

## RESEARCH ARTICLE

# A Reliable Deep Learning Approach for Time-Varying Faults Identification: Spacecraft Reaction Wheel Case Study

ABD-ELSALAM R. ABD-ELHAY<sup>1</sup>, WAEL A. MURTADA<sup>2</sup>, AND MOHAMED I. YOUSSEF<sup>3</sup><sup>1</sup>Egyptian Space Agency (EGSA), Cairo 130, Egypt<sup>2</sup>National Authority for Remote Sensing and Space Sciences (NARSS), Cairo 1564, Egypt<sup>3</sup>Department of Electrical Engineering, Faculty of Engineering, Al-Azhar University, Cairo 11884, Egypt


Corresponding author: Abd-Elsalam R. Abd-Elhay (abdelsalam.hay@egsa.gov.eg)

**ABSTRACT** Reaction wheels are key components for the spacecraft attitude control subsystems. Faults in reaction wheels may lead to high energy consumption, lack of spacecraft attitude control, and in case of failure, loss of the spacecraft. The accurate identification of reaction wheels anomalies is a challenging task due to the internal nonlinearities of the reaction wheels. This study proposes a fast and accurate end-to-end architecture for detecting and identifying the anomalies occurring in spacecraft reaction wheels using One-Dimensional Convolutional Neural Network (1D-CNN) with Long Short-Term Memory (LSTM) network architecture. 1D-CNN is used to capture the useful features from the raw residual signals. The Long Short-Term Memory layer is used due to its effectiveness in handling the time series data and its capabilities for learning long-term dependencies. The proposed architecture is directly trained using the raw torque residual signals captured from a 3-axis attitude control subsystem simulation model. In this way, this scheme eliminates the need for a specific feature extraction method. Results showed that the proposed algorithm represents a reliable and robust anomaly detection and identification mechanism with compact system architecture. Furthermore, the obtained results revealed the superiority and generalizability of the proposed model in diagnosing time-varying reaction wheel faults over other recent approaches. Ultimately, the proposed approach is considered to be a generic fault diagnosis architecture for safety-critical systems. The dataset is available for download at: <https://dx.doi.org/10.21227/jr1c-bm66>

**INDEX TERMS** Reaction wheel, fault detection and identification (FDI), fault diagnosis, 1D-CNN, long short-term memory (LSTM), spacecraft attitude control.

## I. INTRODUCTION

Spacecraft is a complex system that operates in a relentlessly hostile environment. The status of the Attitude Control Subsystem (ACS) has a great impact on the normal operation of the spacecraft. Moreover, the three-axis-stabilized spacecraft requires high pointing accuracy during imaging and commonly uses reaction wheels for attitude control. Consequently, reaction wheels' status has a great effect on the performance of the spacecraft. Furthermore, recently reported failures in satellite attitude control subsystem [1]–[3],

The associate editor coordinating the review of this manuscript and approving it for publication was Mark Kok Yew Ng .

particularly reaction wheels ensure the necessity of a proper fault diagnosis system onboard the spacecraft to avoid these failures. For instance, In July 2012, NASA announced the failure of Kepler Reaction Wheel Assembly 1 (RWA1) and Reaction Wheel Assembly 2 (RWA2) [4]. In September 2007, NASA's Dawn spacecraft was launched for analyzing two of the three known protoplanets of the asteroid belt: Vesta and Ceres. In April 2017, While preparing the spacecraft to observe Ceres, the spacecraft's remaining reaction wheels stopped working. Meanwhile, the spacecraft transferred to safe mode, and the orientation control was assigned to the thrusters. Consequently, it is necessary to detect and identify any kinds of reaction wheel anomalies as early as possible and

implement fault-tolerant operations for minimizing performance degradation and averting critical situations. Therefore, this research investigates the problem of fault diagnosis in spacecraft reaction wheels.

Generally, Fault Detection and Identification (FDI) approaches can be classified into model-based FDI, signal-based FDI, and knowledge-based FDI techniques [5], [6]. Model-based fault diagnosis methods were firstly proposed by Bred in 1971 to replace the traditional hardware redundancy. The main idea of the model-based FDI methods is to replace the hardware redundancy using either physical principles or system identification methods. Therefore, the fault diagnosis process is developed by monitoring the consistency between the model estimated output and the measured output [7]. The major advantage of the model-based fault diagnosis techniques is the simplicity of the fault diagnosis process. However, they need high-fidelity models that are practically unavoidable. Since it is difficult to distinguish anomalies from errors in the model, model-based fault diagnosis approaches can lack robustness [8]. On the other hand, signal-based diagnosis approaches depend on the processing of the measured signals and don't require an explicit input-output model. Thus, in signal-based diagnosis, features of measured signals are captured and diagnosis is based on the symptom analysis and prior knowledge of the symptoms of the healthy systems [9]. The major disadvantage of the signal-based diagnosis approaches is that their performance is highly dependent on the quality and quantity of the process data. Furthermore, Knowledge-based FDI techniques can be categorized into qualitative approaches and quantitative approaches. Qualitative approaches include all techniques that attempt to check for a fault without placing a value on the fault level [9]. Examples of qualitative Knowledge-based approaches include fault matrix, fault trees, diagraphs, and expert systems. Whereas quantitative methods implement the diagnosis process as a pattern recognition problem. In such methods, features can be extracted either by using statistical methods such as Principal Component Analysis (PCA), Partially Least Squares (PLS), Independent Component Analysis (ICA), or non-statistical methods such as Artificial Neural Networks (ANN). The performance of knowledge-based methods depends on the training dataset and the quality of the selected features heavily [10].

In recent years, many studies were proposed for spacecraft Reaction Wheel (RW) fault diagnosis [11]–[29]. For instance, in [11] E. Sobhani *et al.* utilized a group of neural parameter estimators for nonlinear fault detection, isolation, and identification with application to spacecraft reaction wheel. Two reaction wheel faults have been considered in this work, namely the bus voltage and the motor current faults. The main drawbacks of the work in [11] are the extensive computational resources and the sensitivity to the noise due to the use of Kalman filters for states estimation. The authors in [18] proposed a data-driven fault detection method that combines clustering and PCA mechanism. Meanwhile, the clustering was utilized as feature recognition, while PCA was utilized to

establish the relationship of the parameters in each process. The results have shown that the proposed approach has a better performance compared to the traditional PCA model. However, the authors considered only two reaction wheel faults that are the motor friction and the motor magnetic flux faults. Moreover, the proposed approach failed to detect minor motor friction faults. In [21] Chen *et al.* suggested two stages extended Kalman filters for reaction wheel fault estimation. The simulation results demonstrated the effectiveness of the proposed fault estimation methodology. The main issue with the work in [21] is that the use of two cascaded Kalman filters can consume more computational resources. Moreover, the authors didn't provide the accuracy and the architecture of the proposed methodology. Furthermore, the authors in [25] have investigated fault diagnosis for reaction wheels using a feedforward neural network classifier. The authors have proposed the Prony method for feature extraction to distinguish between normal and anomalies states of the reaction wheel. The proposed approach has shown high accuracy in reaction wheel faults diagnosis. But, the main limitation with the proposed approach in [25] is that the Prony method is very sensitive to noise. To address this concern, Wael A. *et al.* [26] utilized the Short Time Energy (STE) with Prony method for extracting discriminative features from noisy fault signals in spacecraft solar panels. The obtained results showed the robustness of the proposed technique. However, the use of a separate feature extraction technique is a computationally expensive operation that may prevent the usage of the aforementioned approach in real-time applications. In a recent effort, Rahimi *et al.* [28] introduced a hierarchical integrated solution to detect, isolate, and identify satellite RW faults. The detection of reaction wheel anomalies is accomplished by monitoring the residual signals with an adaptive threshold technique. To isolate the faults, the authors have proposed online likelihood distribution calculation and Bayes' law. Moreover, the identification is performed using a dual unscented Kalman filter approach for states and parameters estimation. Results have proved that the introduced FDI algorithm can diagnose RW faults with an accuracy of 89.5 % on average. The drawback of this research in [28] is that the result accuracy of 89.5 % can't be accepted in critical components such as reaction wheels. Another limitation in [28] is the complexity of the proposed method. One of the main shortcomings in the former works is the need for a time-consuming feature extraction technique.

Recently, many deep learning approaches have shown superior identification accuracies in fault diagnosis for different applications. These approaches include Deep Belief Networks (DBN) [30]–[33], deep autoencoders [34]–[35], and Generative Adversarial Networks (GAN) [36]–[38]. One of the main drawbacks of the former deep learning approaches is that they require a large number of trainable parameters when they are trained using time-domain raw signals [39]. This drawback would prevent using these approaches in real-time applications such as spacecraft onboard computers due to their limited resources. On the other hand, convolutional

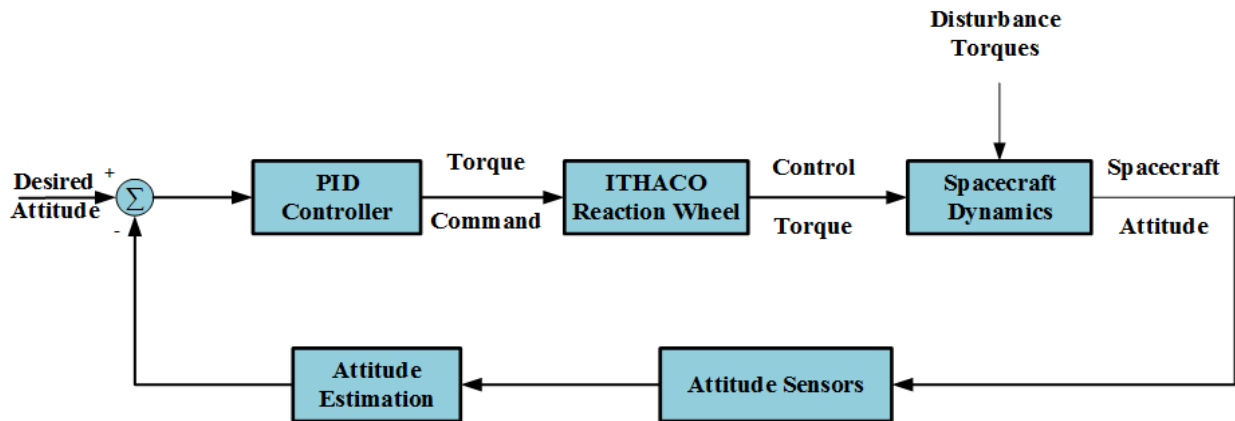


FIGURE 1. Three-axis attitude control subsystem block diagram.

neural networks can reduce the number of the networks' trainable parameters due to parameter sharing and local connectivity strategies [40]. Consequently, in this work, it is appropriate to address these limitations and drawbacks using an end-to-end FDI architecture based on 1D-CNN with LSTM network for the spacecraft reaction wheel. Convolutional Neural Networks (CNNs) were first utilized in pattern recognition applications. Furthermore, CNNs are the most commonly used deep learning methods due to their optimal configuration and their ability to extract discriminative features from raw signals [41]. Besides, LSTM is a variant of Recurrent Neural Networks (RNN) that was proposed to overcome the problem of vanishing gradient in RNN. Moreover, LSTM networks have been widely used to analyze the patterns in long sequence data [42]. Therefore, by combining the 1D-CNN and LSTM networks in one architecture for reaction wheel fault diagnosis, our contribution can be summarized as follows:

- This research proposes a superior deep learning architecture for the spacecraft reaction wheel fault diagnosis that fuses the feature-extraction and anomalies classification phases in one process. This eliminates the need for time-consuming feature extraction methods.
- The proposed 1D-CNN with LSTM fault diagnosis structure can learn long-term dependencies from the reaction wheel residual time-series signals. Therefore, allow the diagnosis of time-varying reaction wheel faults with simple architecture and superior performance.
- The research proposes a compact deep learning architecture for the spacecraft reaction wheel fault diagnosis with a small memory footprint. This makes it convenient for safety-critical real-time applications in general and specifically for the spacecraft embedded onboard computer fault diagnosis tasks.
- With directly learning the discriminative features from the raw signals, the proposed fault diagnosis approach can be adapted to be applicable to other spacecraft subsystems as well as different industrial systems.

