed back to the AHU through the RA ducts using the return fan. The exhaust air (EA) dampers, outside air (OA) dampers, and RA dampers are operated simultaneously to control the fraction of the recirculated air and the ventilation air in order to maintain the indoor air quality.

The zones are supplied from the AHU with the cold air of constant temperature of 13°C and a variable flow rate controlled by the VAV box terminals. The water chiller and the cooling coil are connected via the chilled water tank. The water tank temperature is controlled at 11°C via a water valve regulating the flow of chilled water from the chiller to the tank. The system's variables of interest for the current study are as follows: the zones temperatures T_{z1} , T_{z2} , and T_{z3} , the temperature of the chilled water tank T_t , the temperature of the AHU supply air T_{ao} , the temperature of the return water from the cooling coil to the water tank T_{wo} , the ambient temperature T_{amb} , the zones VAV boxes control signals U_1 , U_2 , and U_3 , and the water tank valve control signal U_4 .

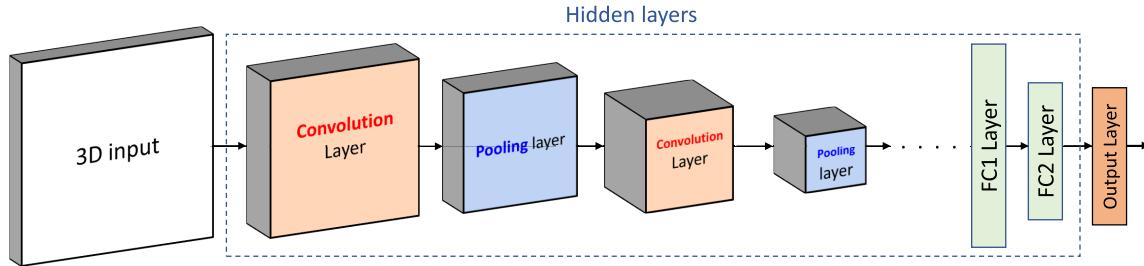


Fig. 3. Typical CNN architecture containing convolution layers, pooling layers, and FC layers.

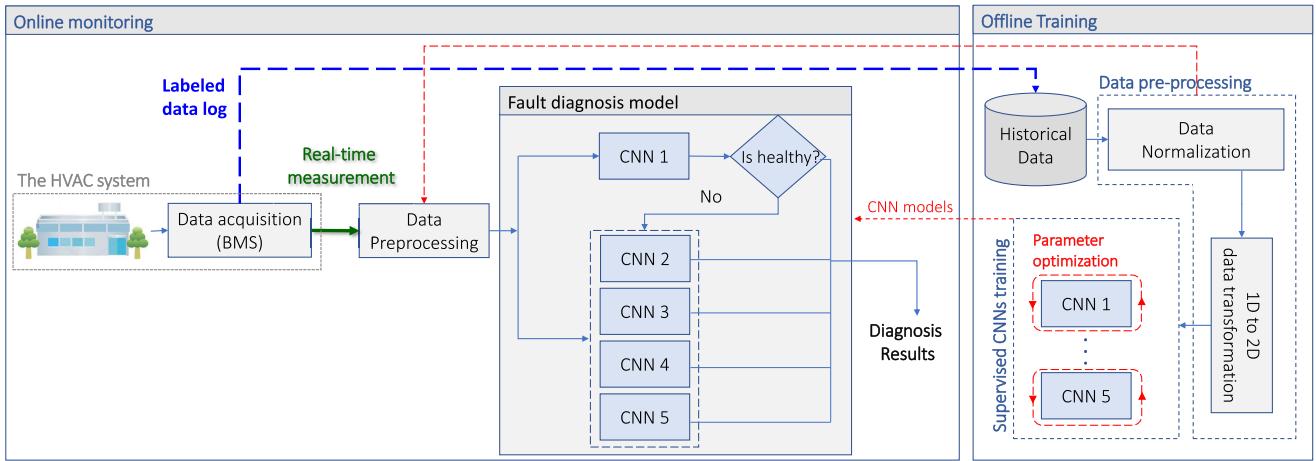


Fig. 4. Proposed 2-D CNN-based HVAC system actuator FD method framework. The labeled historical data of the HVAC system are used for the CNNs' off-line training. The data are normalized and configured into the 2-D configuration and then are used to develop the CNN models. The FD model is composed of the five CNNs, each trained in a supervised learning scheme to perform binary classification—one-versus-all—for each of the five statuses of the system, as described in Table III. Following the completion of the off-line training, the online monitoring is performed using the finalized FD model on the preprocessed real-time system data.

III. CONVOLUTIONAL NEURAL NETWORKS

The CNN is a special type of feed-forward neural network characterized by its internal structure, which contains the convolution (Conv) layer, the pooling layer, and the fully connected (FC) layers, as shown in Fig. 3. It is trained using the backpropagation algorithm to optimize the cross-entropy function given by [37]

$$E = \sum_{i=1}^m \sum_{j=1}^k t_{ij} \ln y_{ij} \quad (1)$$

where m is the number of samples, k is the number of classes, t_{ij} is the j th element of the target vector \mathbf{t}_i , and y_{ij} is the j th element of the prediction output vector $\mathbf{y}_i \in \mathbb{R}^k$. The details of the CNN layers are as follows.

- 1) *Convolution Layer:* There are a number of hyperparameters of the convolution layer, which are as follows.
 - a) The activation function, f , which enables the network to learn and approximate complex functional mappings between the inputs and output targets.
 - b) The size of filter, $k = [k_1, k_2]$ where $k_i \in \mathbb{R}$ defines the field of view of the convolution operation.
 - c) The number of filters, $F_c \in \mathbb{R}$, representing the number of feature maps.
 - d) The strides, $s \in \mathbb{R}$ representing the size of the convolution step.

e) The type of convolution, which can be *same convolution* where zero-padding is performed, *valid convolution* where zero-padding is not performed, or *dilated convolution* that is used to increase the receptive field of the layer.

- 2) *Pooling Layer:* The pooling layer—also known as the subsampling layer—aims to reduce the spatial size of the convolution layer output in order to reduce the number of parameters and the amount of computation.
- 3) *FC Layer:* It has nodes with full pairwise connections with the adjacent layers. The number of neurons in the FC layer $F_{\text{FC}} \in \mathbb{R}$ is a hyperparameter to be set in addition to the type of activation function.

IV. PROPOSED CNN-BASED ACTUATOR FD FRAMEWORK

As illustrated in Fig. 4, the proposed 2-D CNN-based diagnosis framework consists of an off-line stage, in which the CNNs are trained using the historical building data using the Bayesian optimization for hyperparameters tuning, and an on-line stage such that the real-time measurements of the system variables are acquired by the building management system (BMS) and used to determine the diagnosis decision.

