Multi-Dimensional Time-Series Shapelet Based Real-Time Fault Detection and Localization on ISS Electrical Power Distribution System

Qin Liu, *Student Member, IEEE,* Weiran Chen, *Student Member, IEEE,* and Venkata Dinavahi, *Fellow, IEEE*

***Abstract*—International space station (ISS) is a grand invention for human beings to have a chance at exploring the outer space. Its operation is completely dependent on the autonomous power distribution system which transforms energy by solar arrays from the sun. There is a high demand for a reliable monitoring system that can accurately and timely detect and localize faults in its power system for the special working environment of the ISS. In this paper, a fault detection and localization (FDL) based on multi-dimensional time-series trend extracted shapelet (MTES) method was proposed. A fast shapelet discovery was created to accelerate the process of extracting shape features from time-series signals collected from the ISS electrical power distribution system (EPDS). Then the techniques of randomization and information gain were exploited for the further shapelet selection. Finally, multi-dimensional time-series classification for FDL was solved by a designed random forest classifier. The real-time FDL measurement instrument was emulated on the Xilinx® VCU128 FPGA board, while a hardware-in-the-loop (HIL) testing platform was established to verify the effectiveness, execution speed, and accuracy of the MTES method. Comparing with other state-of-the-art data-driven methods, higher accuracy (above 96%) and easier hardware implementation were achieved using MTES.**

***Index Terms*—Electrical power distribution system, fault detection and localization, hardware-in-the-loop tests, multi-dimensional time-series classification, shapelets.**

LIST OF ABBREVIATIONS

BGA Beta Gimbal Assembly
BRAM Block Random Access Memory
DSP Digital Signal Processor
EPCE Electrical Power Consuming Equipment
EPDS Electrical Power Distribution System
ESS Electrical Storage System
FDL Fault Detection and Localization
FF Flip Flop
FFT Fast Fourier Transform
HHT Hilbert-Huang Transform
HIL Hardware-in-the-loop
ISS International Space Station
LUT Lookup Table
LVDC Low-voltage Direct Current
MBSU Main Bus Switching Unit
MG Microgrid
MTES Multi-dimensional Time-series Trend Extracted Shapelet
OSP Optimal Split Point
PP Piecewise Point
PPS Primary Power System
PVM Photovoltaic Power Module
SARJ Solar Alpha Rotary Joint
SAX Symbolic Aggregate approXimation
SPS Secondary Power System
SSU Sequential Shunt Unit
TSC Time-series Classification

This work is supported by the Natural Science and Engineering Research Council of Canada (NSERC). (Qin Liu and Weiran Chen are co-first authors.) (Corresponding author: Qin Liu.)

Q. Liu is with the School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, Beijing, 100044 China (email: 13116338@bjtu.edu.cn)

W. Chen and V. Dinavahi are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, T6G 2V4 Canada (email: weiran4@ualberta.ca, dinavahi@ualberta.ca)

I. INTRODUCTION

Curiosity for the origin of life propels the exploration in outer space. The construction of international space station (ISS) is one of the momentous steps in spaceflight history. The ISS is a multi-nation collaborative project serving as a space environment laboratory for a variety of scientific researches and a low Earth orbit staging base for long-duration missions [2], [3]. The ISS power system is a low-voltage direct current (LVDC) microgrid (MG) supplied by solar arrays. Harsh environment of outer space brings critical challenges to the measurement, instrumentation, monitoring, and power distribution system protection [4], [5]. Any electrical failure could result in unsafe operation and delay in the planned work schedules. For instance, the loss of one Main Bus Switching Unit (MBSU) of the ISS limited the power capacity to 75% of nominal until it could be replaced in 2012 [6]. Therefore, developing an intelligent monitoring system that can detect and localize faults in real-time is crucial for the ISS [7], [8]. Currently, machine learning and data-oriented strategies, which are able to learn intrinsic principles from the measured system dataset, are promising for the field of monitoring, fault detection, and prediction [9]. Artificial neural networks (ANNs) [10] and convolutional neural networks (CNNs) [11] classifiers were applied to identify faults in photovoltaic systems. Aeroengine remaining life was estimated using bidirectional long short-term memory (LSTM) network [12]. A reinforcement-learning-based algorithm was designed