- Validating the generalizability of the proposed model utilizing various testing datasets with different tilting angles and different time-varying conditions that represent the worst-case operation scenarios.
- Evaluating the robustness of the proposed FDI approach using noisy datasets with different noise levels.

This research is organized as follows: In the next section, the spacecraft attitude dynamic model is given. Subsequently, section III presents a brief review of CNN and LSTM networks. Section IV introduces the proposed architecture for the spacecraft reaction wheel fault diagnosis. Experimental results are introduced and evaluated in section V. Ultimately, section VI summarizes the research.

## II. SPACECRAFT ATTITUDE DYNAMICS MODEL

To evaluate the performance of the proposed fault diagnosis method, a three-axis stabilized spacecraft with reaction wheel as an actuator is considered. The attitude dynamics of spacecraft stabilized by reaction wheels are represented by Euler equation as follows [43]:

$$I_s \dot{\omega} = -\omega \times (I_s \omega + I_w \omega_w) - \tau_w \quad (1)$$

where,  $I_s$  represents the spacecraft moment of inertia,  $I_w$  denotes the reaction wheels moment of inertia,  $\omega$  introduces the spacecraft angular velocity vector  $[\omega_x, \omega_y, \omega_z]$ ,  $\tau_w$  is the reaction wheel torque, and  $\omega_w$  is the reaction wheel angular velocity. To validate the effectiveness of the proposed FDI method, an accurate three-axis attitude control simulation model has been developed. ACS model includes the spacecraft attitude dynamics, Proportional Integral Derivative (PID) controller, and a nonlinear RW model as shown in Figure 1.

Reaction wheel is a common choice for active satellite attitude control, particularly with unmanned spacecrafts. It is simply a flywheel driven by a Direct Current (DC) motor. Besides, it is used to provide the satellite with the required torque during attitude maneuvers. Moreover, to perform a maneuver using reaction wheels, each wheel is accelerated

TABLE 1. Reaction wheel parameters.

Parameter	Description	Value
$w_s$	Over-speed Circuit Gain	690 rad/sec
$K_t$	Motor Torque Constant	0.029 N.m/A
$K_e$	Motor Back-EMF	0.029 V/rad/sec
$K_s$	Over-speed Circuit Gain	95 V/ rad/sec
$w_d$	Driver Bandwidth	2000 rad/sec
$K_f$	Voltage Feedback Gain	0.5 V/V
N	Number of Poles	36
$T_c$	Torque command range	-5 to +5 V
$\tau_c$	Coulomb Friction	0.002 N.m
J	Inertia of the flywheel	0.0077 N.m.s <sup>2</sup>
$V_{bus}$	Bus Voltage	8 V
$\tau_v$	Viscous Friction	$3.84 \times 10^{-4}$ Nm/rad/s

in one direction. Therefore, the satellite will be accelerated in the opposite direction [44]. The proposed reaction wheel in this research is ‘ITHACO type A’ and it is produced by Goodrich Corporation. Thus, the proposed high-fidelity mathematical model by Biake [45], has been integrated with the ACS model as shown in Figure 1. To get high precision, there are important loops that should be integrated with the reaction wheel model as illustrated in Figure 2. More details about these loops can be found in [46]. The main parameters of the ‘ITHACO type A’ reaction wheel are listed in TABLE 1.

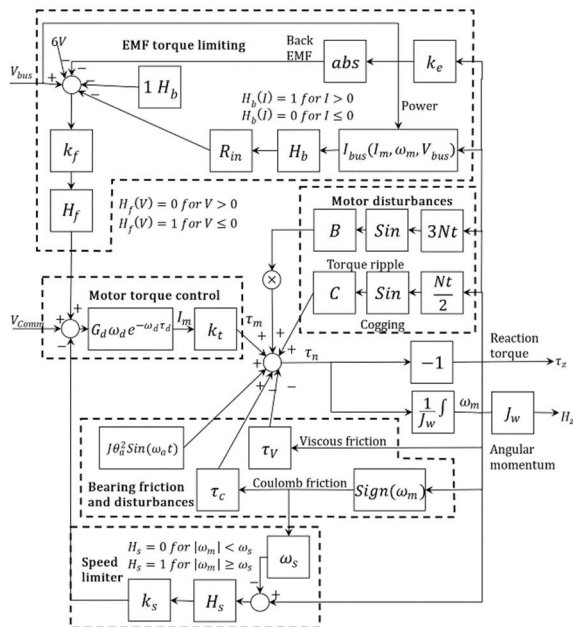


FIGURE 2. RW mathematical model [46].

Due to the internal nonlinearities of the reaction wheels and their complex design, they are vulnerable to faults and failures. Moreover, the most commonly reported faults in

reaction wheels are caused by abnormalities in the power supply line voltages, sudden variations in motor current, and the increase in the bearings’ friction [11]. Therefore, the considered faults in this research are the bus voltage ( $V_{bus}$ ) faults, the faults caused by variations in the motor current ( $I_m$ ), and fault due to increase in the bearing’s friction.

### III. THEORETICAL BACKGROUND

Artificial Neural Network (ANN) is an abstract model of the animal nervous system that can approximate any function from input-output pairs from the function [47]. Furthermore, ANN is commonly used in applications where the functions of interest are difficult to be evaluated directly. In the reaction wheel fault diagnosis task, ANN is utilized for modeling the complex relation between the anomalies and the underlying faults in reaction wheels. Moreover, ANN will train this relation implicitly from dataset samples generated by the ACS simulator. This research proposes a superior end-to-end fault diagnosis methodology for spacecraft reaction wheels using 1D-CNN and LSTM neural networks. Therefore, this section represents a brief overview for the CNN and LSTM network.

#### A. CONVOLUTIONAL NEURAL NETWORKS

Convolutional neural networks were primarily proposed by Fukushima and Miyake [48] for solving image recognition problems. Moreover, CNN is a biologically inspired feed-forward artificial neural network that introduces a simplified model for the mammalian visual cortex. Due to their superior performance, CNNs had become a popular solution for computer vision problems such as image classification and object tracking. Furthermore, CNNs can amalgamate the feature extraction and classification tasks in one learnable block. Therefore, they can be trained to extract the optimal features directly from raw signals to enhance classification accuracy. Generally speaking, CNN is a variant of Multi-Layer Neural Network (MLNN), but the difference is that the convolutional layer has a set of filters that are known as kernels. Figure 3 shows the basic configuration of 2D-CNN for the image classification task. Also, CNN consists of two convolutional pooling layer pairs and a single fully connected layer followed by the output layer. In the convolutional layer, the convolution filters execute convolutions on the inputs to compose the feature output. Thus, the convolutional layer output can be represented as follows [48]:

$$y_j = f(X * W_j + b_j) \tag{2}$$

In equation (2), the \* operator is used for the convolution of the input data matrix  $X$  and the weight matrix  $W_j$  of the  $j - th$  kernel to get the feature map  $y_j$ ;  $b_j$  represents the  $j - th$  bias;  $f$  is a nonlinear activation function that is applied to the result. meanwhile,  $f$  may be hyperbolic tangent, sigmoid, or Rectified Linear Units (ReLU) function. Parameter sharing and local connectivity are two major properties of the convolution layers that can dramatically reduce the memory and computational requirements of CNNs [49].

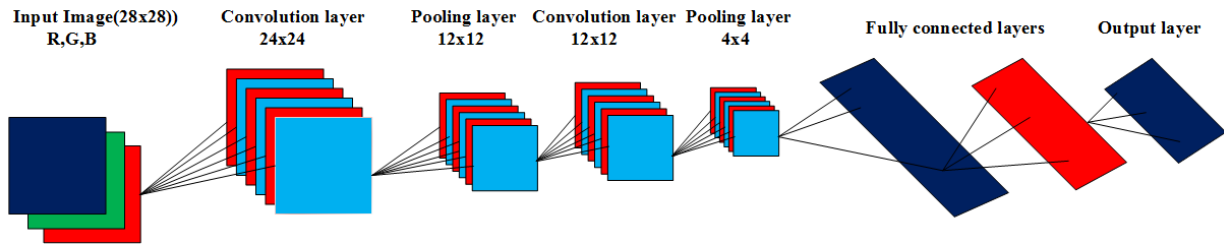


FIGURE 3. Basic configuration of 2D-CNN.

A Pooling or subsampling layer is another important block of CNN, which typically follows the convolution layer. It is mainly used to reduce the dimensionality of the convolutional layer feature maps and hence the number of its parameters. The maximum pooling is the most common subsampling technique in CNNs as it represents the best response. Thus, the  $i$ th pooling layer output can be calculated as follows [50]:

$$x_j^i = f(\beta_j^i \text{down}(x_j^{i-1}) + b_j^i) \quad (3)$$

where  $f$  denotes the activation function,  $\text{down}$  represents the subsampling function,  $b_j^i$  and  $\beta_j^i$  are the additive and multiplicative bias of the  $j$ th kernel. The features are captured through convolutional and pooling layers and then injected into the classification layers. The classification layers are identical to the hidden and output layers of the conventional MLNN that are represented as follows:

$$O^i = f^i(W^i O^{i-1} + b^i) \quad (4)$$

where:  $O^i$  is the output vector of the  $i$ th layer,  $f^i$  is the activation function in layer  $i$ ,  $W^i$  is the Weight matrix,  $b^i$  is the bias between layer  $i$  and  $i - 1$ . Finally, a soft-max classifier is typically used as an output layer to create the classes' labels. The soft-max layer is a common linear classifier, where the output can be calculated as follows [51]:

$$O = \frac{1}{\sum_{j=1}^K \exp(W_j x + b_j)} \begin{bmatrix} \exp(W_1 x + b_1) \\ \exp(W_2 x + b_2) \\ \dots \\ \exp(W_K x + b_K) \end{bmatrix} \quad (5)$$

In equation (5),  $x$  is the soft-max layer input,  $W_j$  represents the weight matrix, and  $b_j$  is the bias.

### B. LONG SHORT-TERM MEMORY NEURAL NETWORK

Recurrent Neural Network is a mathematical abstraction of biological nervous systems which can carry out complicated mappings from input to output sequences [52]. Recently, RNN has been a crucial focus of studies and improvement. The main advantage of RNN, in contrast to FFNN, is the connections from the network's output to its input. As a result, these connections permit a 'memory' of previous inputs to persist within the network's inner state and thereby have an impact on the network output [53]. However, the standard RNN suffers from the vanishing gradient problem which

leads to training failure [54]. Therefore, it is difficult for RNN to train long-sequence data. To address the problem of vanishing gradient in RNN, Hochreiter *et al.* introduced LSTM, which allows the recurrent neural network to process time-series sequences efficiently [55], [56]. The structure of the LSTM includes a set of recurrently connected subnets known as memory blocks. Each memory block includes at least one memory cell and three multiplicative gates, namely input, output, and the forget gates. These gates represent the write, read, and reset operations for the LSTM memory cells. Therefore, those multiplicative units allow LSTM memory blocks to store and access long-distance data.

Figure 4 illustrates the internal architecture of the LSTM unit. In this figure, the internal cell state is represented by  $C_t$ , and the output of the LSTM hidden layer is denoted by  $h_t$ . At time  $t$ , the input data sequence will be  $x(t)$ , the cell state becomes  $C_{t-1}$  and LSTM hidden layer output will be  $h_{t-1}$ . Therefore, the output of the forget unit, input unit, and output unit are calculated as follows [57]:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \quad (8)$$

Thus, the value of the current candidate cell state  $\tilde{C}$ , current cell state, and current LSTM unit output are calculated as follows:

$$\tilde{C} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (9)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C} \quad (10)$$

$$h_t = O_t \cdot \tanh(C_t) \quad (11)$$

here  $\sigma$  represents the sigmoid activation function.  $W_f$ ,  $W_i$ , and  $W_O$  are the weight coefficients of the forget, the input, and the output gates, respectively.  $b_O$ ,  $b_i$ , and  $b_f$  are the bias of output, input, and forget units, respectively.