The proposed diagnosis approach is a multimodel framework that is composed of five two-class CNNs. CNN 1 performs fault detection by determining whether the system is fault-free or not. The remaining CNNs are trained to diagnose

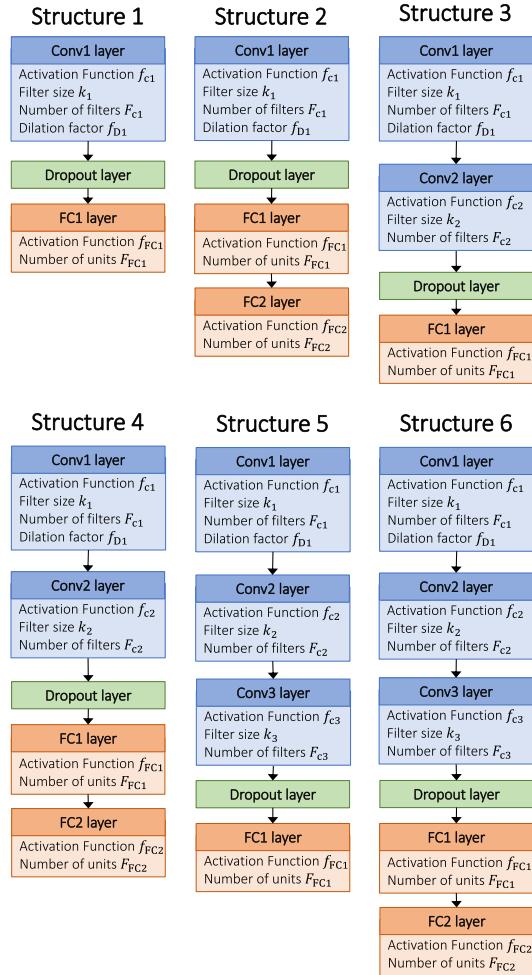


Fig. 5. Exploited network architectures ranging from shallow to deep.

TABLE III
LABELS OF CLASSES ACCORDING TO THE SYSTEM STATE

Actuator type	System state	Class identification
-	Healthy	Class A
Zone 1 VAV box damper	Faulty at actuator 1	Class B
Zone 2 VAV box damper	Faulty at actuator 2	Class C
Zone 3 VAV box damper	Faulty at actuator 3	Class D
Water valve	Faulty at actuator 4	Class E

each type of the concerned actuator faults independently, which are four types. Each CNN was trained to perform binary classification of “belongs to the class or not” such that the network’s outputs were labeled positive for the concerned class and negative otherwise. The HVAC system under this study has four actuators, which are the three zones’ VAV box dampers and the water valve in the chiller-tank link. The system status can be denoted by five possible classes that are healthy, faulty in actuator 1, faulty in actuator 2, faulty in actuator 3, and faulty in actuator 4, as listed in Table III.

Different network architectures were designed, as shown in Fig. 5, based on the depth of the network, i.e., the number of convolution layers and FC layers. The pooling layers for the current application are not needed given the limited dimensionality of the system data. The hyperparameters of the CNNs are tuned using the Bayesian optimization algorithm.

A. 1-D-to-2-D Data Transformation Technique

The 1-D-to-2-D data transformation is performed to enable using the 2-D CNN-based diagnosis approach with 1-D signals. The variables of the system are time-series signals, i.e., temperature, control signals, and so on. They need to be reshaped before being used in the 2-D CNN-based diagnosis model. The first step of the signal-to-matrix conversion is the normalization of the temperature measurements to be in the range of 0–1. The normalized sensor data along with the control signals are configured—with proper zero-padding into a 3-D matrix from each data sample such that three different 2-D configurations are produced, each by using a distinct permutation of the system’s variables.

The purpose of the three-channel 2-D data configuration is to promote extracting diverse correlation characteristics between the variables of the system—i.e., the sensors measurements and the control signals by the different convolution filters. That is, since CNNs exploit the spatial local correlation of the data being examined, the proposed 1-D-to-2-D data transformation technique aims to configure the 1-D data of the system variables into three different 2-D spatial arrangements, as demonstrated in Fig. 6, in which each arrangement represents a channel of the 3-D reshaped data form. Each channel is handled independently by the convolution kernels to extract the useful features toward optimizing the cost function.

It is done in two schemes, which are the static configuration scheme in which each data sample represents a single instant of time, and the dynamic configuration scheme using a sequence of time instants in order to account for the temporal features in the time-series signals. For example, in the static 2-D transformation, the data sample for the system under study is produced by configuring the 11 system variables’ readings of a single time instant t_i into the 3-D matrix of size $3 \times 4 \times 3$. While using the dynamic 2-D transformation, the data sample is produced based on configuring the 11 system variables’ readings of a sequence of time instants a window $W - t = [t_s, W + t_s]$ into the 3-D matrix of size $3 \times 4W \times 3$.

The window size W is a hyperparameter that determines the extent of the examined temporal features of the time series. The arrangement $S \in \mathbb{R}^3$ of the system’s variables into the 3-D matrix is another hyperparameter to be decided. That is, there are enormous possibilities of how the variables can be configured into the 2-D representation, and it corresponds to the number of its permutations. However, the Bayesian optimization is used for the network’s parameters tuning in this work, and hence, this hyperparameter is jointly optimized with the other hyperparameters of the CNN such as the filter size k and dilation factor f_D resulting in reducing the hyperparameter space of S .

B. Training of the CNN Models

The training of the CNNs in the proposed FD framework was carried using all the class data such that the targets were labeled as positive outputs for the concerned class and negative otherwise with proper data balancing. That is, in data-driven classification approaches, a common issue is encountered when data of some classes are scarce (e.g., faulty data), while, for others, data are abundant (e.g., healthy data). The

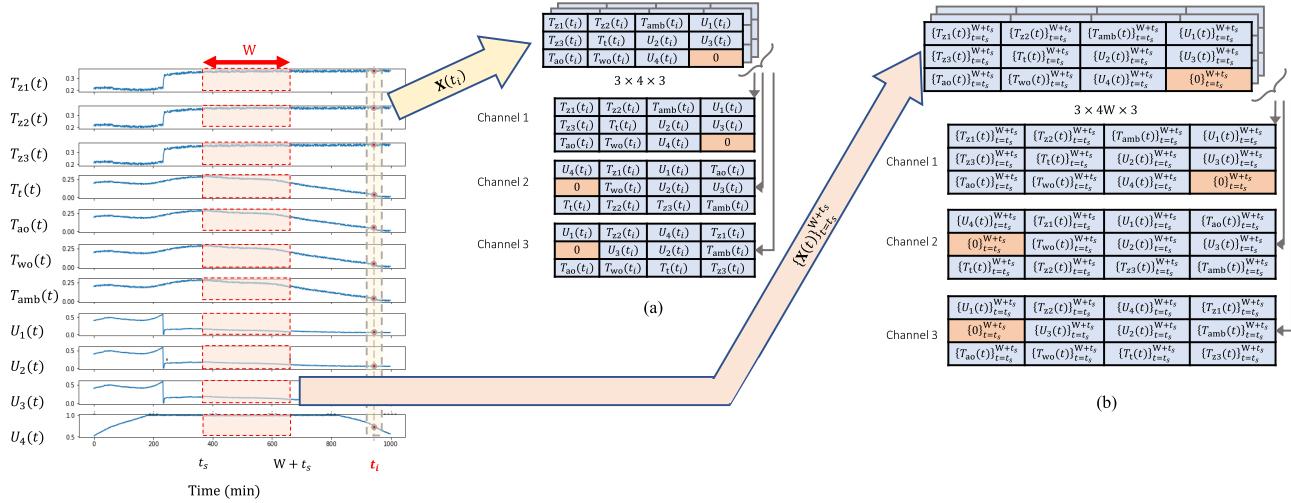


Fig. 6. Demonstration of the proposed 1-D-to-2-D data transformation approach based on reshaping the 11 system's 1-D data samples into a 3-D configuration in both static and dynamic forms. (a) Data sample produced by the static 2-D data transformation, which is based on configuring the 11 system variables' readings of a single time instant t_i into the 3-D matrix of size $3 \times 4 \times 3$. (b) Data sample produced by the dynamic 2-D data transformation, which is based on configuring the 11 system variables' readings of a sequence of time instants—a window $W-t = [t_s, W+t_s]$ into the 3-D matrix of size $3 \times 4W \times 3$.