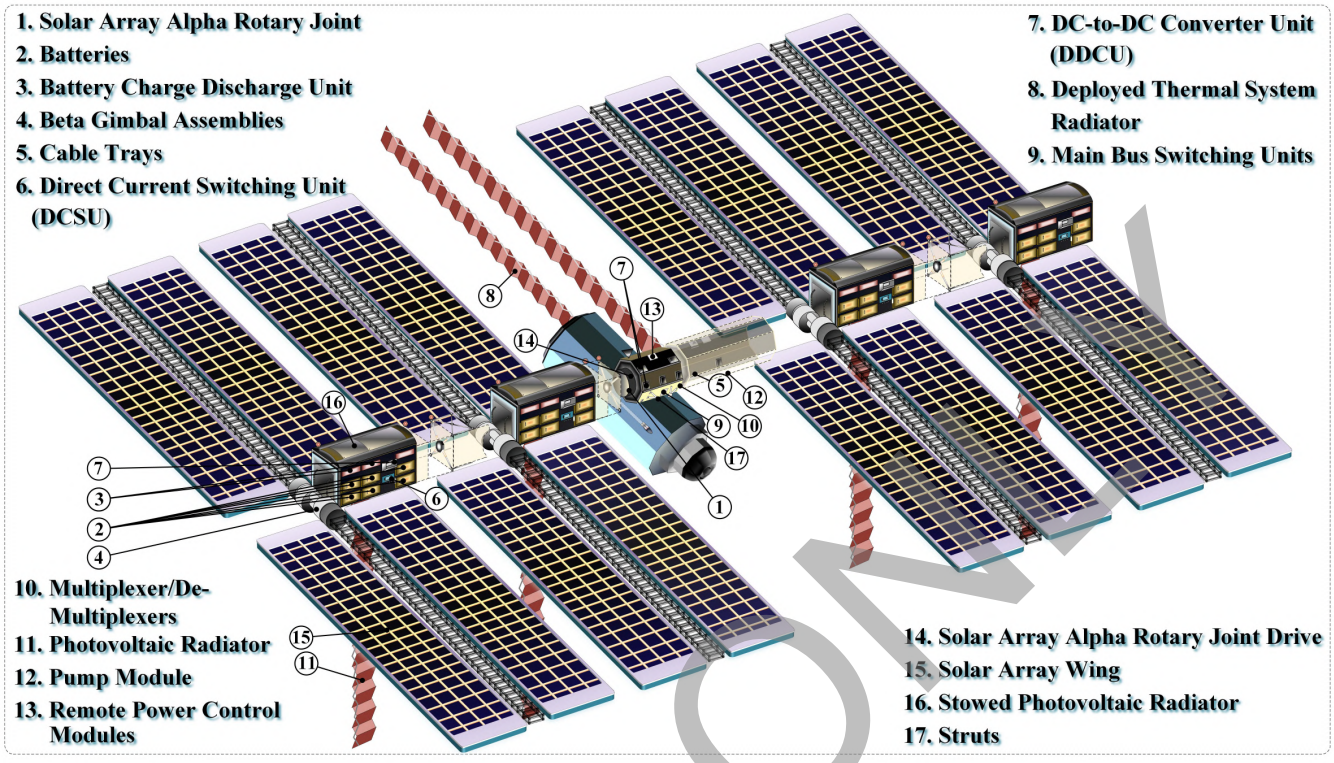

Fig. 1. Overview of the ISS [1].

to make an unmanned aerial vehicles able to adjust its control policy in a dynamic environment [13]. The data collected from the ISS are time series, so the essence of fault detection and localization (FDL) is time-series classification (TSC).

One of key steps of the data-driven TSC is feature engineering which extracts appropriate features from raw data. For time series, the form of features can be time-series, time-domain and frequency-domain as well as shape features. Time-series features are obtained by simply normalizing without complex computations [14], but redundant and disturbing information contained in these features may increase difficulties of learning principles by the FDL method. Time-domain and frequency-domain features are commonly extracted from raw time series using signal processing methods, such as wavelet transform [15], fast Fourier transform (FFT) [16], Hilbert-Huang transform (HHT) [17] and so on. However, signal processing is usually time-consuming and elaborate feature selection is required. Shape features have recently attracted great attention since the concept of shapelet was proposed in 2009 [18]. Shapelet-based methods [19], [20] convey the relationship between inputs and outputs explicitly by seeking out the shape feature representations of each class.

The most challenging step in shapelet-based methods is to discover appropriate shapelets from time series quickly and efficiently. Space search approaches, optimization of objective functions, random selection and so forth have been explored to accelerate the process of shapelet discovery. A symbolic representation using Symbolic Aggregate approXimation (SAX) transformed the time series into a discrete and low-dimensional space, over which shapelet candidates could be searched faster in [21]. Shapelet learning [22], [23] was regarded as a machine learning optimization task. In [24], [25], randomization was applied to reduce the size of the shapelet candidate set to avoid exhaustive searching. Although these methods could give solution to the task, they were not intuitive enough. In addition, most of them were merely applicable to one-dimensional time-series instances. Multiple sensor signals are collected from the ISS power distribution system. Each signal provides complementary information for identifying different faults. These signals form data with multi-dimensional time series which are necessary to be processed simultaneously.

In this paper, a multi-dimensional time-series trend extracted shapelet-based (MTES) method for the FDL of ISS EPDS is proposed. There are three steps: fast shapelet discovery, shapelet selection and classification. In the first step, informative shapelet candidates with multi-trend are elaborately picked out from raw time series data. Randomization and information gain are utilized to find the best top-k shapelets in the second step. Finally, the FDL with multi-dimensional time series is solved by a random forest classifier based on Bayesian

probability features. Besides, correlation-based feature subset selection method [26] is applied to reduce the redundant features (namely associated shapelets) for testing in real time. A hardware-in-the-loop (HIL) real-time emulation platform was established to validate the proposed MTES method. The FDL algorithm is implemented on the Xilinx® VCU128 FPGA board [27], serving as a fault detection instrument, and it was integrated into the VCU118 FPGA board based ISS real-time emulator for FDL. To summarize, the main contributions of this paper are:

- We design a fast shapelet discovery method to intuitively pick out informative shapelet candidates with multi-trend from raw time series data. Piecewise linearization quickly separates a long time series into linear subsequences. Then multiple consecutive subsequences together constitute a multi-trend shapelet candidate. Shapelets merging is utilized to eliminate redundant candidates in advance.
- We select the best top-k shapelets using information gain and randomization.
- We train a random forest classifier based on a Bayesian probability table deduced from multi-dimensional time series. Besides, correlation-based feature subset selection method is applied to reduce the number of shapelets at the testing stage for the FDL in a computationally efficient manner.
- We conduct HIL real-time emulation experiment on FPGA to verify the effectiveness of the real-time FDL application. In addition, comparison with other state-of-the-art methods proved the advantages of the proposed MTES method in both accuracy and resource consumption.