### IV. PROPOSED APPROACH FOR REACTION WHEEL FAULT DIAGNOSIS

Nowadays, various CNN-LSTM architectures [58]–[60] have been proposed for fault diagnosis in different applications. However, most of these architectures suffer from computational complexity and high memory requirements due to utilizing sophisticated configurations. These drawbacks make

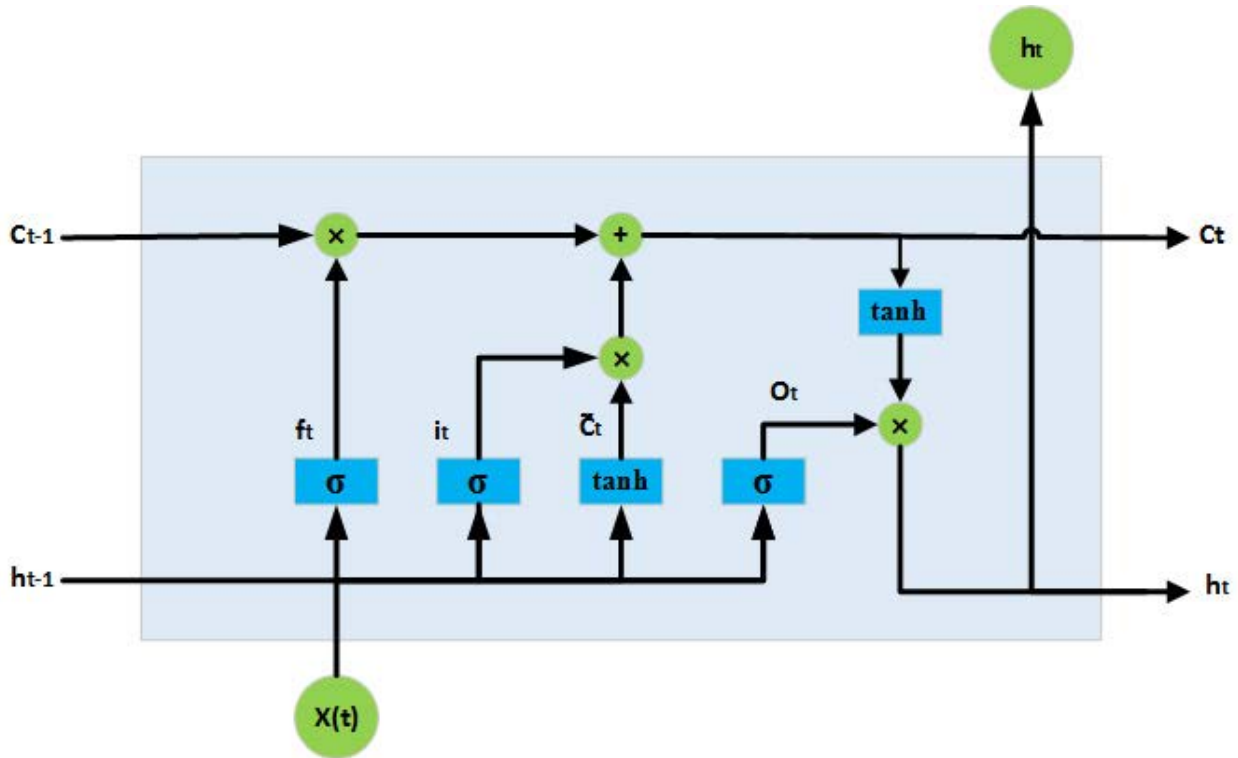


FIGURE 4. LSTM memory block with a single memory cell.

these architectures aren't appropriate for spacecraft onboard computer fault diagnosis tasks. Motivated by these shortcomings in the existing architectures in the literature, this research introduces a simple end-to-end Deep Learning (DL) architecture for spacecraft reaction wheel fault diagnosis. This section presents the proposed fault detection and identification approach for the spacecraft reaction wheel.

#### A. PROPOSED FAULT DIAGNOSIS SCHEME

Figure 5 depicts the proposed FDI scheme for the spacecraft reaction wheel. There are two phases for processing. One is the residual generation phase, where the residual signal is generated as the difference between the measured torque and the estimated torque. In order to ensure the efficacy of the proposed FDI scheme, the data integrity of the training dataset should be considered as proposed in [61]. Therefore, in this work, the torque is estimated using a high accuracy Radial Basis Function Neural Network (RBFNN) model for the reaction wheel, which is proposed in [62]. The aforementioned model is a compact model with high accuracy and low computational requirements. Consequently, it is well-suited for the spacecraft onboard computer applications in terms of low system complexity and low memory footprint. The other phase is the proposed CNN-LSTM model, where the residual time-series signals are used to extract distinctive features of different reaction wheel states. In the next subsection, the proposed CNN-LSTM model will be introduced.

#### B. PROPOSED CNN-LSTM MODEL FOR REACTION WHEEL FAULT DIAGNOSIS

This research proposes an end-to-end robust fault diagnosis architecture for spacecraft reaction wheels based on 1D-CNN and LSTM networks (CNN-LSTM). The motivation for using CNN is its ability to fulfill the feature extraction process without reliance on any traditional feature extraction technique as in the former researches. However, LSTM is used due to its efficacy to memorize the time series data with the help of its hidden memory units. Therefore, the combination of CNN and LSTM in one model allows us to diagnose the time-varying reaction wheel faults with high accuracy. Meanwhile, the raw residual time-series signals will be used directly as the input for the proposed fault diagnosis model without any preprocessing. In such a case, the fault diagnosis process will be accomplished in a single process using an end-to-end architecture. Thanks for this advantage that helps to minimize the model's complexity. The architecture of the proposed model is depicted in Figure 6. As can be observed from Figure 6, the proposed architecture includes four layers that are the convolutional layer, the pooling layer, LSTM layer, and the fully connected layer. First of all, the dataset is partitioned into training, validation, and testing subsets to be submitted to the proposed model. The raw residual signals are then normalized and applied to the convolutional layer. Meanwhile, multiple convolutional filters are applied to the input signal to extract the distinct features. Therefore, the output will be the convolution between the input signal and

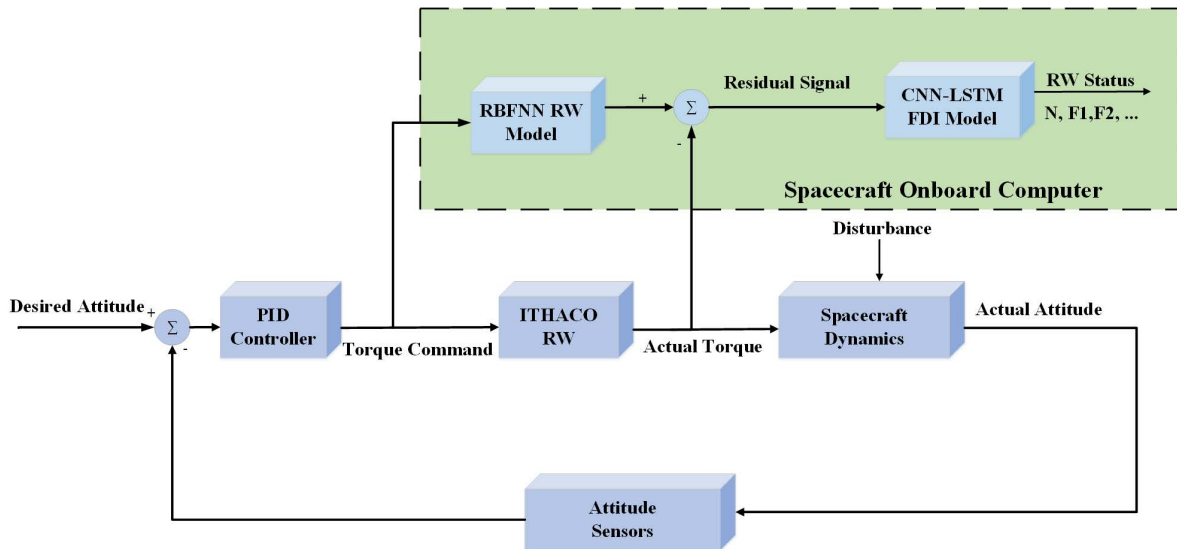


FIGURE 5. Proposed FDI scheme.

the applied filters as in equation (2). Further, the obtained feature maps will be downsampled through Max-pooling layer to reduce the dimensions of the input feature maps. The output of the pooling layer is then passed to the LSTM layer, which captures the internal features of the residual signal. Thus, the output of the LSTM layer can be described as follows [53]:

$$h_t = f(h_{t-1}, x_t, W) \quad (12)$$

where  $W$  represents the LSTM trainable parameters. The features of the input signals are captured through the LSTM layer, and finally, fed to the fully connected layers. The proposed FDI scheme is summarized in Algorithm 1.

**Algorithm 1** The Proposed Fault Identification Scheme

- Step 1. Import the dataset
- Step 2. Splitting the dataset into training, validation, and testing
- Step 3. Normalize the dataset
- Step 4. Select the number of LSTM hidden neurons
- Step 5. Set the Model hyperparameters
- Step 6. Train the models using the training dataset
  - for N in range (Number of filters) do
  - for M in range (filter Size) do
  - for I in range (Number of trials) do
  - model.fit(Training\_input, Training\_output)
  - end
  - end
  - end
- Step 7. Evaluate the trained models using the testing dataset model.evaluate (Testing\_input, Testing\_Output)
- Step 8. Cross-Validation for the fault types and normal state of the reaction wheel residual torque signals.

**V. RESULTS AND DISCUSSION**

This section demonstrates the effectiveness of the proposed fault diagnosis architecture as in the next subsections.

**A. TRAINING DATASET PREPARATION**

In this research, five different reaction wheel states are considered as shown in TABLE 2, namely Normal state (Healthy reaction wheel), low bus voltage Fault (F1), high bus voltage Fault (F2), motor current losses Fault (F3), and high friction Fault (F4). Furthermore, the low and high bus voltage faults are modeled and injected as a drop or increase in the reaction wheel bus voltage by 50% of the nominal bus voltage value. Besides, the motor current losses are modeled as 50% losses in the motor current. Ultimately, the friction faults are modeled as an increase in the bearing’s friction ( $\tau_v$ ) by three times of nominal friction value.

To train the proposed architecture, a sufficient and high-quality dataset should be provided. The residual time-series signals; are utilized as the input for the training process. Moreover, the training dataset is obtained by driving the ACS simulation model with tilting angles in the range of  $[-90^\circ, 90^\circ]$ . Meanwhile, the tilting angle is changed with  $0.1^\circ$  step. Therefore, 1801 samples are obtained for each reaction wheel class. These samples are randomly partitioned into 7205 samples (80%) for training, 900 samples (10%) for validation, and 900 samples (10%) for testing. An example of the reaction wheel residual torque signals is illustrated in Figure 7. The description of different classes and their corresponding labels are presented in TABLE 3. Furthermore, to reveal the capabilities of the introduced approach for diagnosing the time-varying reaction wheel faults, the proposed architecture is trained using samples of faults that are injected at 580 second only. However, the model is tested for faults that were injected at different instances of time (0, 100, 200, 300, 400, 500, and 550 seconds) as shown in TABLE 6.