TABLE IV

HYPERPARAMETERS' RANGES USED FOR THE CNN MODEL'S TUNING USING THE BAYESIAN OPTIMIZATION

Parameter	Range
Window size (W)	5 - 20
Filter size (k)	2 - 5
Dilation factor (f_D)	1 - 5
Number of Conv layer's filters (F_c)	10 - 60
Number of FC layer's units (F_{FC})	10 - 200
Activation function (f)	ReLU, Leaky ReLU, Sigmoid, Tanh

classification performance of the network is altered such that a bias is introduced to the network predictions since the network model will pay more attention to the majority class. The effect of this issue can be minimized by modifying the sampling rate, which can be done by majority undersampling, minority oversampling, or a combination of both with the proper design of the network structure to avoid overfitting problems.

For example, for a CNN trained to identify Class A, the data samples of Class B to Class E were labeled as negative outputs, and the healthy data were identified with positive labels and downsampled to avoid data imbalance, such that the number of negative and positive samples is equal. For training the CNNs for Class B to Class E, both majority downsampling for the healthy data and minority upsampling for the concerned class data were used to balance the data set. It is worth noting that data balancing was performed on the training data set only. Using MATLAB, the CNNs were trained using the Adam optimization algorithm with piecewise learning decay to optimize the cost function using an initial learning rate of 0.03, a minibatch size of 128, and a maximum number of epochs of 5, utilizing a PC of 64-GB RAM and 12-core AMD Ryzen 9 3900X CPU with 3.8-GHz speed using a 64-bit Windows 10 Pro OS.

The data used for the network training and validation were collected from the TRNSYS system simulator. The total raw data set size is 230k samples with 150k samples corresponding to the normal system operation and 20k samples for each of

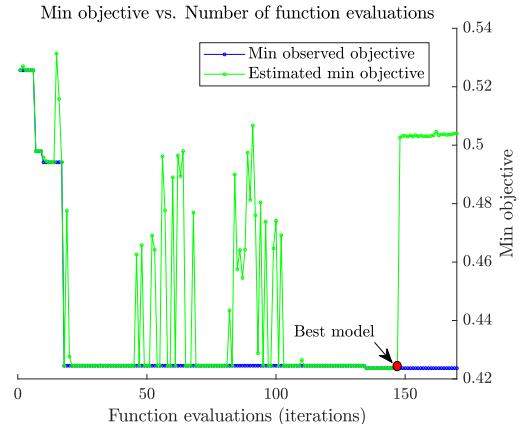


Fig. 7. Progress of the hyperparameter tuning of CNN 1 using Bayesian optimization.

the four types of the actuator fault. The actuators in the system under study are three VAV box dampers and one water valve. The data were collected at a sampling time of 1 min for two months of the healthy system operation. The faults data were acquired for 13 days per fault type for different fault scenarios, which are stuck at fully closed (FC) position, stuck at fully open (FO) position, and stuck at partially open (PO) position with different severity levels.

The Bayesian optimization was used to tune the hyperparameters of the convolution neural networks for the ranges presented in Table IV using a maximum number of iterations of 300. The training was conducted using sevenfold cross-validation to evaluate the networks' generalization ability and hence avoid overfitting problems. Figs. 7 and 8 demonstrate a sample of the tuning process of the CNNs for Class A and Class D using the Bayesian optimization algorithm, respectively. The blue graph represents the actual objective function. The green graph represents the estimated objective function, which is an approximation of the objective

TABLE V
DETAILS OF THE BEST TRAINED CNNs FOR THE STATIC 2-D CNN-BASED FD FRAMEWORK

Network	Convolution layers	Fully connected layers
S-CNN-S1-A	Conv1: $k = [2,2]$, $F_c = 17$, $f_D = 1$, $f = \text{ReLU}$	FC1: $F_{FC} = 51$, $f = \text{Leaky ReLU}$
S-CNN-S2-B	Conv1: $k = [2,2]$, $F_c = 19$, $f_D = 1$, $f = \text{ReLU}$	FC1: $F_{FC} = 55$, $f = \text{Leaky ReLU}$ FC2: $F_{FC} = 11$, $f = \text{Leaky ReLU}$
S-CNN-S1-C	Conv1: $k = [2,2]$, $F_c = 20$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 120$, $f = \text{Leaky ReLU}$
S-CNN-S1-D	Conv1: $k = [2,2]$, $F_c = 16$, $f_D = 1$, $f = \text{ReLU}$	FC1: $F_{FC} = 172$, $f = \text{Leaky ReLU}$
S-CNN-S3-E	Conv1: $k = [2,2]$, $F_c = 17$, $f_D = 1$, $f = \text{Leaky ReLU}$ Conv1: $k = [1,2]$, $F_c = 34$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 172$, $f = \text{ReLU}$

TABLE VI
DETAILS OF THE BEST TRAINED CNNs FOR THE DYNAMIC 2-D CNN-BASED FD FRAMEWORK

Network	W	Convolution layers	Fully connected layers
D-CNN-S2-A	9	Conv1: $k = [2,4]$, $F_c = 16$, $f_D = 1$, $f = \text{ReLU}$	FC1: $F_{FC} = 97$, $f = \text{Leaky ReLU}$ FC2: $F_{FC} = 19$, $f = \text{ReLU}$
D-CNN-S2-B	6	Conv1: $k = [2,4]$, $F_c = 19$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 185$, $f = \text{ReLU}$ FC2: $F_{FC} = 37$, $f = \text{ReLU}$
D-CNN-S2-C	5	Conv1: $k = [2,2]$, $F_c = 19$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 138$, $f = \text{Leaky ReLU}$ FC2: $F_{FC} = 27$, $f = \text{ReLU}$
D-CNN-S2-D	5	Conv1: $k = [2,4]$, $F_c = 18$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 102$, $f = \text{ReLU}$ FC2: $F_{FC} = 20$, $f = \text{Leaky ReLU}$
D-CNN-S2-E	9	Conv1: $k = [2,3]$, $F_c = 18$, $f_D = 2$, $f = \text{ReLU}$	FC1: $F_{FC} = 150$, $f = \text{ReLU}$ FC2: $F_{FC} = 30$, $f = \text{ReLU}$