The remainder of this paper is organized as follows: Section II describes the model of the ISS power system and various normal and faulty conditions simulated in Matlab. Section III elaborates the definitions and notations involved in the proposed MTES method. The detailed algorithms and procedures for FDL step-by-step are illustrated in Section IV. Section V provides an elaborate account of the experimental design for HIL testing of the FDL. Section VI presents the discussion of the experimental results including data preparation, selected shapelet, FDL results for the ISS system and comparison with other methods, followed by the conclusions in Section VII.

II. Overview of ISS EPDS

The general electrical power distribution system (EPDS) of ISS employs multi-stage arrangements of power sources, energy storage, control systems, and electronic devices that all operate cooperatively at various system dispositions. Overview of the ISS structure is presented in Figure 1 [1]. For simplicity, only the left half of the ISS EPDS is modeled as in Figure 2 [28], since the EPDS basically adopts a bilaterally symmetric structure. A typical multi-bus channelized electrical distribution system is adopted in MG of ISS, which generally consists of eight basically self-sustained independent power channels, two of which constitute a photovoltaic power module (PVM), coupling via transformers to transport electricity based on requirements and system capacity. Each power channel

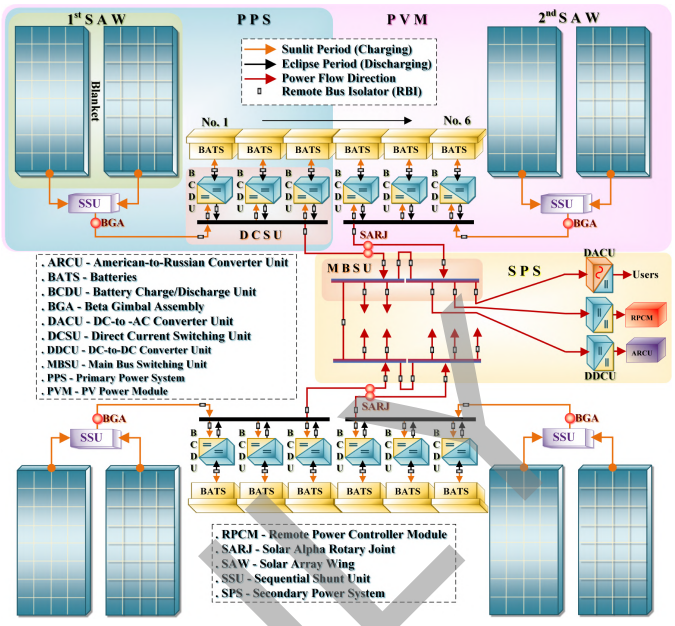

Fig. 2. Detailed half EPDS of the ISS [28].

consists of the primary power system (PPS), solar array wing (SAW), sequential shunt units (SSUs), energy storage system (ESS), secondary power system (SPS), and numerous electronic components to produce, store, and transfer electricity as an isolated source. The PPS features $160\,V$ bus voltage, stores and provides electricity during insolation and eclipse respectively. For tracking the sun position and adjusting the angle as the ISS orbits the Earth, Solar Alpha Rotary Joints (SARJs) and Beta Gimbal Assemblies (BGAs) of the pointing system are designed. The sequential shunt units (SSUs) enable each SAW severally isolated or connected from PPS. The SPS couples with PPS through MBSUs which convert the main bus voltage to $120\,V$ and provide electricity downstream to international segments and loads. In this work, one PVM in ISS is established in Matlab/Simulink [29] as a study case to verify the proposed FDL method.

A. Selected Conditions

Out of a total of 22 conditions, 1 normal and 21 faulty conditions, which may occur in the SAW, PVM, SSU, MBSU, SPS, ESS, pointing system, and electrical power consuming equipments (EPCEs), are modeled, simulated and summarized in Table I. All aforementioned cases might be caused by miss-orientation from sun [30], out of range temperature, shading [31], sensor failures [32], switch position failures [33], loss of components [34], limits condition mismatches [35] etc., which would have serious consequences for ISS's in-orbit life span.

III. Definitions and Notations

A. Time-Series Dataset

$\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$ is a single-dimensional time series (a sensor signal), which contains n temporally ordered values with equal time intervals. Multi-dimensional time series $\boldsymbol{M} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_d\}$ (an instance) are composed of d series. A

TABLE I
CONDITIONS MODELED IN THE ISS EPDS.