**B. MODEL TRAINING AND HYPERPARAMETERS SELECTION**

The configuration of the proposed CNN-LSTM model is characterized by a group of hyper-parameters including the

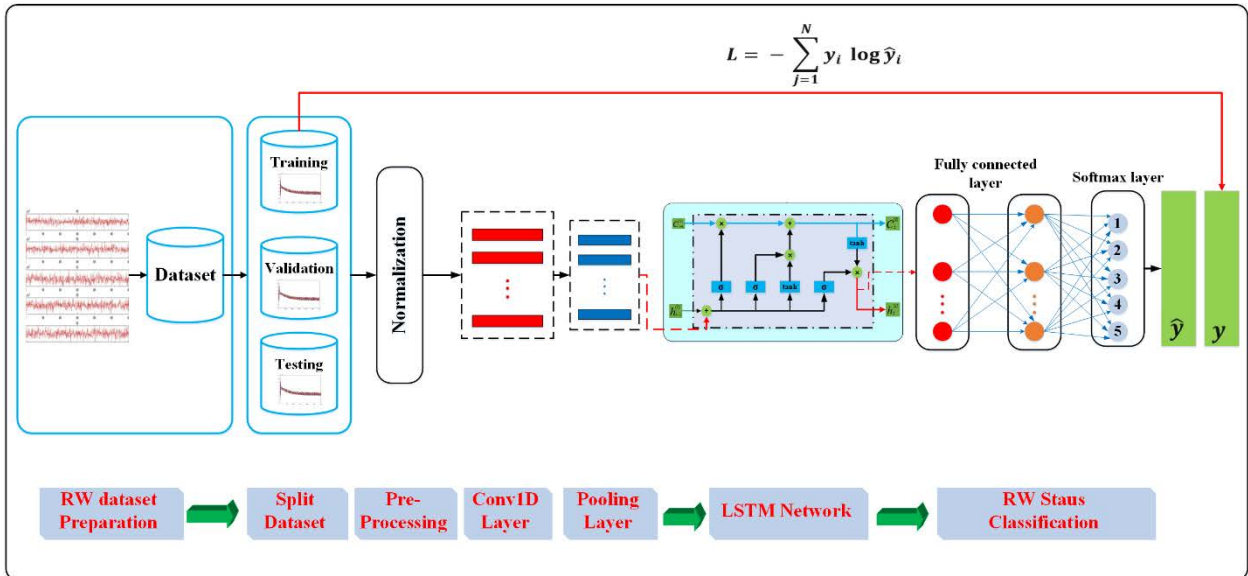


FIGURE 6. The proposed architecture for reaction wheel fault diagnosis.

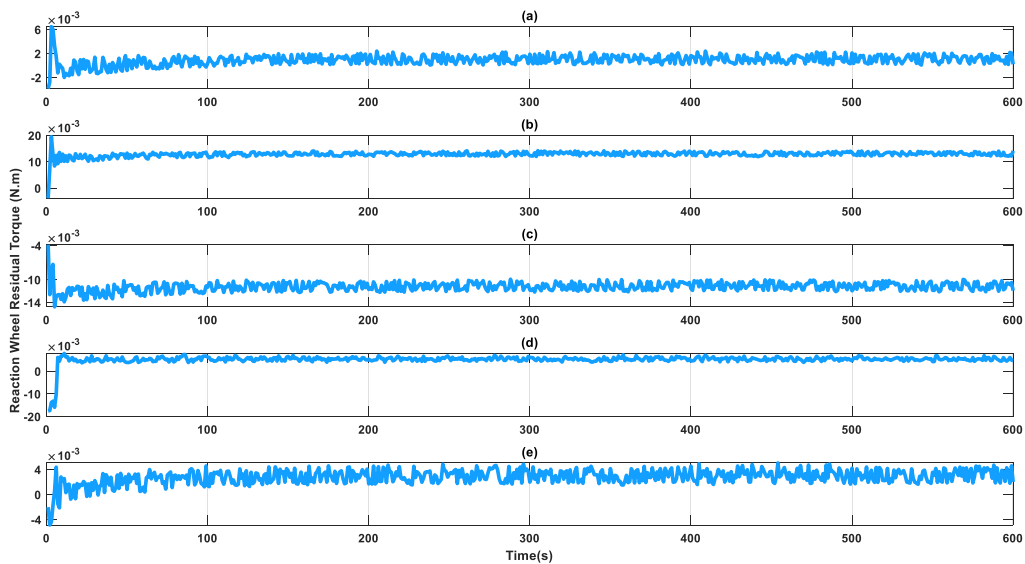


FIGURE 7. Raw residual signals waveforms for the five classes of the reaction wheel states: (a) normal signal; (b) low bus voltage fault; (c) high bus voltage fault; (d) motor current loss fault; (e) high friction fault.

TABLE 2. Reaction wheel states.

RW state	Description	RW affected Parameter
Normal (N)	The reaction wheel is healthy	
Fault 1 (F1)	Low bus voltage fault	$V_{bus}$
Fault 2 (F2)	High bus voltage fault	$V_{bus}$
Fault 3 (F3)	Motor current loss fault	$I_m$
Fault 4 (F4)	High friction fault	$\tau_v$

TABLE 3. Reaction wheel training dataset description.

Reaction wheel state	Number of samples			Label category
	Training	Validation	Testing	
N	1441	180	180	1
F1	1441	180	180	2
F2	1441	180	180	3
F3	1441	180	180	4
F4	1441	180	180	5

size and number of filters used in the convolutional layer, pooling layer factors, number of LSTM neurons, and the number of fully connected layers and neurons. Moreover,

selecting appropriate values for these hyper-parameters will produce high identification accuracy and protect the model against overfitting. The first hyperparameter that needs to be



configured is the size of the convolutional filters. A common practice in images applications is to utilize 2D filters of  $3 \times 3$  size [54]. However, utilizing 1D-CNN with  $3 \times 1$  filters to capture the faults' features from reaction wheel residual signals is impractical. Moreover, increasing the number and size of the convolutional kernels can help the model to extract more features from the raw residual signals. However, this will require more computational capabilities that aren't suitable for the spacecraft onboard computer systems. Consequently, in this research, we started with a 1D filter of size seven and then increase the number and size of filters gradually to ensure the maximum fault identification accuracy and model simplicity. Another hyper-parameter that should be determined is the down-sampling type and factors. A pooling layer with a large down-sampling window will significantly decrease the dimension of the feature maps. Nevertheless, it may result in a loss of some significant features. Moreover, in image classification applications, maximum pooling is more popular than others because it provides the strongest response [49]. Therefore, this research utilizes a Max-pooling as a down-sampling layer. Another motivation behind using such pooling type is when a fault occurs in the reaction wheel the ACS controller tries to compensate the response due to the effect of the fault by increasing or decreasing the torque voltage signal. As a result, the amplitude of the residual signal will increase or decrease according to the fault class. This means that the reaction wheel faults can be identified based on the amplitude of the residual signal. The number of LSTM hidden neurons is another hyperparameter that affects the model accuracy. In this work, we started with LSTM network with 10 hidden neurons and increase the number of hidden neurons gradually to get the maximum fault identification accuracy under different operating conditions. Based on the former principles, to select the best configuration for the proposed fault diagnosis model, various models with different hyperparameters were investigated. All these models were trained using Adam optimization algorithm. Further, this algorithm is a straightforward and computationally efficient algorithm that allocates a small memory [63]. Furthermore, all the models were trained using 100 training samples per batch through 50 epochs with a learning rate of 0.001. Moreover, the used loss function measures the gap between the actual output and the predicted output of the model. Consequently, the cross-entropy loss function can be used to evaluate the proposed models as illustrated in Figure 6 as follows [64]:

$$L = - \sum_{j=1}^N y_i \log \hat{y}_i \quad (13)$$

where  $N$  presents the total number of the training examples,  $y$  is the target distribution, and  $\hat{y}$  is the predicted distribution. Meanwhile, the training process is automatically stopped when the loss isn't decreased for 10 epochs. To reduce the effect of the perturbation, each CNN-LSTM model was trained for 10 times. All models in this research

were implemented using the Tensor flow toolbox in ANA-CONDA Spider for Python development environment on a machine with Intel Core i5 M520 CPU @ 2.40GHz and 8GB RAM. TABLE 4 and Figure 8 illustrate the iterative trials to find the optimal configuration for the proposed model. Firstly, the hyperparameters of the convolutional layer are investigated. As illustrated by TABLE 2, the accuracy of the model increases dramatically as the convolutional layer hyperparameters increase. The models with only one filter (models no. 1, 2, 3, 10, and 11) have a poor performance and will not be accepted in the design of safety-critical systems such as spacecrafts. On the other hand, the models with two and three filters have a better accuracy compared with the models with one filter. It can be shown that the best accuracy is achieved in model 18, by a convolutional layer with 3 filters of size 51. The number of neurons in the LSTM layer is investigated as shown in TABLE 4. The best performance is obtained for the models with 15 neurons in the LSTM layer. Models 19 to 27 show that the further increase in the size of the LSTM layer doesn't enhance the model performance, but increases the model complexity. Furthermore, the best results are obtained for models with a max-pooling layer of size 8. Models 20 and 21 indicate that reducing the size of the max-pooling layer to 4 will reduce the model performance. Therefore, model 18 is recommended for the reaction wheel fault diagnosis. Finally, the proposed configuration with 3 convolutional filters of size 51, max pooling layer of size 8, and LSTM layer with 15 neurons is recommended for the reaction wheel fault diagnosis task.

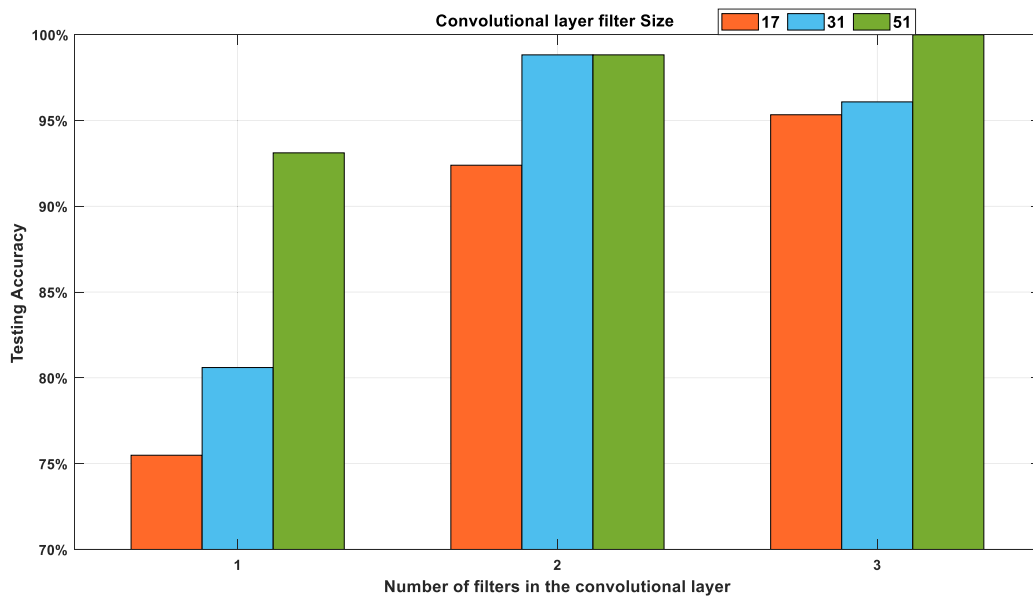
TABLE 5 shows the optimal parameters of the proposed FDI model. This model takes the reaction wheel torque residual signal as an input. The first convolutional layer uses 3 kernels of size 51 that convolve with the residuals to generate the feature maps. Then the rectified linear (ReLU) function is applied to these feature maps to generate the output feature vectors. In the maximum pooling layer, the feature maps are transformed into 71 scalars representing the maximum values of the 3 feature maps from the former convolutional layer. Further, the features are fed into the LSTM layer with 15 neurons. Thus, the output of the LSTM layer is fed to 8 neurons fully connected layer that is followed by 5 neurons SoftMax layer to encode the faults that occurred in the reaction wheel.

### C. ABLATION STUDY

An ablation study is a set of trials where various layers of a machine learning model are replaced or removed to measure the impact of these layers on the performance [65]. Therefore, to determine which parts of the proposed FDI model are critical for fault diagnosis accuracy, an ablation study is presented in TABLE 7. The first observation from TABLE 7 is that the LSTM layer has the most significant impact on the model accuracy. It can be seen that removing the LSTM layer would reduce the average testing accuracy from 99.99% to 75%. It can be shown also from TABLE 7 that the best activation function for the convolutional and the Fully Connected (FC) layer is the (ReLU) function. On the other

**TABLE 4.** Different CNN-LSTM models and corresponding testing accuracy.

Model No.	Number of filters in the convolutional layer	Size of the convolutional layer filter	No. Of LSTM neurons	Size of Max pooling layer	Testing Accuracy
1	1	17	10	8	86.77%
2	1	31	10	8	77.17%
3	1	51	10	8	94.19%
4	2	17	10	8	96.20%
5	2	31	10	8	96.28%
6	2	51	10	8	97.28%
7	3	17	10	8	98.04%
8	3	31	10	8	94.93%
9	3	51	10	8	98.89%
10	1	17	15	8	75.49%
11	1	31	15	8	80.60%
12	1	51	15	8	93.11%
13	2	17	15	8	92.39%
14	2	31	15	8	98.82%
15	2	51	15	8	98.82%
16	3	17	15	8	95.33%
17	3	31	15	8	96.08%
<b>18</b>	<b>3</b>	<b>51</b>	<b>15</b>	<b>8</b>	<b>99.99%</b>
19	3	17	15	4	95.52%
20	3	31	15	4	93.72%
21	3	51	15	4	99.23%
22	3	31	20	8	97.14%
23	2	31	20	8	98.15%
24	2	51	20	8	97.39%
25	3	31	20	8	97.14%
26	4	31	20	8	97.07%
27	4	61	20	8	99.49%



**FIGURE 8.** Testing accuracies of various models.

hand, using (elu) function as an activation for the convolutional layer has decreased the accuracy to 89.1%. Besides, using (elu) function as an activation for the fully connected layer would decrease the accuracy to 97.4%. Furthermore, TABLE 7 illustrates that utilizing the (AveragePooling1D) layer instead of (Maxpooling1D) would decrease the testing accuracy to 95%. Moreover, it is shown that removing the conv1D layer would decrease the accuracy to 93%.