TABLE VII
DETAILS OF THE BEST TRAINED CNNs FOR THE 1-D CNN-BASED FD FRAMEWORK

Network	W	Convolution layers	Fully connected layers
CNN-S2-A	17	Conv1: $k = [1,3]$, $F_c = 20$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 154$, $f = \text{ReLU}$ FC2: $F_{FC} = 30$, $f = \text{Leaky ReLU}$
CNN-S2-B	11	Conv1: $k = [1,5]$, $F_c = 18$, $f_D = 2$, $f = \text{ReLU}$	FC1: $F_{FC} = 74$, $f = \text{Leaky ReLU}$ FC2: $F_{FC} = 14$, $f = \text{ReLU}$
CNN-S3-C	19	Conv1: $k = [1,2]$, $F_c = 20$, $f_D = 2$, $f = \text{Leaky ReLU}$ Conv2: $k = [1,2]$, $F_c = 40$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 96$, $f = \text{Leaky ReLU}$
CNN-S2-D	15	Conv1: $k = [1,3]$, $F_c = 16$, $f_D = 2$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 156$, $f = \text{ReLU}$ FC2: $F_{FC} = 31$, $f = \text{ReLU}$
CNN-S2-E	13	Conv1: $k = [1,3]$, $F_c = 16$, $f_D = 1$, $f = \text{Leaky ReLU}$	FC1: $F_{FC} = 71$, $f = \text{Leaky ReLU}$ FC2: $F_{FC} = 14$, $f = \text{ReLU}$

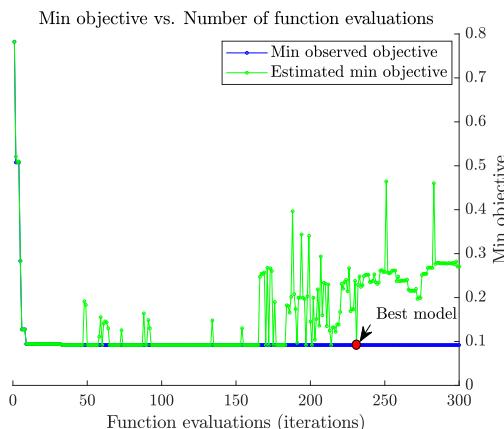


Fig. 8. Progress of the hyperparameter tuning of CNN 4 using the Bayesian optimization.

function using a probabilistic model for the hyperparameters mapping to a probability of a score on the actual objective function.

As the Bayesian optimization-based tuning process progresses, the estimated objective function converges to the actual one toward finding the best hyperparameters set that minimizes the model's cost function. It can be seen that the tuning process converged at around 150th and 240th

iterations with a final cost value of 0.42 and 0.1 for CNN 1 and CNN 4, respectively. Tables V–VII list the details of the developed CNNs in terms of the hyperparameters' values obtained from the Bayesian optimization-based tuning process.

V. EVALUATION AND DISCUSSION

This section presents the evaluation of the proposed framework in terms of the demonstration and discussion of the results obtained.

A. Evaluation Metrics

The confusion matrix is used to evaluate the performance of the CNN. It is a form of contingency table with two dimensions identified as True and Predicted and a set of classes in both dimensions, as presented in Table VIII. The following performance metrics are derived from the confusion matrix [38].

- 1) Accuracy (ACC): It is a measure of the closeness between the predicted result and the true value. Accuracy is given by

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (2)$$

TABLE VIII
TABLE OF CONFUSION FOR TWO-CLASS PROBLEM

		Predicted	
		Positive	Negative
True	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

2) Precision: It is also called positive predictive value (PPV), which is a measure of the closeness the set of predicted results, and it is expressed as

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (3)$$

3) Sensitivity: It is called true positive rate (TPR) or recall and is calculated by

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (4)$$

4) Specificity: It is also known as true negative rate (TNR) and is expressed as

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (5)$$

5) F₁-score: It is the harmonic average of the precision and recall, where it is at its best at a value of 1 meaning perfect precision and recall, and it is given by

$$\text{TNR} = 2 \times \frac{\text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}}. \quad (6)$$

B. Comparison Results

Tables IX and X present the performance of the best CNNs for the static and dynamic 2-D CNN-based FD models, respectively. The performance results of the CNNs represent the average performance of the seven trained models given that sevenfold cross-validation was used, in which the data set was divided equally into seven random subsets. Then, the CNN was trained seven times, and each time, one subset was used as the validation set and the remaining ones for training.

The first observation was that the sevenfold cross-validation evaluation results were high, meaning that the developed models generalize well on the training data set. In addition, the classification performances of both the static and dynamic 2-D CNN-based diagnosis models are akin that can be attributed to the fact that the system is slow and broadly exhibits a steady behavior. The classification accuracy for CNN 2 to CNN 5 is about 99% while it is around 95% for CNN 1 with around 3% false alarm rate and 5% missed alarm rate. The reason is that the data used to train CNN 1 are difficult as they contain all types of faults with their variants in the negative class. On the other hand, the static 2-D CNNs were more computationally efficient than the ones in the dynamic 2-D CNN-based diagnosis framework given the increased computational complexity of the latter approach resulted from accounting for the temporal features of the system's measurements.

It was found that CNN's structure 1 and structure 2, which contain a single Conv layer, were the ones used for most of the best-trained networks with a dilation factor of 1 and the number of feature mapping between five and six times the original number of input features. The filter size used in the static 2-D CNNs was consistently 2 × 2, while, for the

TABLE IX
PERFORMANCE RESULTS OF THE BEST FIVE CNNs AMONG ALL TRAINED NETWORKS FOR THE STATIC 2-D CNN-BASED FD FRAMEWORK

Network	Accuracy	Sensitivity	Specificity	Precision	F1-score
S-CNN-S1-A1	94.80%	95.95%	93.87%	92.60%	94.24%
S-CNN-S2-A2	94.73%	94.16%	95.19%	93.98%	94.07%
S-CNN-S1-A3	94.57%	95.31%	93.99%	92.68%	93.97%
S-CNN-S1-A4	94.27%	92.11%	96.00%	94.84%	93.46%
S-CNN-S2-A5	94.20%	94.69%	93.81%	92.43%	93.55%
S-CNN-S2-B1	99.92%	99.98%	99.40%	99.93%	99.95%
S-CNN-S1-B2	99.79%	99.93%	98.68%	99.84%	99.88%
S-CNN-S1-B3	99.76%	99.98%	97.99%	99.75%	99.86%
S-CNN-S1-B4	99.72%	99.93%	98.08%	99.76%	99.85%
S-CNN-S1-B5	99.63%	99.74%	98.68%	99.83%	99.79%
S-CNN-S1-C1	99.39%	99.84%	95.82%	99.48%	99.66%
S-CNN-S2-C2	99.35%	99.68%	96.70%	99.59%	99.63%
S-CNN-S2-C3	99.33%	99.86%	95.06%	99.39%	99.62%
S-CNN-S1-C4	99.28%	99.87%	94.59%	99.33%	99.60%
S-CNN-S2-C5	99.24%	99.86%	94.28%	99.29%	99.57%
S-CNN-S1-D1	99.64%	99.90%	97.52%	99.69%	99.80%
S-CNN-S3-D2	99.60%	99.96%	96.70%	99.59%	99.77%
S-CNN-S1-D3	99.59%	99.84%	97.61%	99.70%	99.77%
S-CNN-S2-D4	99.58%	99.90%	97.01%	99.63%	99.76%
S-CNN-S1-D5	99.56%	99.88%	96.98%	99.62%	99.75%
S-CNN-S3-E1	98.70%	98.54%	100.00%	100.00%	99.26%
S-CNN-S3-E2	98.66%	98.50%	100.00%	100.00%	99.24%
S-CNN-S3-E3	98.51%	98.33%	100.00%	100.00%	99.16%
S-CNN-S1-E4	98.25%	98.03%	100.00%	100.00%	99.00%
S-CNN-S3-E5	98.24%	98.03%	100.00%	100.00%	99.00%