Locations	Descriptions	Conditions
None	No fault	F0
SAW	Shading condition	F1
	Temperature out of range ($> 60°$C)	F2
	Temperature out of range ($< -180°$C)	F3
	Solar cells line-to-line fault	F4
	Bypass diode fault	F5
	Degradation	F6
	Solar cells bridging fault	F7
	Maximum power point tracking fault	F8
	Converter open circuit	F9
PVM	Power channel open circuit	F10
SSU	SSU short circuit	F11
MBSU	MBSU open circuit	F12
SPS	SPS open circuit	F13
ESS	Loss of orbital replacement unit (ORU)	F14
	Battery cells line-to-line fault	F15
	Cells balancing fault	F16
	Converter open circuit	F17
Pointing system	SARJ and BGA open circuit	F18
	Sensor failures (control fault)	F19
EPCEs	DDCU open circuit	F20
	DACU open circuit	F21

time-series dataset includes m instances, expressed as $\boldsymbol{S_M} = \{\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_m\}$.

B. Shapelet

Shapelet is a discriminative subsequence of time series belonging to a specific class that can maximally represent the shape feature of the class, proposed by Ye and Keogh for the first time in 2009 [18].

C. Distance Between Time Series and Subsequence

$\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$ is a time series and $\boldsymbol{SX} = \{sx_1, sx_2, \cdots, sx_l\}$ is a subsequence with length l. $\boldsymbol{X}_{k,l} = \{x_k, x_{k+1}, \cdots, x_{k+l-1}\}\,(1 \le k \le n-l+1)$ is a partial series of the time series $\boldsymbol{X}$ with the same length l. Then the distance between $\boldsymbol{X}$ and $\boldsymbol{SX}$ is defined as the minimum of distances between $\boldsymbol{X}_{k,l}$ and $\boldsymbol{SX}$:

$$d\left(\boldsymbol{X}, \boldsymbol{SX}\right) = \min\left(D\left(\boldsymbol{X}_{k,l}, \boldsymbol{SX}\right)\right). \tag{1}$$

The distance between $\boldsymbol{X}_{k,l}$ and $\boldsymbol{SX}$ is given as:

$$D\left(\boldsymbol{X}_{k,l}, \boldsymbol{SX}\right) = \sqrt{\frac{1}{l}\sum_{i=1}^{l}\left(x_i - sx_i\right)^2}. \tag{2}$$

D. Entropy

A time series dataset $\boldsymbol{S_M}$ contains two classes (A and B) of instances. If the proportion of subset made up of class A is $p\left(A\right)$ and the proportion of class B is $p\left(B\right)$, then the entropy of $\boldsymbol{S_M}$ is expressed as:

$$E\left(\boldsymbol{S_M}\right) = -p\left(A\right)\log\left(p\left(A\right)\right) - p\left(B\right)\log\left(p\left(B\right)\right). \tag{3}$$

E. Information Gain

Suppose that a time series dataset $\boldsymbol{S_M}$ is split into two subsets $\boldsymbol{S_{M1}}$ and $\boldsymbol{S_{M2}}$, the information gain is computed as:

$$G\left(\boldsymbol{S_M}\right) = E\left(\boldsymbol{S_M}\right) - \frac{\left|\boldsymbol{S_{M1}}\right|}{\left|\boldsymbol{S_M}\right|}E\left(\boldsymbol{S_{M1}}\right) - \frac{\left|\boldsymbol{S_{M2}}\right|}{\left|\boldsymbol{S_M}\right|}E\left(\boldsymbol{S_{M2}}\right). \tag{4}$$

F. Optimal Split Point

For notation ease, we assume that $\boldsymbol{S_M}$ is a single-dimensional time series dataset here. Calculate distances between the shapelet $\boldsymbol{SX}$ and each series in $\boldsymbol{S_M}$ and sort them in non-descending order as:

$$OrderList = \left\{d_1, d_2 \cdots d_m\right\}, d_1 \le d_2 \le \cdots \le d_m. \tag{5}$$

Obtain split points $SP(i)$ by computing thresholds $d_{SP(i)} = (d_i + d_{i+1})/2$, $i = 1, 2, \cdots, m-1$. Then $\boldsymbol{S_M}$ can be split into two subsets according to the $SP(i)$, i.e. the series satisfying $d \le d_{SP(i)}$ belong to $\boldsymbol{S_{M1}}$, otherwise belong to $\boldsymbol{S_{M2}}$. Information gains of all the $SP(i)$ can be calculated and the one with maximum information gain is defined as the optimal split point (OSP).

G. Trend Feature

Suppose that the linear regression equation $y = kx + b$ can fit a piece of time series. The discretized slope k is regarded as the trend feature of the time series. The discretized slope k is derived by the rule as follows: if the fitting line parallels with x-axis, $k = 0$; if the angle between the line and x-axis in $(10°, 20°)$, $k = 2$; if the angle in $(-20°, -10°)$, $k = -2$, and so on. Therefore, the set of k is $\{k | k = -9, -8, \cdots, 0, \cdots 8, 9\}$.

H. Overall Significance of Linear Regression Equation

Linear regression is used to fit a time series. Only when the correlation coefficient R^2 and P-value of F-test satisfy $R^2 \ge 0.5$ and $P \le 0.05$ [36], the overall significance level can achieve. Otherwise, we call the linear regression equation is not statistically significant.