**D. REACTION WHEEL FAULT DIAGNOSIS USING THE PROPOSED MODEL**

To ensure the capability of the proposed approach to provide significant fault diagnosis accuracy and avert overfitting, various datasets under different operating conditions were evaluated as illustrated in TABLE 6. Moreover, the dataset in TABLE 6 is another example that differs from the training dataset that was used to train the introduced model. Thus,

Predicted label	N	181 20%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	181 100%
	F1	0 0.0%	181 20%	0 0.0%	0 0.0%	0 0.0%	181 100%
	F2	0 0.0%	0 0.0%	181 20%	0 0.0%	0 0.0%	181 100%
	F3	0 0.0%	0 0.0%	0 0.0%	181 20%	0 0.0%	181 100%
	F4	0 0.0%	0 0.0%	0 0.0%	0 0.0%	181 20%	181 100%
		181 100%	181 100%	181 100%	181 100%	181 100%	905 100%
	N	F1	F2	F3	F4	Sum	
	Target label						

FIGURE 9. Results of reaction wheel fault diagnosis at tilting angle  $[-90,90]$  for faults injected at 0 second.

TABLE 5. Recommended parameters of the proposed model.

No.	Parameter	Value
1	Number of Convolutional layer filters	3
2	Size of kernel in convolutional layer	51
3	Convolutional layer activation function	ReLU
4	Size of max-pooling layer	8
5	Number of LSTM layer neurons	15
6	Number fully connected layer neurons	8
7	Fully connected layer activation function	ReLU
8	Number of SoftMax layer neurons	5

TABLE 6. Dataset used to test the proposed model.

RW state	Tilting Angle	Dataset Size	Fault injection Time (Sec.)	Label category
N	$[-90^\circ, 90^\circ]$	181	-	1
F1	$[-90^\circ, 90^\circ]$	181	0/50/100/200/300/400/550	2
	$[-180^\circ, 180^\circ]$	362		
F2	$[-90^\circ, 90^\circ]$	181	0/50/100/200/300/400/550	3
	$[-180^\circ, 180^\circ]$	362		
F3	$[-90^\circ, 90^\circ]$	181	0/50/100/200/300/400/550	4
	$[-180^\circ, 180^\circ]$	362		
F4	$[-90^\circ, 90^\circ]$	181	0/50/100/200/300/400/550	5
	$[-180^\circ, 180^\circ]$	362		

the accuracy of the proposed model will be investigated at different time-varying conditions. Furthermore, to validate the generalizability of the proposed model, a dataset of tilting angles in the range of  $[-180^\circ, 180^\circ]$  is used.

The confusion matrix is an efficient visualization tool that is used for evaluating the performance of the classification scheme [64]. Therefore, the confusion matrix is used in this research to analyze the performance of the proposed method. The model is tested using a dataset of faults that are injected at an inception time of 0 second with tilting angles in the range of  $[-90^\circ, 90^\circ]$  and  $[-180^\circ, 180^\circ]$ . This dataset is chosen as the worst-case scenario because the model is trained using a set of faults that were injected at 580 second with tilting angles in the range of  $[-90^\circ, 90^\circ]$ . Figure 9 and Figure 10 illustrate the confusion matrices using the proposed approach for the classification of the reaction wheel testing samples. The results indicate that the overall classification accuracies of the normal class and the four faults classes are almost 100%. This performance proves the superiority of the proposed algorithm, which strongly identifies the reaction wheel states. Moreover, this performance confirms the requirements of safety-critical systems like spacecraft in which the miss classification may cause the failure of the system. The results of reaction wheel fault diagnosis using the testing dataset of TABLE 6 are summarized in Figure 11, which reveals the superior performance of the proposed fault diagnosis model.

E. SENSITIVITY ANALYSIS

This section investigates the robustness of the proposed FDI approach. In order to show the robustness of the proposed

**TABLE 7.** Average fault diagnosis accuracy utilizing different model configuration.

Model layers						Average Testing accuracy
Conv1D	Conv1D Activation (ReLU)	Maxpooling1D	LSTM	FC	FC activation (ReLU)	99.99%
-	-	-	LSTM	FC	FC activation (ReLU)	93%
Conv1D	Conv1D Activation (ReLU)	Maxpooling1D	-	FC	FC activation (ReLU)	75%
Conv1D	Conv1D Activation (ReLU)	AveragePooling1D	LSTM	FC	FC activation (ReLU)	95%
Conv1D	Conv1D Activation (sigmoid)	Maxpooling1D	LSTM	FC	FC activation (ReLU)	98.2
Conv1D	Conv1D Activation (elu)	Maxpooling1D	LSTM	FC	FC activation (ReLU)	89.1
Conv1D	Conv1D Activation (ReLU)	Maxpooling1D	LSTM	FC	FC activation (sigmoid)	92.6
Conv1D	Conv1D Activation (ReLU)	Maxpooling1D	LSTM	FC	FC activation (elu)	97.34

<b>Predicted label</b>	<b>N</b>	<b>361</b> <b>20%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>100%</b>
	<b>F1</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>20%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>100%</b>
	<b>F2</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>20%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>100%</b>
	<b>F3</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>20%</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>100%</b>
	<b>F4</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>0</b> <b>0.0%</b>	<b>361</b> <b>20%</b>	<b>361</b> <b>100%</b>
		<b>361</b> <b>100%</b>	<b>361</b> <b>100%</b>	<b>361</b> <b>100%</b>	<b>361</b> <b>100%</b>	<b>361</b> <b>100%</b>	<b>1805</b> <b>100%</b>
	<b>N</b>	<b>F1</b>	<b>F2</b>	<b>F3</b>	<b>F4</b>	<b>Sum</b>	
	<b>Target label</b>						

**FIGURE 10.** Results of reaction wheel fault diagnosis at tilting angle  $[-180,180]$  for faults injected at 0 second.

model against noise, its performance is evaluated using a testing dataset with different noise levels. Furthermore, additive white Gaussian noise with different levels is added to the raw residual signals using the MATLAB function `awgn()` as in [66]–[68]. Figure 12 shows the noisy residual signals with different noise levels ranging from 10% to 50% of the original raw signal amplitude. The results of reaction wheel fault

diagnosis using the proposed approach under different noise levels are depicted in TABLE 8. It can be seen from TABLE 8 that the identification accuracy is less satisfactory for noisy residual signals with noise levels of 40% and 50% of the original signals. The main reason is that the residual signals are highly distorted at these high noise levels as shown in Figure 12. Moreover, it is evident from TABLE 8 that the pro-

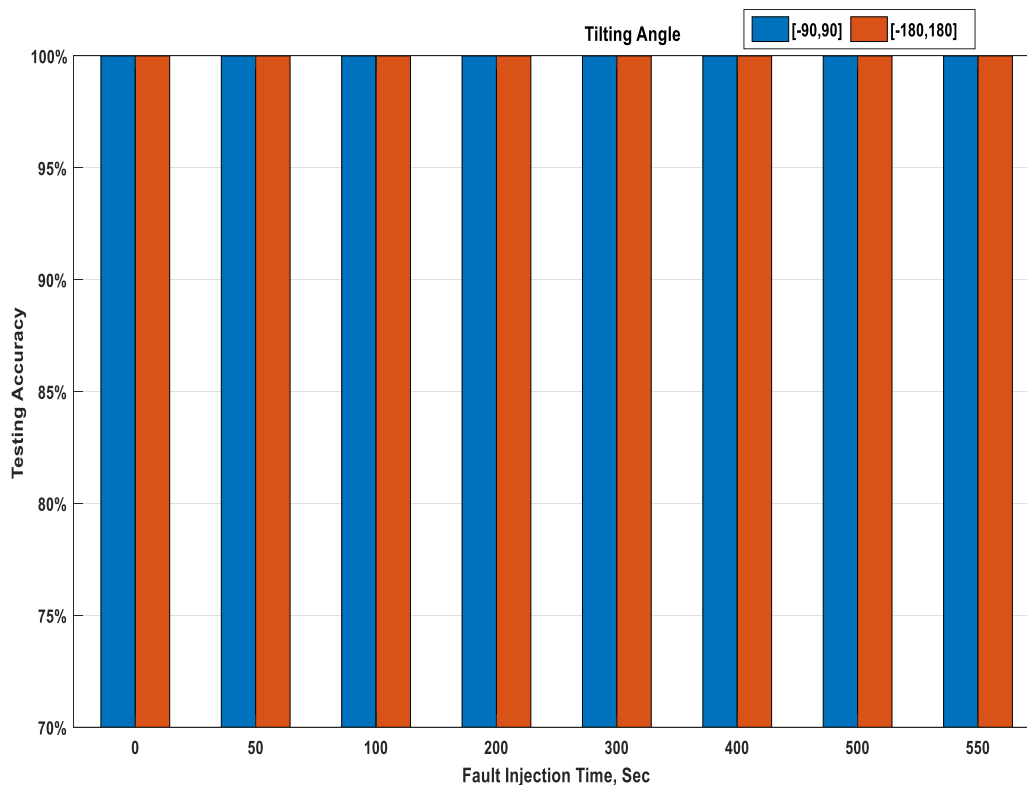


FIGURE 11. Results of reaction wheel fault diagnosis using dataset in Table 6.

TABLE 8. Reaction wheel fault diagnosis accuracies at different noise amplitudes.

RW State	Accuracies at different noise levels (%)					
	Original	10%	20%	30%	40%	50%
N	100%	100%	100%	100%	87.29%	83.45%
F1	100%	100%	100%	100%	100%	96.38%
F2	100%	100%	100%	100%	100%	94.37%
F3	100%	100%	100%	100%	96.13%	89.19%
F4	100%	100%	100%	100%	86.74%	82.26%
Average	100%	100%	100%	100%	94.03%	89.13%

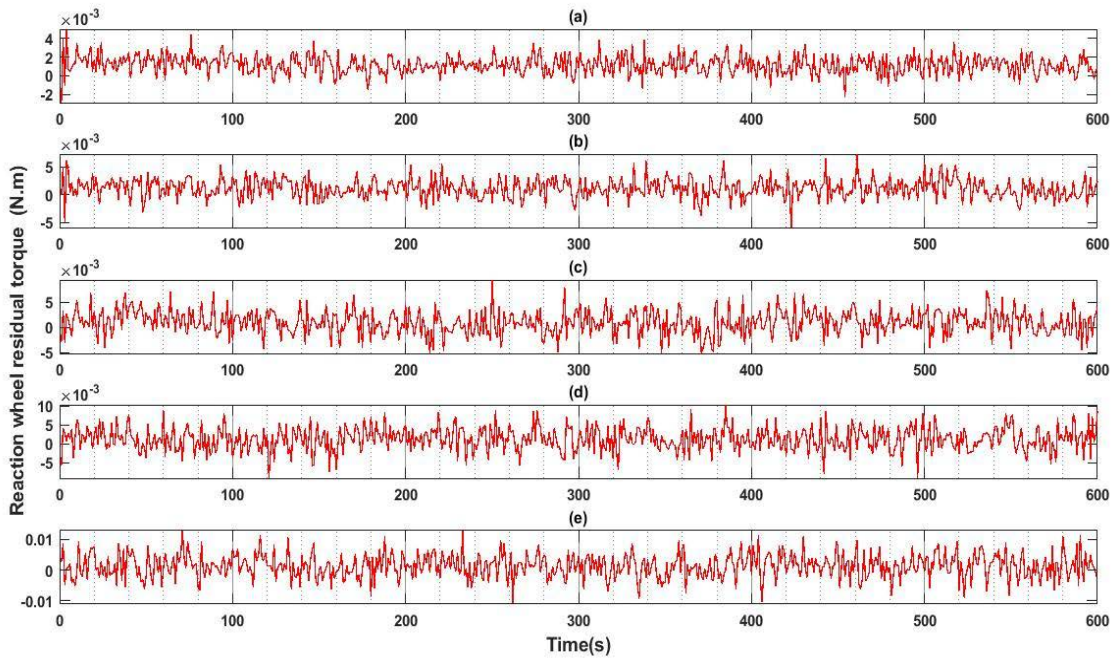
posed FDI method can achieve an average fault identification accuracy of 100% for noisy signals with noise amplitudes up to 30%. The obtained results reveal the robustness of the proposed scheme against strong noise. Besides, the proposed approach can perform well in harsh environments without utilizing computationally expensive techniques such as Deep Belief Network (DBN) and Sparse Autoencoder (SAE).