TABLE X
PERFORMANCE RESULTS OF THE BEST FIVE CNNs AMONG ALL TRAINED NETWORKS FOR THE DYNAMIC 2-D CNN-BASED FD FRAMEWORK

Network	Accuracy	Sensitivity	Specificity	Precision	F1-score
D-CNN-S2-A1	96.12%	96.52%	95.79%	94.82%	95.66%
D-CNN-S2-A2	95.27%	94.47%	95.91%	94.85%	94.66%
D-CNN-S2-A3	95.06%	92.66%	96.98%	96.08%	94.34%
D-CNN-S2-A4	94.76%	93.80%	95.53%	94.37%	94.08%
D-CNN-S4-A5	94.75%	90.96%	97.77%	97.02%	93.89%
D-CNN-S2-B1	99.82%	99.89%	99.21%	99.90%	99.90%
D-CNN-S3-B2	99.81%	99.96%	98.58%	99.82%	99.89%
D-CNN-S2-B3	99.76%	99.99%	97.89%	99.74%	99.86%
D-CNN-S2-B4	99.58%	99.62%	99.31%	99.91%	99.76%
D-CNN-S2-B5	99.53%	99.90%	96.60%	99.58%	99.74%
D-CNN-S2-C1	99.38%	99.91%	95.12%	99.39%	99.65%
D-CNN-S2-C2	99.31%	99.92%	94.40%	99.31%	99.61%
D-CNN-S2-C3	99.29%	99.85%	94.84%	99.36%	99.60%
D-CNN-S2-C4	99.29%	99.76%	95.60%	99.45%	99.60%
D-CNN-S2-C5	99.26%	99.84%	94.62%	99.33%	99.59%
D-CNN-S2-D1	99.67%	99.90%	97.83%	99.73%	99.81%
D-CNN-S4-D2	99.66%	99.99%	97.04%	99.63%	99.81%
D-CNN-S3-D3	99.61%	99.88%	97.42%	99.68%	99.78%
D-CNN-S3-D4	99.57%	100.00%	96.10%	99.52%	99.76%
D-CNN-S2-D5	99.52%	99.85%	96.89%	99.61%	99.73%
D-CNN-S2-E1	98.80%	98.68%	99.72%	99.96%	99.32
D-CNN-S2-E2	98.64%	98.53%	99.50%	99.94%	99.23%
D-CNN-S3-E3	98.55%	98.43%	99.53%	99.94%	99.18%
D-CNN-S2-E4	98.46%	98.31%	99.65%	99.96%	99.13%
D-CNN-S2-E5	98.44%	98.27%	99.84%	99.98%	99.12%

dynamic 2-D CNN, it was mostly 2 × 4. The window size of at most $W = 9$ for the latter approach was convenient to achieve the observed network performance. In addition, the best performing CNNs were the ones, in which ReLU and Leaky ReLU functions were used for the layers' activations. That is, the latter activation functions are characterized by their fast and improved convergence performance in addition to their computational efficiency compared to the others, e.g., sigmoid and hyperbolic tangent.

The static 2-D CNN-based approach was compared with NN- and SVM-based approaches, as presented in Table XI, such that the proposed FD model was evaluated using NN- and SVM-based classifiers instead of the CNNs. It was found that the static 2-D CNN-based FD diagnosis scheme demonstrated an improved performance such that the average sensitivity of the latter approach is about 99%, while it is 95% for

TABLE XI
COMPARISON RESULTS BETWEEN USING NN, SVM, AND CNN
FOR THE PROPOSED FD FRAMEWORK

Classifier	Accuracy	Sensitivity	Specificity	Precision	F_1 -score
NN 1	90.10%	85.16%	94.04%	91.94%	88.42%
NN 2	98.23%	98.74%	94.18%	99.27%	99.00%
NN 3	98.38%	99.13%	92.36%	99.05%	99.09%
NN 4	98.16%	99.94%	83.93%	98.03%	98.98%
NN 5	92.98%	92.11%	99.94%	99.99%	95.89%
SVM 1	92.58%	98.32%	85.39%	89.40%	93.65%
SVM 2	97.94%	95.60%	98.23%	87.11%	91.15%
SVM 3	97.03%	94.31%	97.37%	81.74%	87.57%
SVM 4	98.67%	88.55%	99.93%	99.40%	93.66%
SVM 5	93.89%	98.46%	93.32%	64.78%	78.15%
S-CNN 1	94.80%	95.95%	93.87%	92.60%	94.24%
S-CNN 2	99.92%	99.98%	99.40%	99.93%	99.95%
S-CNN 3	99.39%	99.84%	95.82%	99.48%	99.66%
S-CNN 4	99.64%	99.90%	97.52%	99.69%	99.80%
S-CNN 5	98.70%	98.54%	100.00%	100.00%	99.26%

TABLE XII
ACTUATOR FAULT SCENARIOS USED FOR THE EVALUATION OF THE
PROPOSED 2-D CNN-BASED FD APPROACH

Fault identifier	Description
F-VAV1-High	Stuck damper at FO position at the beginning of daytime
F-VAV2-High	Stuck damper at FC position during night
F-VAV3-High	Stuck damper at FC position during night
F-Valve-High	Stuck valve at FC position during minimum thermal load
F-VAV1-Low	Stuck damper at PO at 18% opening
F-VAV2-Low	Stuck damper at PO at 56% opening

SVM- and NN-based approaches with an overall performance improvement of around 4%. The value of recall is inversely proportional to the missed alarm rate, and hence, using CNNs elevates the reliability of the FD. In addition, the average diagnosis accuracy is higher using CNN with about 99%, while it is 96% and 95% for SVM and NN, respectively. The average diagnosis precision of the CNN-based approach is 98% with an increase of 14% and 2% compared to SVM and NN, respectively.

C. Case Studies

The performance of the proposed approach was evaluated using a dedicated test data set that was generated for the six different actuator fault scenarios listed in Table XII, in which increased measurement noise was introduced to the data. The noise was emulated by an additive random uniformly distributed variable with a maximum level of 0.5 °C.

The performance evaluation was conducted on the proposed 2-D CNN-based FD framework using both the static and the dynamic schemes, and it was compared with the 1-D CNN-based diagnosis scheme. The evaluation results are summarized in Table XIII presenting the average performance of the five CNNs under each fault scenario. The evaluation was made mainly upon the diagnosis accuracy and the false alarm rate. It was based on an observation window of one day under the assumption that the fault source will be abolished within that time period. It was found that the accuracy of the CNNs was high proving that the networks did not overfit to the training data, and the developed models have adequate generalization ability.