I. Piecewise Point

When the linear regression for a time series can not reach the significance level, piecewise linearization is required to separate the series into several sections. The point separating two sections is called piecewise point (PP). The mathematical definition is the point having the maximum of absolute value of the differences between the series ($\boldsymbol{X} = \{x_1, x_2, \cdots, x_n\}$) and regression line ($\boldsymbol{Y} = \{y_1, y_2, \cdots, y_n\}$):

$$\left|y_{PP} - x_{PP}\right| = \max\left(\left|\mathbf{Y} - \mathbf{X}\right|\right). \tag{6}$$

J. Informatization Representation of Shapelet

In this work, a shapelet is composed of multiple sequential subsequences with different trend features. Assume that the trend feature vector is $\boldsymbol{K} = \{k_1, k_2, k_3\}$, the length vector is $\boldsymbol{L} = \{l_1, l_2, l_3\}$ and the data point vector is $\boldsymbol{X} = \{x_1, x_2, \cdots, x_{l_1+l_2+l_3}\}$, then we represent the shapelet as $\boldsymbol{IX} = \{\boldsymbol{K}, \boldsymbol{L}, \boldsymbol{X}\}$ (i.e. informatization representation of the shapelet).

K. Similar Shapelets

$\boldsymbol{IX}_i = \{\boldsymbol{K}_i, \boldsymbol{L}_i, \boldsymbol{X}_i\}$ and $\boldsymbol{IX}_j = \{\boldsymbol{K}_j, \boldsymbol{L}_j, \boldsymbol{X}_j\}$ are two shapelets. If $\boldsymbol{K}_i = \boldsymbol{K}_j$ and $\boldsymbol{L}_i = \boldsymbol{L}_j$, then we call $\boldsymbol{IX}_i$ and $\boldsymbol{IX}_j$ are similar, denoted as $\boldsymbol{IX}_i \sim \boldsymbol{IX}_j$.

L. Shapelets Merging

If u shapelets are similar $\boldsymbol{IX}_i \sim \boldsymbol{IX}_j \sim \ldots \sim \boldsymbol{IX}_p$, i.e. $\boldsymbol{K}_i = \boldsymbol{K}_j = \cdots = \boldsymbol{K}_p = \boldsymbol{K}$ and $\boldsymbol{L}_i = \boldsymbol{L}_j = \cdots = \boldsymbol{L}_p = \boldsymbol{L}$, then we will merge them into a new shapelet $\boldsymbol{IX}_{fuse} = \{\boldsymbol{K}, \boldsymbol{L}, \boldsymbol{X}_{fuse}\}$, where $\boldsymbol{X}_{fuse} = \left\{x_1^{fuse}, x_2^{fuse}, \cdots, x_{l_1+l_2+l_3}^{fuse}\right\}$, $x_q^{fuse} = \left(x_q^i + x_q^j + \cdots + x_q^p\right)/u$, $q = 1, 2, \cdots, l_1 + l_2 + l_3$.

M. Shapelet Information Integration

$\boldsymbol{IX} = \{\boldsymbol{K}, \boldsymbol{L}, \boldsymbol{X}\}$ is a shapelet, where $\boldsymbol{K} = \{k, k, k\}$ and $\boldsymbol{L} = \{l_1, l_2, l_3\}$. The trend features of the subsequences made up of the shapelet are the same, so we integrate its information and re-write it as $\boldsymbol{IX} = \{k, l_1 + l_2 + l_3, \boldsymbol{X}\}$.

N. Time-Series Splitting

$\boldsymbol{SX} = \{k, l, \boldsymbol{X}\}$ is the information representation of a time series. We will split it into three subsequences when necessary as $\boldsymbol{SX}_1 = \{k, l/3, \boldsymbol{X}_1\}$, $\boldsymbol{SX}_2 = \{k, l/3, \boldsymbol{X}_2\}$ and $\boldsymbol{SX}_3 = \{k, l/3, \boldsymbol{X}_3\}$, where $\boldsymbol{X}_1 = \left\{x_1, x_2, \cdots, x_{l/3}\right\}$, $\boldsymbol{X}_2 = \left\{x_{l/3+1}, x_{l/3+2}, \cdots, x_{2l/3}\right\}$, $\boldsymbol{X}_3 = \left\{x_{2l/3+1}, x_{2l/3+2}, \cdots, x_l\right\}$. These subsequences have the same trend feature as the original time series, $1/3$ of the length and equally distributed data points.

IV. Proposed Method

The essence of the FDL for the ISS power system is multi-dimensional TSC since the sensor signals collected from the system are time series data. Shapelet-based methods have recently attracted great attention in the field of TSC. Most machine learning methods establish the intrinsic rules between input time series and output classes implicitly, but shapelet-based methods can present the relationship explicitly by seeking out the shape features corresponding to each class. In this work, a multi-dimensional time-series trend extracted shapelet-based method to solve the FDL problem for the ISS is proposed. Details of the method are presented in this section.

A. Fast Shapelet Discovery

The most important but time-consuming step of shapelet-based methods is collecting appropriate shapelet candidates from raw time series dataset, called the shapelet discovery procedure. In the MTES method, a shapelet discovery approach is proposed, which can speedily pick out contributing shape features and narrow the range of shapelet candidates. It includes three steps: time series piecewise linearization, multi-trend extracted based shapelet discovery and shapelet merging. At the training stage, the goal is to find shapelets of each sensor signal, so only single-dimensional time series shapelet extraction is considered at this point, while the multi-dimensional time series will be involved in Section III E at the testing stage.