F. COMPARATIVE STUDY

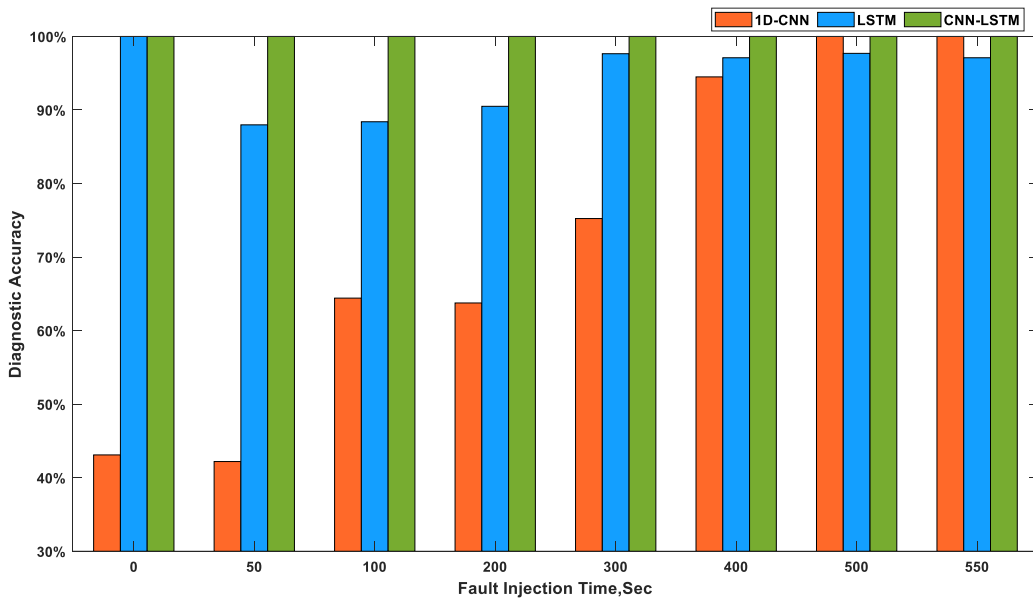
To further certify the efficacy of the proposed fault diagnosis model and reveal the superiority of the proposed approach, its results have been compared with two recent fault diagnosis techniques. These techniques include the 1D-CNN-based fault diagnosis method [69], and the LSTM-based fault diagnosis presented in [70]. Meanwhile, the three techniques were employed to diagnose the reaction wheel faults using the testing dataset in TABLE 6. Figure 13 illustrates the diagnosis

accuracies obtained by the 1D-CNN, LSTM, and the proposed method. As can be observed from Figure 13, it is interesting that the proposed method can achieve a superior diagnostic accuracy of 100% for different reaction wheel states and at all the time-varying conditions. Although the proposed model was trained only for faults that were injected at 580 second, it can diagnose faults that were injected at 0, 100, 200, 300, 400, 500, and 550 seconds with 100% accuracy. 1D-CNN-based approach achieves 100% in diagnosing faults that are occurred at 550 and 500 seconds. However, it achieves poor diagnosis accuracies (less than 80%) for faults that were injected at 0, 50,100,200, and 300 seconds. In that way, the average classification accuracy of the 1D-CNN is 75%. On the other hand, the LSTM reaches better accuracies (more than 85%) under all the time-varying conditions and an average accuracy of 94.5%. This accuracy demonstrates the efficacy of the LSTM networks in capturing the long-term time dependencies from the input signals, thanks to the hidden LSTM cells. However, this accuracy is not suitable for reaction wheel fault diagnosis because the space application should not have any jitter in fault identification due to its criticality. By combining the 1D-CNN and LSTM architectures in one model, this research utilizes the advantages of both architectures and the accuracy is improved to 100%.

To further investigate the performance of the proposed approach, it is compared to the recent reaction wheel fault diagnosis approaches in the literature such as [22] and [25].



**FIGURE 12.** Noisy residual signals with different noise levels amplitudes: (a) 10% of the original signal amplitude; (b) 20% of the original signal amplitude; (c) 30% of the original signal amplitude; (d) 40% of the original signal amplitude; (e) 50% of the original signal amplitude.



**FIGURE 13.** Comparison between the identification accuracies of different methods at different instances.

Contrary to the proposed CNN-LSTM model, the proposed approaches in [22] and [25] have used time-consuming feature extraction techniques. In [22], the authors relied upon the Autoregressive Moving Average (ARMA) model as a feature extraction method and MLP neural network as faults' classifier. In [25], Prony method is used as a feature extraction technique and FFNN is used for faults identification. Although these methods can achieve 99% and 100%

identification accuracies respectively, these accuracies can be achieved only for faults that are injected at a certain inception time. However, to reach a high accuracy using these approaches at different inception times, the network designer has to use a huge dataset and a complex network design. The proposed methodology is compared also with a machine learning approach that is introduced in [29]. The former technique can classify reaction wheel faults with

**TABLE 9. Comparison between different deep learning methods utilized in fault diagnosis.**

Method	Description	Pros	Cons
Deep Belief Network (DBN)	<ul style="list-style-type: none"> <li>Comprises of many Restricted Boltzmann Machines (RBM)</li> <li>Recent approaches have been proposed for fault diagnosis [30]-[33]</li> </ul>	<ul style="list-style-type: none"> <li>Can extract the deep difference between normal and faulty data [40]</li> <li>Revealed high accuracy in image recognition, and speech recognition [40]</li> <li>Utilizes greedy learning method for initializing the model [71]</li> <li>Good generalizability [72]</li> </ul>	<ul style="list-style-type: none"> <li>Computationally expensive [40]</li> <li>Requires many trainable parameters [72]</li> <li>Suffer from the issue of gradient disappearing in reverse fine-tuning [73]</li> <li>Accuracy is less significant when used in fault diagnosis [74]</li> <li>Not appropriate for safety-critical systems due to the time required for intensive calculations.</li> </ul>
Sparse Autoencoder (SAE)	<ul style="list-style-type: none"> <li>Unsupervised learning utilized mainly to extract features or reduce feature dimension</li> <li>Comprises of encoder that extracts features and a decoder that feeds the features to the hidden layers</li> <li>many approaches were proposed for fault diagnosis [75]-[77]</li> </ul>	<ul style="list-style-type: none"> <li>Remove the dependency on separate feature extraction techniques [40]</li> <li>Labeled data isn't required for training [74]</li> <li>Robust features [76]</li> </ul>	<ul style="list-style-type: none"> <li>Performance depends on a large number of hyper-parameters [40]</li> <li>May suffer from the vanishing of errors [74]</li> <li>Requires a pre-training phase [75]</li> <li>Large memory footprint [77]</li> <li>Not appropriate for Embedded computers due to the limited resources nature of embedded systems.</li> </ul>
Deep neural networks such as the Deep temporal Dictionary Learning (DTDLD)	<ul style="list-style-type: none"> <li>LSTM autoencoder that has tunable time-based states [78]</li> </ul>	<ul style="list-style-type: none"> <li>A dictionary matrix is trained to capture significant features [78]</li> <li>Can extract temporal features of signals [78]</li> </ul>	<ul style="list-style-type: none"> <li>high computational cost</li> <li>A large number of learned parameters</li> <li>Not suitable for real-time applications in general and spacecraft onboard computer due to its limited resources</li> </ul>
Generative Adversarial Networks (GAN)	<ul style="list-style-type: none"> <li>Contain two parts: the generator and discriminator</li> <li>Applied to fault diagnosis in different areas [79]-[80]</li> </ul>	<ul style="list-style-type: none"> <li>Superior generative Capabilities [80]</li> <li>Doesn't require Monte Carlo approximation to be trained [71]</li> <li>Can be used to solve the incomplete dataset problem [61]</li> </ul>	<ul style="list-style-type: none"> <li>the training process is unstable [71]</li> <li>A single GAN model can't guarantee acceptable diagnostic accuracy [79]</li> <li>High computational complexity</li> </ul>

an accuracy of only  $\sim 59\%$  which isn't accepted for space applications.

The proposed approach is compared as well with the recent Deep Learning (DL) approaches in literature such as Deep Belief Network (DBN), Sparse Autoencoder (SAE), and the Deep temporal Dictionary Learning (DTDLD). TABLE 9 briefly represents the advantages and disadvantages of these approaches. In contrast to the approaches that are summarized in TABLE 9, the proposed FDI method is a compact architecture that can achieve fault identification accuracies of 100% at different operating conditions.

Fortunately, the proposed approach is a robust method that can operate with superior performance in noisy environments with noise levels up to 30% of the original signals. Finally, the former advantages of the proposed scheme ensure that the proposed approach is most appropriate for spacecraft reaction wheel fault diagnosis.

## VI. CONCLUSION

This research introduces a new end-to-end architecture to detect and identify spacecraft RW faults. The proposed architecture combines 1D-CNN and LSTM in one network architecture and uses the advantages of both networks to further improve the fault identification accuracy. The proposed model presents the following advantages:

- It is trained directly using the raw residual signals without reliance upon any other time-consuming feature extraction techniques
- Compact deep learning structure that has low computational complexity with a small memory footprint. Consequently, makes it suitable for real-time applications

in general and specifically for spacecraft onboard applications;

- The proposed method achieves fault identification accuracy of almost 100% under different reaction wheel operating conditions;
- The proposed model reveals superior generalizability in diagnosing all reaction wheel states under worst-case tilting angles in the range of  $[-180^\circ, 180^\circ]$  and different faults inception times.
- A robust fault diagnosis model that can achieve an average identification accuracy of 100 % in noisy environments with noise amplitudes up to 30% of the original signals.

Furthermore, the proposed fault diagnosis scheme is compared with 1D-CNN and LSTM-based fault diagnosis methods. 1D-CNN method achieved an average accuracy of 75%, whereas the LSTM-based approach achieved an average accuracy of 94.5%. The obtained identification accuracies demonstrated that the proposed architecture can capture highly discriminative features directly from residual signals. The proposed fault diagnosis approach can be adapted to be applicable to other spacecraft subsystems as well as different safety-critical systems as in the industry.