1) *Case Study 1—High Severity Faults:* The first case study is evaluating the actuator faults with high severity level fault,

which are the scenarios F-VAV1-High to F-Valve-High. For instance, the VAV damper was stuck at a fully closed position at night during which the load was minimal causing the HVAC system to fail to meet the cooling load in the daytime and vice versa. It was found that the static 2-D CNN-based diagnosis scheme is more robust and reliable with an average diagnosis accuracy of 99.06% and an average false alarm rate of 0.31%. For the dynamic 2-D CNN-based scheme and the 1-D CNN-based approach, the average diagnosis accuracies are lower as 96.48% and 94.89%, respectively, with an increase in the average false alarm rate of 2.64% and 4.32%, respectively, compared with the static 2-D CNN-based scheme.

It was observed that the dynamic 2-D CNN-based and the 1-D CNN-based approaches are more prone to noise since they rely on extracting the temporal features in the time-series signals. The increased noise level altered the CNNs' capability in accurate diagnosis resulting in reduced diagnosis accuracy and increased false alarm rate. The performance of the three CNN-based approaches in terms of the missed alarm rate and the detection time was found similar and acceptable. Figs. 9–12 demonstrate the diagnosis of the high impact faults using the static 2-D CNN-based scheme, which was quick and accurate for the four severe fault scenarios and akin to the dynamic 2-D CNN-based and the 1-D CNN-based approaches, as presented in Table XIII. It is worth noting that the plots of CNN outputs are downsampled for demonstration purposes.

2) *Case Study 2—Low Severity Faults:* The second case study presents low impact fault events, in which the actuators were stuck at a partially open position that did not severely impede the HVAC system operation in fulfilling the thermal load requirement. The 2-D CNN-based schemes outperformed the 1-D CNN-based diagnosis approach. The average diagnosis accuracy of the two 2-D CNN-based diagnosis schemes was found to be 89.21% with an average false alarm rate of 1.49%, while, for the 1-D CNN-based diagnosis approach, they were 83.38% and 7.25%, respectively.

Figs. 13 and 14 represent the diagnosis results of fault incidents with low impact level using the static 2-D CNN-based approach. The damper in each of the two scenarios was stuck at a partially open position. For the F-VAV1-Low fault, the damper was stuck at 18% open position shortly after the HVAC system started the daytime operation. As demonstrated in Fig. 13, the diagnosis framework was able to identify the fault within half an hour for the subsequent 8 h, while, for the F-VAV2-Low fault, the fault was diagnosed after 14 h, which was at the start of the night operation, since the 50% faulty damper position did not alter the HVAC system operation during the daytime. The missed alarm rate for the low-severity faults was higher than it was for the ones of high severity given the fact that for the former fault scenarios, the faulty actuator's opening was in occasions sufficient to accommodate the thermal load and, hence demonstrated reduced impact on the system's variables.

D. Computational Complexity

A simplified computational complexity analysis was carried out for comparing the three CNN-based diagnosis approaches, as illustrated in Fig. 15. The figure shows the average

TABLE XIII

EVALUATION RESULTS OF THE PROPOSED CNN-BASED FD FRAMEWORK ON THE ACTUATOR FAULT SCENARIOS—CASE STUDIES

Fault identifier	FD Approach	Diagnosis accuracy	False alarm rate	Missed alarm rate	Diagnosis time
F-VAV1-High	Static 2D CNN	98.21%	0.90%	0.89%	66 min
	Dynamic 2D CNN	98.21%	0.90%	0.89%	63 min
	1D CNN	98.32%	0.77%	0.92%	61 min
F-VAV2-High	Static 2D CNN	98.57%	0.25%	1.18%	68 min
	Dynamic 2D CNN	91.07%	8.02%	0.91%	65 min
	1D CNN	95.61%	3.68%	0.71%	54 min
F-VAV3-High	Static 2D CNN	99.92%	0.07%	0.00%	0 min
	Dynamic 2D CNN	97.94%	2.06%	0.00%	0 min
	1D CNN	87.52%	12.35%	0.31%	10 min
F-Valve-High	Static 2D CNN	99.54%	0.01%	0.45%	20 min
	Dynamic 2D CNN	98.72%	0.82%	0.46%	35 min
	1D CNN	98.12%	1.69%	0.19%	14 min
F-VAV1-Low	Static 2D CNN	94.93%	0.05%	5.02%	34 min
	Dynamic 2D CNN	91.71%	3.61%	4.67%	29 min
	1D CNN	83.88%	10.88%	5.24%	66 min
F-VAV2-Low	Static 2D CNN	86.43%	1.11%	12.46%	14 hours
	Dynamic 2D CNN	83.76%	1.17%	15.08%	15 hours
	1D CNN	82.87%	3.63%	13.50%	12 hours

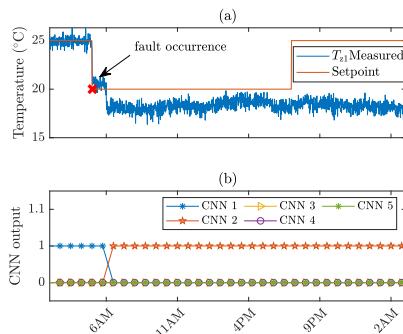


Fig. 9. Performance of the static 2-D CNN-based diagnosis method due to stuck F-VAV1-high fault at 100% open position at the beginning of the day. (a) Zone 1 temperature. (b) CNN outputs.

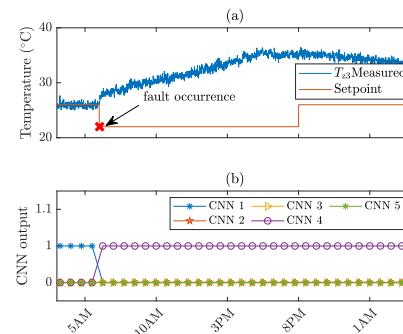


Fig. 11. Performance of the static 2-D CNN-based diagnosis method due to stuck F-VAV3-high fault at 100% closed position at night. (a) Zone 3 temperature. (b) CNN outputs.

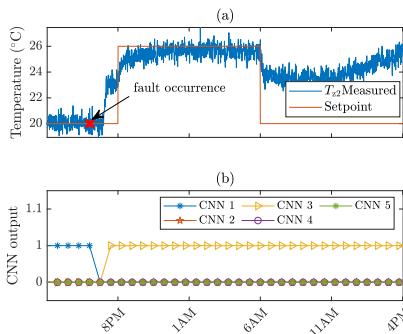


Fig. 10. Performance of the static 2-D CNN-based diagnosis method due to stuck F-VAV2-high fault at 100% closed position at night. (a) Zone 2 temperature. (b) CNN outputs.

number of learnable parameters of the five CNNs, which roughly reflects the space and time complexity of the model against its diagnosis performance represented by the average F1-score value for the static 2-D CNN-based, the dynamic 2-D CNN-based, and the 1-D CNN-based detection approaches for the optimized CNNs in the FD model. It was observed that the dynamic 2-D CNN-based approach is the highest in the computational complexity with an average number of learnable parameters of 140348, while the static 2-D CNN-based is the most efficient in terms of the models' computational requirement. Even though the 1-D CNN-based and the dynamic 2-D CNN-based schemes have higher average F1-score values of around 98.8%, the improvement in the diagnosis performance

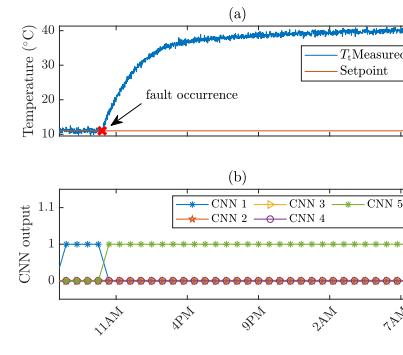


Fig. 12. Performance of the static 2-D CNN-based diagnosis method due to stuck F-valve-high fault at 100% closed position. (a) Tank temperature. (b) CNN outputs.

is as minor as 0.3% against an increase in the computational complexity of 1.4% and 8.9%, respectively, compared to the static 2-D CNN-based approach.