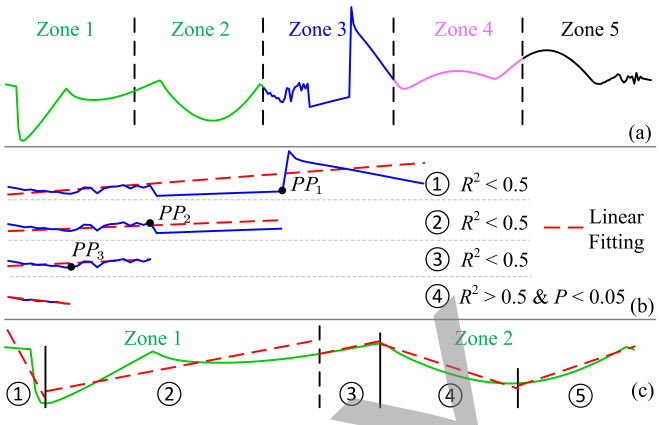

Fig. 3. Time-series piecewise linearization and multi-trend extracted based shapelet discovery step.

1) Time Series Piecewise Linearization: Trend changing is regarded as the shape feature (or trend feature) of time series. Piecewise linearization is utilized to analyze these features. First, a time series is equally distributed into five zones. Second, linear regression is implemented in each zone respectively. When the linear regression equation is not statistically significant, a PP is selected and new linear regression is applied on the section divided by the PP. According to the above division principle, each zone can be separated into several pieces by PPs. One piece of the subsequence, named linear subsequence, can be fit by a linear equation with information of trend feature k, length l and data points $\boldsymbol{X}$. An example of piecewise linearization step in Zone 3 is depicted in Figure 3 (a) and (b). Finally, all the linear subsequences together recombine the original time series.

2) Multi-Trend Extracted Based Shapelet Discovery: Shapelet is a discriminative subsequence of time series belonging to a specific class that can maximally represent the shape feature of the class, so it should not be a straight line segment. In this paper, multiple sequential linear subsequences are connected to be selected as a shapelet candidate because adjacent linear subsequences usually have different trend features. For instance, the first two zones of the time series are separated into 5 pieces as shown in Figure 3 (c). If the number of linear sequences comprising a shaplet candidate is set to 3, then linear subsequences ①②③, ②③④ and ③④⑤ are respectively connected as three candidates.

3) Shapelets Merging in One Class: In the training dataset, we assume that all instances belonging to the same class should have similar shapelets. To find these similar shapelets, one instance is randomly chosen and shapelet candidates are extracted through the aforementioned steps. Suppose that $\boldsymbol{IX}_1$ is a shapelet, then the similar shapelet $\boldsymbol{IX}$ in other instance $\boldsymbol{X}$ is the specific subsequence $\boldsymbol{SX}$. The distance between $\boldsymbol{SX}$ and $\boldsymbol{IX}_1$ equals the distance between $\boldsymbol{X}$ and $\boldsymbol{IX}_1$, expressed as:

$$\boldsymbol{IX} = \{\boldsymbol{SX} | D(\boldsymbol{SX}, \boldsymbol{IX}_1) = d(\boldsymbol{X}, \boldsymbol{IX}_1)\}. \quad (7)$$

Finally, shapelet merging is applied to all the similar shapelets belonging to the same class and the set of candidates is obtained.

The proposed fast shapelet discovery approach based on multi-trend extraction is summarized in Algorithm 1.

Algorithm 1 Fast shapelet discovery approach based on multi-trend extraction

Input: training dataset DS_{train}

for i = 1:N_{class} **do**
  X_1 = RandomOneSample($DS_{\text{train}}(i)$)
  for j = 1:N_{signal} **do**
    SX_1= PiecewiseLinearization($X_1(j)$)
    SC_1 = ShapeletDiscovery(SX_1)
    for p = 1:size(SC_1) **do**
      for q = 2:N_{sample} **do**
        X = AnotherSample($DS_{\text{train}}(i)$)
        IX_1 = $SC_1(p)$
        $IX(q)$ = $\{SX|D(SX, IX_1)$ = $d(X(j), IX_1)\}$
    $SC(i,j,p)$ = ShapeletMerge(IX_1, IX)
return the set of shapelet candidates SC

B. Key Points of Fast Shapelet Discovery

The main points for appropriately applying the proposed fast shapelet discovery approach are emphasized in this subsection.

1) Minimum Length of Shapelet: When the time series has high frequency fluctuation, the linear subsequences separated by piecewise linearization may contain too small number of data points, which leads to shapelet candidates with not enough data information. For example, the time series below in Figure 4 (a) is divided into 12 small pieces marked in different colors. The three linear subsequences, pointed by the arrows, are supposed to be connected as a shapelet, but each of these subsequences only includes 2 data points and thus the shapelet has simply 4 points in total, which is definitely not enough. Therefore, we set the minimum length of a shapelet. If the length cannot reach the minimum level, more linear subsequences will be combine to form a shapelet, as the shapelet candidate above in Figure 4 (a).

2) Maximum Length of Shapelet: By contrast, when the time series is too smooth, the linear subsequence may contain too many data points, which causes too long shapelet. To solve this problem, firstly, the entire time series is equally distributed into five zones. Secondly, we set the maximum length of a linear sequence that should less than 2/3 of the zone length. Thus the maximum length of a shapelet can be guaranteed less than 2/5 of the whole series. Although the maximum length, which should be considered on a case-by-case basis, may be exclusively suitable for this study, the similar rule could be applied for other cases. In Figure 4 (b), the linear regression for the subsequence in zone 4 can reach significance level. In this case, time-series splitting is used to further separate it into 3 smaller linear subsequences marked in different colors.