## REFERENCES

- M. Tafazoli, "A study of on-orbit spacecraft failures," *Acta Astronaut.*, vol. 64, nos. 2-3, pp. 195-205, Jan. 2009, doi: 10.1016/j.actaastro.2008.07.019.
- S. A. Jacklin, "Small-satellite mission failure rates," NASA Ames Res. Center, Moffett Field, CA, USA, Tech. Rep. NASA/TM-2018-220034, 2019. [Online]. Available: <https://ntrs.nasa.gov/>

- [3] X.-Y. Ji, Y.-Z. Li, G.-Q. Liu, J. Wang, S.-H. Xiang, X.-N. Yang, and Y.-Q. Bi, "A brief review of ground and flight failures of Chinese spacecraft," *Prog. Aerosp. Sci.*, vol. 107, pp. 19–29, May 2019, doi: [10.1016/j.paerosci.2019.04.002](https://doi.org/10.1016/j.paerosci.2019.04.002).
- [4] J. Kampmeier, R. Larsen, L. F. Migliorini, and K. A. Larson, "Reaction wheel performance characterization using the Kepler spacecraft as a case study," in *Proc. 15th Int. Conf. Space Oper.*, Marseille, France, May 2018, pp. 1–17, doi: [10.2514/6.2018-2563](https://doi.org/10.2514/6.2018-2563).
- [5] G. Vachtsevanos, F. L. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Hoboken, NJ, USA: Wiley, 2006.
- [6] D. Ma, X. Hu, and H. Zhang, "A hierarchical event detection method based on spectral theory of multidimensional matrix for power system," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 51, no. 4, pp. 2173–2186, Apr. 2021, doi: [10.1109/TSMC.2019.2931316](https://doi.org/10.1109/TSMC.2019.2931316).
- [7] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015, doi: [10.1109/TIE.2015.2417501](https://doi.org/10.1109/TIE.2015.2417501).
- [8] R. Patton, P. Frank, and R. Clark, *Issues of Fault Diagnosis for Dynamic Systems*. London, U.K.: Springer-Verlag, 2006.
- [9] X. Dai and Z. Gao, "From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2226–2238, Nov. 2013, doi: [10.1109/TII.2013.2243743](https://doi.org/10.1109/TII.2013.2243743).
- [10] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, Jun. 2015, doi: [10.1109/TIE.2015.2419013](https://doi.org/10.1109/TIE.2015.2419013).
- [11] N. Tudoroiu and K. Khorasani, "Satellite fault diagnosis using a bank of interacting Kalman filters," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 4, pp. 1334–1350, Oct. 2007, doi: [10.1109/TAES.2007.4441743](https://doi.org/10.1109/TAES.2007.4441743).
- [12] S. P. Zhao and K. Khorasani, "A recurrent neural network based fault diagnosis scheme for a satellite," in *Proc. 33rd Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2007, pp. 2660–2665, doi: [10.1109/IECON.2007.4459995](https://doi.org/10.1109/IECON.2007.4459995).
- [13] N. Tudoroiu, E. Sobhani-Tehrani, K. Khorasani, T. Letia, and R. Tudoroiu, "Real-time embedded fault detection estimators in a satellite's reaction wheels," in *Proc. Int. Multiconf. Comput. Sci. Inf. Technol.*, Oct. 2010, pp. 759–766, doi: [10.1109/IMCSIT.2010.5679730](https://doi.org/10.1109/IMCSIT.2010.5679730).
- [14] P. Baldi, P. Castaldi, S. Simani, and G. Bertoni, "Fault diagnosis and control reconfiguration for satellite reaction wheels," in *Proc. Conf. Control Fault-Tolerant Syst. (SysTol)*, Oct. 2010, pp. 143–148, doi: [10.1109/SYSTOL.2010.5676066](https://doi.org/10.1109/SYSTOL.2010.5676066).
- [15] M. H. Amoozgar, Y. M. Zhang, J. Lee, and A. Ng, "A fault detection and diagnosis technique for spacecraft in formation flying," *IFAC Proc. Volumes*, vol. 45, no. 20, pp. 301–306, Jan. 2012, doi: [10.3182/20120829-3-MX-2028.00269](https://doi.org/10.3182/20120829-3-MX-2028.00269).
- [16] H. Talebi and K. Khorasani, "A neural network-based multiplicative actuator fault detection and isolation of nonlinear systems," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 3, pp. 842–851, May 2013, doi: [10.1109/TCST.2012.2186634](https://doi.org/10.1109/TCST.2012.2186634).
- [17] E. Sobhani-Tehrani, H. A. Talebi, and K. Khorasani, "Hybrid fault diagnosis of nonlinear systems using neural parameter estimators," *Neural Netw.*, vol. 50, pp. 12–32, Feb. 2014, doi: [10.1016/j.neunet.2013.10.005](https://doi.org/10.1016/j.neunet.2013.10.005).
- [18] O. Nasri, I. Gueddi, K. Benothman, and P. Dague, "Fault diagnosis of spacecraft reaction wheels based on principal component analysis," in *Proc. 4th Int. Conf. Syst. Control (ICSC)*, Apr. 2015, pp. 91–96, doi: [10.1109/ICoSC.2015.7152772](https://doi.org/10.1109/ICoSC.2015.7152772).
- [19] R. Wang, X. Gong, M. Xu, and Y. Li, "Fault detection of flywheel system based on clustering and principal component analysis," *Chin. J. Aeronaut.*, vol. 28, no. 6, pp. 1676–1688, Dec. 2015.
- [20] I. Gueddi, O. Nasri, K. Benothman, and P. Dague, "VPCA-based fault diagnosis of spacecraft reaction wheels," in *Proc. Int. Conf. Inf., Commun. Autom. Technol. (ICAT)*, Oct. 2015, pp. 1–6, doi: [10.1109/ICAT.2015.7340524](https://doi.org/10.1109/ICAT.2015.7340524).
- [21] X. Chen, R. Sun, W. Jiang, Q. Jia, and J. Zhang, "A novel two-stage extended Kalman filter algorithm for reaction flywheels fault estimation," *Chin. J. Aeronaut.*, vol. 29, no. 2, pp. 462–469, Apr. 2016, doi: [10.1016/j.cja.2016.01.008](https://doi.org/10.1016/j.cja.2016.01.008).
- [22] E. A. Omran and W. A. Murtada, "Fault detection and identification of spacecraft reaction wheels using autoregressive moving average model and neural networks," in *Proc. 12th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2016, pp. 77–82, doi: [10.1109/ICENCO.2016.7856449](https://doi.org/10.1109/ICENCO.2016.7856449).
- [23] C. U. Mba, H. A. Gabbar, S. Marchesiello, and A. Fasana, "Fault diagnosis in flywheels: Case study of a reaction wheel dynamic system with bearing imperfection," *Int. J. Performability Eng.*, vol. 13, pp. 362–373, Jun. 2017, doi: [10.23940/ijpe.17.04.p3.362373](https://doi.org/10.23940/ijpe.17.04.p3.362373).
- [24] V. Izadi, M. Abedi, and H. Bolandi, "Supervisory algorithm based on reaction wheel modelling and spectrum analysis for detection and classification of electromechanical faults," *IET Sci., Meas. Technol.*, vol. 11, no. 8, pp. 1085–1093, Nov. 2017, doi: [10.1049/iet-smt.2017.0137](https://doi.org/10.1049/iet-smt.2017.0137).
- [25] E. A. Omran and W. A. Murtada, "Efficient anomaly classification for spacecraft reaction wheels," *Neural Comput. Appl.*, vol. 31, no. 7, pp. 2741–2747, Oct. 2017, doi: [10.1007/s00521-017-3226-y](https://doi.org/10.1007/s00521-017-3226-y).
- [26] W. A. Murtada and E. A. Omran, "Robust anomaly identification algorithm for noisy signals: Spacecraft solar panels model," *Neural Comput. Appl.*, vol. 32, no. 16, pp. 12281–12294, Aug. 2019, doi: [10.1007/s00521-019-04407-2](https://doi.org/10.1007/s00521-019-04407-2).
- [27] H. A. Nozari, P. Castaldi, H. D. Banadaki, and S. Simani, "Novel non-model-based fault detection and isolation of satellite reaction wheels based on a mixed-learning fusion framework," *IFAC-PapersOnLine*, vol. 52, no. 12, pp. 194–199, 2019, doi: [10.1016/j.ifacol.2019.11.222](https://doi.org/10.1016/j.ifacol.2019.11.222).
- [28] A. Rahimi, K. D. Kumar, and H. Alighanbari, "Fault isolation of reaction wheels for satellite attitude control," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 1, pp. 610–629, Feb. 2020, doi: [10.1109/TAES.2019.2946665](https://doi.org/10.1109/TAES.2019.2946665).
- [29] A. Rahimi and A. Saadat, "Fault isolation of reaction wheels onboard three-axis controlled in-orbit satellite using ensemble machine learning," *Aerosp. Syst.*, vol. 3, no. 2, pp. 119–126, Mar. 2020, doi: [10.1007/s42401-020-00046-x](https://doi.org/10.1007/s42401-020-00046-x).
- [30] C. Tao, X. Wang, F. Gao, and M. Wang, "Fault diagnosis of photovoltaic array based on deep belief network optimized by genetic algorithm," *Chin. J. Electr. Eng.*, vol. 6, no. 3, pp. 106–114, Sep. 2020, doi: [10.23919/CJEE.2020.000024](https://doi.org/10.23919/CJEE.2020.000024).
- [31] C. Zhang, Y. B. Zhang, C. X. Hu, Z. B. Liu, L. Y. Cheng, and Y. Zhou, "A novel intelligent fault diagnosis method based on variational mode decomposition and ensemble deep belief network," *IEEE Access*, vol. 8, pp. 36293–36312, 2020, doi: [10.1109/ACCESS.2020.2969412](https://doi.org/10.1109/ACCESS.2020.2969412).
- [32] D. Zhu, X. Cheng, L. Yang, Y. Chen, and S. X. Yang, "Information fusion fault diagnosis method for deep-sea human occupied vehicle thruster based on deep belief network," *IEEE Trans. Cybern.*, early access, Mar. 11, 2021, doi: [10.1109/TCYB.2021.3055770](https://doi.org/10.1109/TCYB.2021.3055770).
- [33] X. Huang, X. Zhang, Y. Xiong, H. Liu, and Y. Zhang, "A novel intelligent fault diagnosis approach for early cracks of turbine blades via improved deep belief network using three-dimensional blade tip clearance," *IEEE Access*, vol. 9, pp. 13039–13051, 2021, doi: [10.1109/ACCESS.2021.3052217](https://doi.org/10.1109/ACCESS.2021.3052217).
- [34] X. Yan, D. She, Y. Xu, and M. Jia, "Deep regularized variational autoencoder for intelligent fault diagnosis of rotor-bearing system within entire life-cycle process," *Knowl.-Based Syst.*, vol. 226, Aug. 2021, Art. no. 107142, doi: [10.1016/j.knsys.2021.107142](https://doi.org/10.1016/j.knsys.2021.107142).
- [35] X. Wu, Y. Zhang, C. Cheng, and Z. Peng, "A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery," *Mech. Syst. Signal Process.*, vol. 149, Feb. 2021, Art. no. 107327, doi: [10.1016/j.ymsp.2020.107327](https://doi.org/10.1016/j.ymsp.2020.107327).
- [36] B. Zhao and Q. Yuan, "Improved generative adversarial network for vibration-based fault diagnosis with imbalanced data," *Measurement*, vol. 169, Feb. 2021, Art. no. 108522, doi: [10.1016/j.measurement.2020.108522](https://doi.org/10.1016/j.measurement.2020.108522).
- [37] X. Hu, H. Zhang, D. Ma, and R. Wang, "A tGAN-based leak detection method for pipeline network considering incomplete sensor data," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021, doi: [10.1109/TIM.2020.3045843](https://doi.org/10.1109/TIM.2020.3045843).
- [38] L. Ma, Y. Ding, Z. Wang, C. Wang, J. Ma, and C. Lu, "An interpretable data augmentation scheme for machine fault diagnosis based on a sparsity-constrained generative adversarial network," *Expert Syst. Appl.*, vol. 182, Nov. 2021, Art. no. 115234, doi: [10.1016/j.eswa.2021.115234](https://doi.org/10.1016/j.eswa.2021.115234).
- [39] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep learning for anomaly detection in time-series data: Review, analysis, and guidelines," *IEEE Access*, vol. 9, pp. 120043–120065, 2021, doi: [10.1109/ACCESS.2021.3107975](https://doi.org/10.1109/ACCESS.2021.3107975).