E. Summary

A summary of the main characteristics of the three CNN-based schemes is presented in Table XIV. The 1-D CNN-based and the dynamic 2-D CNN-based schemes account for the effects of the temporal features of the time-series signals on the diagnosis performance. As summarized in Table XIII, it can be observed that the performance of the 2-D CNN-based diagnosis framework is superior to the one of the 1-D CNN as the latter is more focused on

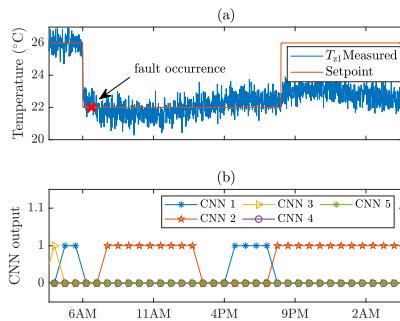


Fig. 13. Performance of the static 2-D CNN-based diagnosis method due to stuck F-VAV1-low fault at 18% open position. (a) Zone 1 temperature. (b) CNN outputs.

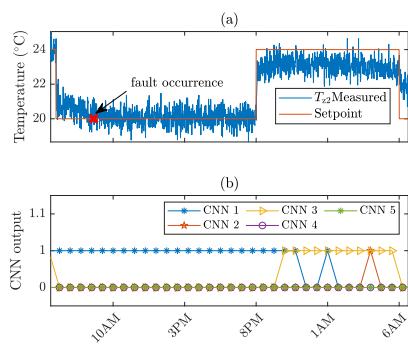


Fig. 14. Performance of the static 2-D CNN-based diagnosis method due to stuck F-VAV2-low fault at 56% open position. (a) Zone 2 temperature. (b) CNN outputs.

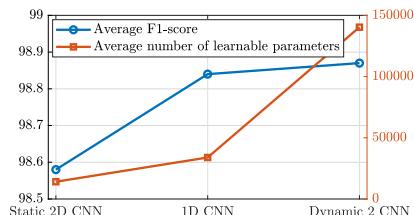


Fig. 15. Comparison between the different CNN-based FD schemes in terms of the diagnosis performance and the computation complexity.

TABLE XIV
SUMMARY OF THE MAIN CHARACTERISTICS OF THE DIFFERENT CNN-BASED DIAGNOSIS SCHEMES

Scheme	Temporal features	Spatial Correlation	Limitations
1D CNN	Yes	Yes	Focused on the independent features of the time-series signals
Dynamic 2D CNN	Yes	Yes	High computational complexity
Static 2D CNN	No	Yes	Temporal features of the time-series signals are not considered

the independent features of the time-series signals, i.e., the temperature measurements and the control signals.

That is, the different variables of the systems are correlated as the case of the temperature of the zones (T_{zi}) that are tied to the temperature of the AHU supply air (T_{ao}), which in turn, affects the AHU return water temperature (T_{wo}), and so on. The use of the 2-D CNN enables exploiting spatially the local correlation between the system's variables that were made possible in the static and the dynamic 2-D CNN-based

diagnosis schemes by the proposed 1-D-to-2-D data transformation technique. In addition, the best diagnosis performance was achieved by the static 2-D CNN-based framework given that the dynamical behavior of the HVAC system is slow and broadly exhibits a steady behavior.

VI. CONCLUSION

In this work, the common actuator faults affecting the HVAC systems were investigated by means of the TRNSYS simulation tool. A 2-D CNN-based actuator FD framework was proposed for single stuck damper and valve faults in the HVAC system based on an efficient 1-D-to-2-D data transformation approach. The performance of the CNNs was ensured by an optimal tuning of its significant hyperparameters using the Bayesian optimization algorithm toward maximizing the classification accuracy.

Two schemes of the proposed framework were exploited, which are the static scheme and the dynamic scheme, in which the latter accounts for the temporal features of the system's variables. Extensive simulations demonstrated the effective performance of the proposed 2-D CNN-based HVAC actuator FD for faults of different severity levels that were also compared with the 1-D CNN-based approach. The diagnosis performance of the two schemes was found similar, but the static 2-D CNN-based diagnosis scheme was more computationally efficient.

In addition, the CNN-based approach was compared with SVM and the standard neural network, and the proposed method demonstrated improved performance in terms of the diagnosis accuracy, precision, and recall. The use of the CNN for the HVAC system actuator FD was proven to be successful in improving the accuracy, reliability, and precision of the HVAC system actuator FD by an average of 4%, 11%, and 10%, respectively, compared to the standard neural networks and SVM, which are commonly used algorithms.

The proposed FD framework is a data-driven approach that is independent of the building technical details and does not utilize the knowledge of the system mathematical model. It can be applied to more complex buildings with more and different types of zones while obtaining the same performance in terms of accuracy and reliability, which is conditioned that sufficient amount and adequate quality of data are available for training. For large-scale buildings, it is suggested to use a multiagent FD solution to compensate for the increased computational requirement of the algorithm with respect to the building size.

Further work in this subject includes the following:

- 1) considering multiple simultaneous actuator faults;
- 2) developing a data-driven FD strategy for simultaneous sensor and actuator FD;
- 3) conducting energy performance assessment of the HVAC system to demonstrate the diagnosis approach capability in improving the system efficiency.

REFERENCES

- [1] *Energy Technology Perspectives* 2017. [Online]. Available: <https://www.iea.org/etp2017>