3) Eliminating Redundancy: Although shapelet 1 and shapelet 2 in Figure 5 are not exactly the same, they have similar shape or trend, so it is not necessary to keep both of them in the candidate set. To eliminating this kind of redundancy, we use shapelets merging to create a new shapelet,

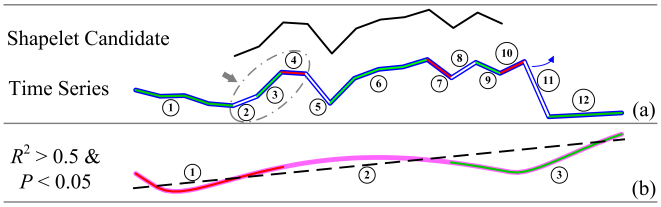

Fig. 4. Shapelet discovery from smooth time series.

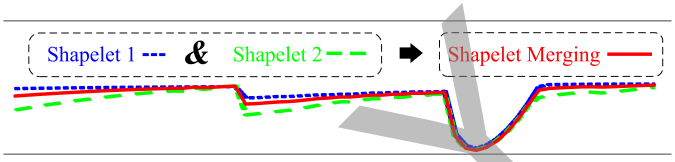

Fig. 5. Merging of similar shapelets.

like the one on the bottom in figure, as a representation of all similar shapelets.

4) Coordinate Transformation: The criteria to distinguish whether a time series belongs to the class, from which a shapelet is extracted, is whether the distance between the series and shapelet is less than the preset threshold. However, even though the shapes or trends of time series look similar, the ranges may have large discrepancy, which could result in misclassification of the similar time series if directly comparing the distances using raw data. Therefore, coordinate transformation is implemented to normalize the range of subsequences of the time series and the shapelet in [0,1] before computing the distances, as displayed in Figure 6.

C. Shapelet Selection

The shapelet candidate set is acquired after preceding the fast shapelet discovery. Then the best top-k shapelets for each class are required to be further selected. Three steps for this purpose are describing as follows.

1) Determination of Distance Threshold: The distance threshold δ for each shapelet is extremely important. If the distance between the time series of a instance and the shapelet from a specific class is not larger than δ, it is more likely that the instance belongs to the class. Otherwise, it should belong to other classes. We set the threshold on the optimal split point (i.e. $\delta = d_{\text{OSP}}$) and preserve the maximum information gain of the shapelet.

2) Calculation of Bayesian Probability: In the training dataset, Bayesian probability can provide the probability of correct classification of a shapelet according to its distance threshold. Two events are defined as:

Event Y: the distance between a time series and a shapelet is not larger than δ.

Event N: the distance between a time series and a shapelet is larger than δ.

Suppose that S is the instance space, $F_1, F_2, \cdots, F_n$ are diverse classes among S. On the basis of the Bayes' theorem,

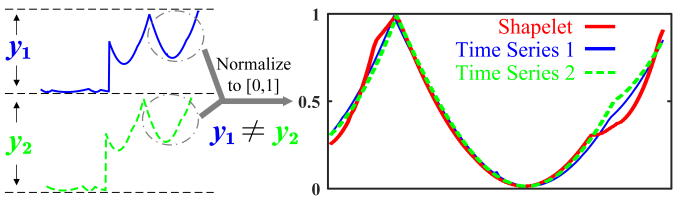

Fig. 6. Coordinate transformation for similarity computation.

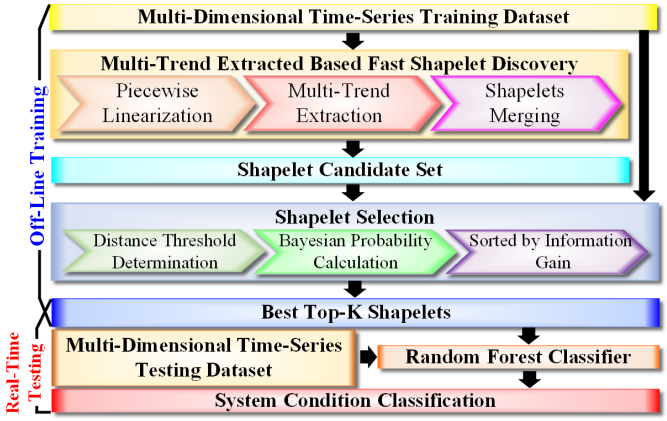

Fig. 7. Proposed multi-dimensional time-series trend extracted shapelet-based FDL method for ISS EPDS.

the probabilities of the instance belonging to a class when the event Y or N occurs are:

$$P\left(F_i|Y\right) = \frac{P\left(Y|F_i\right)P\left(F_i\right)}{\sum_{j=1}^{n} P\left(Y|F_j\right)P\left(F_j\right)}, \tag{8}$$

$$P\left(F_i|N\right) = \frac{P\left(N|F_i\right)P\left(F_i\right)}{\sum_{j=1}^{n} P\left(N|F_j\right)P\left(F_j\right)}. \tag{9}$$

These probabilities are kept as information with a shapelet and will be exploited at the subsequent testing stage.