- [40] B. Wei and D. Zhang, *Learning Control: Applications in Robotics and Complex Dynamical Systems*. Cambridge, MA, USA: Elsevier, 2021.
- [41] S. Liu, H. Ji, and M. C. Wang, "Nonpooling convolutional neural network forecasting for seasonal time series with trends," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 8, pp. 2879–2888, Aug. 2020, doi: [10.1109/TNNLS.2019.2934110](https://doi.org/10.1109/TNNLS.2019.2934110).
- [42] A. Bhukya and C. Koley, "Bi-long short-term memory networks for radio frequency based arrival time detection of partial discharge signals," *IEEE Trans. Power Deliv.*, vol. 37, no. 3, pp. 2024–2031, Jun. 2022, doi: [10.1109/TPWRD.2021.3102937](https://doi.org/10.1109/TPWRD.2021.3102937).
- [43] P. C. Hughes, *Spacecraft Attitude Dynamics*. Mineola, NY, USA: Dover, 2004.
- [44] C. Brown, *Elements of Spacecraft Design*. Reston, VA, USA: Wren Software, 2002.
- [45] B. Bialke, "High fidelity mathematical modeling of reaction wheel performance," in *Proc. 21st Annu. Rocky Mountain Guidance Control Conf.*, Breckenridge, CO, USA, 1998, pp. 483–496.
- [46] A. Rahimi, K. D. Kumar, and H. Alighanbari, "Fault estimation of satellite reaction wheels using covariance based adaptive unscented Kalman filter," *Acta Astronaut.*, vol. 134, pp. 159–169, May 2017, doi: [10.1016/j.actaastro.2017.02.003](https://doi.org/10.1016/j.actaastro.2017.02.003).
- [47] A. N. Okon, S. E. Adewole, and E. M. Uguma, "Artificial neural network model for reservoir petrophysical properties: Porosity, permeability and water saturation prediction," *Model. Earth Syst. Environ.*, vol. 7, pp. 2373–2390, Oct. 2020, doi: [10.1007/s40808-020-01012-4](https://doi.org/10.1007/s40808-020-01012-4).
- [48] K. Fukushima and S. Miyake, "Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position," *Pattern Recognit.*, vol. 15, no. 6, pp. 455–469, Dec. 1982, doi: [10.1016/0031-3203\(82\)90024-3](https://doi.org/10.1016/0031-3203(82)90024-3).
- [49] U. Michelucci, *Applied Deep Learning: A Case-Based Approach to Understanding Deep Neural Networks*. Dübendorf, Switzerland: Apress, 2018, doi: [10.1007/978-1-4842-3790-8](https://doi.org/10.1007/978-1-4842-3790-8).
- [50] Y. Sun, H. Zhang, T. Zhao, Z. Zou, B. Shen, and L. Yang, "A new convolutional neural network with random forest method for hydrogen sensor fault diagnosis," *IEEE Access*, vol. 8, pp. 85421–85430, 2020, doi: [10.1109/ACCESS.2020.2992231](https://doi.org/10.1109/ACCESS.2020.2992231).
- [51] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017, doi: [10.1016/j.measurement.2017.07.017](https://doi.org/10.1016/j.measurement.2017.07.017).
- [52] T. Vignesh, K. K. Thyagarajan, R. B. Jeyavathana, and K. V. Kanimozhi, "Land use and land cover classification using recurrent neural networks with shared layered architecture," in *Proc. Int. Conf. Comput. Commun. Informat. (ICCCI)*, Jan. 2021, pp. 1–6, doi: [10.1109/ICCCI50826.2021.9402638](https://doi.org/10.1109/ICCCI50826.2021.9402638).
- [53] A. Graves, *Supervised Sequence Labelling With Recurrent Neural Networks*. Heidelberg, Germany: Springer, 2012.
- [54] R. Venkatesan and B. Li, *Convolutional Neural Networks in Visual Computing*. London, U.K.: CRC Press, 2018.
- [55] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, 2019, doi: [10.1109/ACCESS.2019.2923006](https://doi.org/10.1109/ACCESS.2019.2923006).
- [56] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [57] Y. Zhang, R. Xiong, H. He, and M. G. Pecht, "Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5695–5705, Jul. 2018, doi: [10.1109/TVT.2018.2805189](https://doi.org/10.1109/TVT.2018.2805189).
- [58] B. Zhao, C. Cheng, Z. Peng, X. Dong, and G. Meng, "Detecting the early damages in structures with nonlinear output frequency response functions and the CNN-LSTM model," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9557–9567, Dec. 2020, doi: [10.1109/TIM.2020.3005113](https://doi.org/10.1109/TIM.2020.3005113).
- [59] T. H. Mohamad, A. Abbasi, E. Kim, and C. Nataraj, "Application of deep CNN-LSTM network to gear fault diagnostics," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2021, pp. 1–6, doi: [10.1109/ICPHM51084.2021.9486591](https://doi.org/10.1109/ICPHM51084.2021.9486591).
- [60] X. Chen, B. Zhang, and D. Gao, "Bearing fault diagnosis base on multi-scale CNN and LSTM model," *J. Intell. Manuf.*, vol. 32, no. 4, pp. 971–987, Apr. 2021, doi: [10.1007/s10845-020-01600-2](https://doi.org/10.1007/s10845-020-01600-2).
- [61] X. Hu, H. Zhang, D. Ma, and R. Wang, "Hierarchical pressure data recovery for pipeline network via generative adversarial networks," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 3, pp. 1960–1970, Jul. 2022, doi: [10.1109/TASE.2021.3069003](https://doi.org/10.1109/TASE.2021.3069003).
- [62] A.-E.-R. Abd-Elhay, W. A. Murtada, and M. I. Yosof, "A high accuracy modeling scheme for dynamic systems: Spacecraft reaction wheel model," *J. Eng. Appl. Sci.*, vol. 69, no. 1, Jan. 2022, doi: [10.1186/s44147-021-00056-2](https://doi.org/10.1186/s44147-021-00056-2).
- [63] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, pp. 425–436, Jan. 2017, doi: [10.3390/s17020425](https://doi.org/10.3390/s17020425).
- [64] R. Yu, Y. Wang, Z. Zou, and L. Wang, "Convolutional neural networks with refined loss functions for the real-time crash risk analysis," *Transp. Res. C, Emerg. Technol.*, vol. 119, Oct. 2020, Art. no. 102740, doi: [10.1016/j.trc.2020.102740](https://doi.org/10.1016/j.trc.2020.102740).
- [65] M. Khodayar, G. Liu, J. Wang, O. Kaynak, and M. E. Khodayar, "Spatiotemporal behind-the-meter load and PV power forecasting via deep graph dictionary learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 10, pp. 4713–4727, Oct. 2021, doi: [10.1109/TNNLS.2020.3042434](https://doi.org/10.1109/TNNLS.2020.3042434).
- [66] M. Saffari, M. Khodayar, and M. Teshnehlab, "Random weights rough neural network for glaucoma diagnosis," in *Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery (Lecture Notes on Data Engineering and Communications Technologies)*, vol. 89, Q. Xie, L. Zhao, K. Li, A. Yadav, and L. Wang, Eds. Cham, Switzerland: Springer, 2021, doi: [10.1007/978-3-030-89698-0\\_55](https://doi.org/10.1007/978-3-030-89698-0_55).
- [67] Z. Xu, C. Li, and Y. Yang, "Fault diagnosis of rolling bearing of wind turbines based on the variational mode decomposition and deep convolutional neural networks," *Appl. Soft Comput.*, vol. 95, Oct. 2020, Art. no. 106515, doi: [10.1016/j.asoc.2020.106515](https://doi.org/10.1016/j.asoc.2020.106515).
- [68] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770–2779, Dec. 2017, doi: [10.1109/TII.2017.2730846](https://doi.org/10.1109/TII.2017.2730846).
- [69] X. Wu, Z. Peng, J. Ren, C. Cheng, W. Zhang, and D. Wang, "Rub-impact fault diagnosis of rotating machinery based on 1-D convolutional neural networks," *IEEE Sensors J.*, vol. 20, no. 15, pp. 8349–8363, Aug. 2020, doi: [10.1109/JSEN.2019.2944157](https://doi.org/10.1109/JSEN.2019.2944157).
- [70] Z. Y. Xue, K. S. Xiahou, M. S. Li, T. Y. Ji, and Q. H. Wu, "Diagnosis of multiple open-circuit switch faults based on long short-term memory network for DFIG-based wind turbine systems," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 8, no. 3, pp. 2600–2610, Sep. 2020, doi: [10.1109/JESTPE.2019.2908981](https://doi.org/10.1109/JESTPE.2019.2908981).
- [71] C. Chen, Y. Ma, and G. Ren, "Hyperspectral classification using deep belief networks based on conjugate gradient update and pixel-centric spectral block features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4060–4069, 2020, doi: [10.1109/JSTARS.2020.3008825](https://doi.org/10.1109/JSTARS.2020.3008825).
- [72] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020, doi: [10.1109/ACCESS.2020.2972859](https://doi.org/10.1109/ACCESS.2020.2972859).
- [73] T. Sun, G. Yu, M. Gao, L. Zhao, C. Bai, and W. Yang, "Fault diagnosis methods based on machine learning and its applications for wind turbines: A review," *IEEE Access*, vol. 9, pp. 147481–147511, 2021, doi: [10.1109/ACCESS.2021.3124025](https://doi.org/10.1109/ACCESS.2021.3124025).
- [74] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017, doi: [10.1109/JBHI.2016.2636665](https://doi.org/10.1109/JBHI.2016.2636665).
- [75] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017, doi: [10.1109/ACCESS.2017.2728010](https://doi.org/10.1109/ACCESS.2017.2728010).
- [76] X. Zhao, M. Jia, and Z. Liu, "Semisupervised deep sparse auto-encoder with local and nonlocal information for intelligent fault diagnosis of rotating machinery," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021, doi: [10.1109/TIM.2020.3016045](https://doi.org/10.1109/TIM.2020.3016045).
- [77] R. Fei, J. Sha, Q. Xu, B. Hu, K. Wang, and S. Li, "A new deep sparse autoencoder for community detection in complex networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–25, Dec. 2020, doi: [10.1186/s13638-020-01706-4](https://doi.org/10.1186/s13638-020-01706-4).

- [78] M. Khodayar, J. Wang, and Z. Wang, "Energy disaggregation via deep temporal dictionary learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1696–1709, May 2020, doi: [10.1109/TNNLS.2019.2921952](https://doi.org/10.1109/TNNLS.2019.2921952).
- [79] Y. Ding, L. Ma, J. Ma, C. Wang, and C. Lu, "A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions," *IEEE Access*, vol. 7, pp. 149736–149749, 2019, doi: [10.1109/ACCESS.2019.2947194](https://doi.org/10.1109/ACCESS.2019.2947194).
- [80] W. He, Y. He, and B. Li, "Generative adversarial networks with comprehensive wavelet feature for fault diagnosis of analog circuits," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 9, pp. 6640–6650, Sep. 2020, doi: [10.1109/TIM.2020.2969008](https://doi.org/10.1109/TIM.2020.2969008).



**ABD-ELSALAM R. ABD-ELHAY** received the B.Sc. degree in electronics and communications engineering from Menoufia University, in 1999, and the M.Sc. degree in electronics and communication engineering from Al-Azhar University, Cairo, Egypt, in 2014, where he is currently pursuing the Ph.D. degree in electronics and communication engineering. He is currently a Satellite Systems Engineer with the Egyptian Space Agency (EGSA). His current research interests include optimization, swarm intelligence, Artificial Intelligence (AI), deep learning, satellite attitude control, Fault Detection and Isolation (FDI), and embedded systems design.



**WAEEL A. MURTADA** received the B.Sc. degree in electronics and communications engineering from Benha University, in 1993, the M.Sc. degree in electronics and communications engineering from Cairo University, in 2005, and the Ph.D. degree in electronics and communications from Al-Azhar University, in 2013. He is currently the Head of the Space and Strategic Studies Division, Spacecraft On-Board Computers and Space Software Department, National Authority for Remote Sensing and Space Sciences (NARSS). His research interests include artificial intelligence; real-time embedded software design; embedded hardware design; metaheuristic optimization algorithms; deep learning; data science; and spacecraft fault detection, isolation, identification, and recovery (FDIIR).



**MOHAMED I. YOUSSEF** received the Ph.D. degree from Ruhr University, Bochum, Germany, in 1988. He is a Professor of communications with Al-Azhar University. His research interests include communication systems, digital signal processing, digital filters, and coding techniques.

...