- [2] P. Wang and R. X. Gao, "Automated performance tracking for heat exchangers in HVAC," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 634–645, Apr. 2017.
- [3] M. Bonvini, M. D. Sohn, J. Granderson, M. Wetter, and M. A. Piette, "Robust on-line fault detection diagnosis for HVAC components based on nonlinear state estimation techniques," *Appl. Energy*, vol. 124, pp. 156–166, Jul. 2014.
- [4] P. M. Papadopoulos, V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Distributed diagnosis of actuator and sensor faults in HVAC systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 4209–4215, 2017.
- [5] A. Beghi, L. Cecchinato, L. Corso, M. Rampazzo, and F. Simmini, "Process history-based fault detection and diagnosis for VAVAC systems," in *Proc. IEEE Int. Conf. Control Appl. (CCA)*, Aug. 2013, pp. 1165–1170.
- [6] S. Li and J. Wen, "Application of pattern matching method for detecting faults in air handling unit system," *Autom. Construct.*, vol. 43, pp. 49–58, Jul. 2014.
- [7] Z. Du, B. Fan, X. Jin, and J. Chi, "Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis," *Building Environ.*, vol. 73, pp. 1–11, Mar. 2014.
- [8] K.-P. Lee, B.-H. Wu, and S.-L. Peng, "Deep-learning-based fault detection and diagnosis of air-handling units," *Building Environ.*, vol. 157, pp. 24–33, Jun. 2019.
- [9] M. Najafi, D. M. Auslander, P. L. Bartlett, P. Haves, and M. D. Sohn, "Application of machine learning in the fault diagnostics of air handling units," *Appl. Energy*, vol. 96, pp. 347–358, Aug. 2012.
- [10] D. Li, Y. Zhou, G. Hu, and C. J. Spanos, "Handling incomplete sensor measurements in fault detection and diagnosis for building HVAC systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 833–846, Apr. 2019.
- [11] H. Wang and Y. Chen, "A robust fault detection and diagnosis strategy for multiple faults of VAV air handling units," *Energy Buildings*, vol. 127, pp. 442–451, Sep. 2016.
- [12] W. H. Allen, A. Rubaai, and R. Chawla, "Fuzzy neural network-based health monitoring for HVAC system variable-air-volume unit," *IEEE Trans. Ind. Appl.*, vol. 52, no. 3, pp. 2513–2524, May 2016.
- [13] Y. Zhao, J. Wen, and S. Wang, "Diagnostic Bayesian networks for diagnosing air handling units faults—Part II: Faults in coils and sensors," *Appl. Thermal Eng.*, vol. 90, pp. 145–157, Nov. 2015.
- [14] K. Verbert, R. BabuÅlka, and B. D. Schutter, "Combining knowledge and historical data for system-level fault diagnosis of HVAC systems," *Eng. Appl. Artif. Intell.*, vol. 59, pp. 260–273, 2017.
- [15] D. Li, Y. Zhou, G. Hu, and C. J. Spanos, "Identifying unseen faults for smart buildings by incorporating expert knowledge with data," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 3, pp. 1412–1425, Jul. 2019.
- [16] Y. Yan, P. B. Luh, and K. R. Pattipati, "Fault diagnosis of HVAC air-handling systems considering fault propagation impacts among components," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 705–717, Apr. 2017.
- [17] M. Elnour and N. Meskin, "Actuator fault diagnosis in multi-zone HVAC systems using 2D convolutional neural networks," in *Proc. IEEE Int. Conf. Informat. IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 404–409.
- [18] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
- [19] O. Abdeljaber, O. Avci, M. S. Kiranyaz, B. Boashash, H. Sodano, and D. J. Inman, "1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data," *Neurocomputing*, vol. 275, pp. 1308–1317, Jan. 2018.
- [20] S. Kiranyaz, A. Gastli, L. Ben-Brahim, N. Al-Emadi, and M. Gabbouj, "Real-time fault detection and identification for MMC using 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 11, pp. 8760–8771, Nov. 2019.
- [21] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.
- [22] Y. Fu *et al.*, "Machining vibration states monitoring based on image representation using convolutional neural networks," *Eng. Appl. Artif. Intell.*, vol. 65, pp. 240–251, Oct. 2017.
- [23] K. B. Lee, S. Cheon, and C. O. Kim, "A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 2, pp. 135–142, May 2017.
- [24] O. Janssens *et al.*, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, Sep. 2016.
- [25] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, Jan. 2019.
- [26] S. Guo, T. Yang, W. Gao, and C. Zhang, "A novel fault diagnosis method for rotating machinery based on a convolutional neural network," *Sensors*, vol. 18, no. 5, p. 1429, May 2018.
- [27] V. Tra, J. Kim, S. A. Khan, and J.-M. Kim, "Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal Levenberg-Marquardt algorithm," *Sensors*, vol. 17, no. 12, p. 2834, Dec. 2017.
- [28] R. Chen, X. Huang, L. Yang, X. Xu, X. Zhang, and Y. Zhang, "Intelligent fault diagnosis method of planetary gearboxes based on convolution neural network and discrete wavelet transform," *Comput. Ind.*, vol. 106, pp. 48–59, Apr. 2019.
- [29] A. Gupta, G. Gurrala, and P. S. Sastry, "An online power system stability monitoring system using convolutional neural networks," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 864–872, Mar. 2019.
- [30] S. Hamdaoui, M. Hamdaoui, A. Allouhi, R. El Alaiji, T. Kousksou, and A. El Bouardi, "Energy demand and environmental impact of various construction scenarios of an office building in morocco," *J. Cleaner Prod.*, vol. 188, pp. 113–124, Jul. 2018.
- [31] S. A. Klein *et al.*, "TRNSYS 18: A transient system simulation program," Solar Energy Lab., Univ. Wisconsin, Madison, WI, USA, 2017. [Online]. Available: <http://sel.me.wisc.edu/trnsys>
- [32] P. M. Van Every, M. Rodriguez, C. B. Jones, A. A. Mammioli, and M. Martínez-Ramón, "Advanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models," *Energy Buildings*, vol. 149, pp. 216–224, Aug. 2017.
- [33] R. Yan, Z. Ma, G. Kokogiannakis, and Y. Zhao, "A sensor fault detection strategy for air handling units using cluster analysis," *Autom. Construct.*, vol. 70, pp. 77–88, Oct. 2016.
- [34] Z. Du, B. Fan, J. Chi, and X. Jin, "Sensor fault detection and its efficiency analysis in air handling unit using the combined neural networks," *Energy Buildings*, vol. 72, pp. 157–166, Apr. 2014.
- [35] S. Wang, Q. Zhou, and F. Xiao, "A system-level fault detection and diagnosis strategy for HVAC systems involving sensor faults," *Energy Buildings*, vol. 42, no. 4, pp. 477–490, Apr. 2010.
- [36] L. Chang, H. Wang, and L. Wang, "Fault detection and diagnosis of an HVAC system using artificial immune recognition system," in *Proc. IEEE PES Asia-Pacific Power Energy Eng. Conf. (APPEEC)*, Dec. 2013, pp. 1–5.
- [37] H.-I. Suk, *Deep Learning for Medical Image Analysis*, S. K. Zhou, H. Greenspan, and D. Shen, Eds. New York, NY, USA: Academic, 2017.
- [38] K. M. Ting, *Confusion Matrix*. Boston, MA, USA: Springer, 2010, p. 209.



Mariam Elnour received the B.Sc. and M.Sc. degrees from Qatar University, Doha, Qatar, in 2015 and 2019, respectively.

She was a Graduate Assistant at Qatar University from September 2016 to January 2018, where she is currently a Research Assistant. Her research interests include machine learning, signal processing, and fault diagnosis.



Nader Meskin (Senior Member, IEEE) received the B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 1998, the M.Sc. degree from the University of Tehran, Tehran, in 2001, and the Ph.D. degree in electrical and computer engineering from Concordia University, Montreal, QC, Canada, in 2008.

He was a Post-Doctoral Fellow at Texas A&M University at Qatar, Doha, Qatar, from January 2010 to December 2010. He is currently an Associate Professor at Qatar University, Doha, and an Adjunct Associate Professor at Concordia University. He has published more than 220 refereed journal and conference papers. His research interests include FDI, multiagent systems, active control for clinical pharmacology, cybersecurity of industrial control systems, and linear parameter varying systems.