3) Best Top-K Shapelet Selection: To further accelerate the training speed, shapelets are randomly sampled from the candidate set with upper-limit quantity. Then these sampled shapelets are sorted in descending order in terms of the information gain, and the front k shapelets are the best top-k selection. The procedure of the shapelet selection is explained in Algorithm 2.

D. Random Forest Classifier Based on Bayesian Probabilities for Multi-dimensional Time Series Classification

In this subsection, a random forest classifier is designed to solve multi-dimension time series classification at the testing stage. Taking one instance as an example. First, distances between sensor signal time series and corresponding top-k shapelets are calculated. Second, Bayesian probabilities are obtained in accordance to the relationship between distances and thresholds. Third, the average of k Bayesian probabilities represents the probability that the instance belongs to a specific class with respect to each signal. At this point, a Bayesian

Algorithm 2 Best top-k shapelets selection

Input: training dataset DS_{train}, shapelet candidates set SC, and upper-limit of randomly sampled $TotalShapelets$

for j = 1:N_{signal} **do**
  $X = DS_{\text{train}}(i)$
  for i = 1:N_{class} **do**
    N_{shapelet} = size$(SC(i,j))$
    $p = 0$
    while $p < N_{\text{shapelet}}$ && $p < TotalShapelets$ **do**
      $p++$
      IX = RandomOneShapelet$(SC(i,j))$
      for q = 1:N_{sample} **do**
        $Distances = d(X(q), IX)$
      $Gain(p)$, $\delta(p)$ = FindOSP$(Distances)$
      $Probabilities(p)$
        = BayesianProbability$(\delta, Distances)$
      $Quality$ = AssessCandidate$(Gain)$
      $KbestShapelets(i,j)$
        = AddShapelet$(IX, Quality)$
return the best top-k shapelets $KbestShapelets$,
  distance threshold of each shapelet δ,
  information gain $Gain$,
  and Bayesian probability $Probabilities$

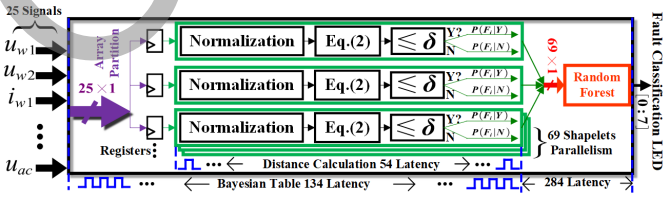

Fig. 8. Top-level fault measuring system FPGA hardware execution latency.

probability table with the size of $N_{\text{class}} \times N_{\text{signal}}$ is acquired. Finally, each probability element in the table is treated as a single feature for the random forest classifier, which is exploited to solve the multi-dimensional FDL problem. While training the classifier, the worth of features are evaluated by considering the individual predictive ability of each feature in accordance with the correlation-based feature subset selection [26] approach. The selected features (namely associated shapelets) excluding redundancy are retained at the subsequent testing stage.

E. Multi-Dimensional Time-Series Trend Extracted Shapelet-Based Method

In summary, the proposed MTES method for FDL of the ISS power system is illustrated in Figure 7. First, a multi-trend extracted based fast shapelet discovery approach is implemented to collected the set of shapelet candidates. Second, the best top-k shapelets for each class is further selected in terms of information gain. Finally, the system condition is evaluated by the random forest classifier based on Bayesian probabilities.

V. Comprehensive Hardware-in-the-Loop Experimental Setup for the Proposed FDL Method

This section presents the validation of the proposed FDL algorithm through HIL real-time emulation testing. The hardware platform used for testing consists of two Xilinx® FPGA boards. The first board, VCU118 [37], hosts the mathematical model of the ISS, enables it in real-time and provides realistic virtual stimuli for the measurement hardware. The second board, VCU128 [38], serves as the measuring instrument for implementing the proposed algorithm. It collects data from the emulator and provides fault classification information at the user observation terminal.

A. ISS System Emulator

The ISS real-time emulation platform is deployed on the VCU118, consisting of a substantial allocation of computational resources including 1,182,240 LUTs, 2,364,480 FFs, and 6,840 DSPs. The electrical components of the ISS, such as PVs, switches, and converters, are described by mathematical models in the form of ordinary differential equations (ODEs) in state-space representation. The nonlinear elements are solved using the implicit Trapezoidal integration rule, with a prescribed integration step-size of 10μs. The proposed model adheres to the electrical topology and voltage bus specifications of the ISS [29]. It utilizes real solar radiation data as the excitation source and generates signals in real-time, employing the IEEE single-precision floating-point format. These signals are then transmitted to the fault detector for analysis and detection. Two external interfaces are employed by the emulator, namely FPGA mezzanine card (FMC) and dual small form-factor pluggable (QSFP), as illustrated on the left side of Figure 9. The FMC interface converts discrete simulation signals into analog signals through a digital to analog (DAC) adapter, which are then displayed on the Tektronix® DPO 7054 oscilloscope for real-time observation of key circuit voltage and current waveforms by the user. Simultaneously, the circuit signals are stored in random access memory (RAM) and transmitted to the faults measuring instrument via an appropriate transmission control protocol/internet protocol (TCP/IP), which will be elaborated in the next section.

B. Fault Measuring Instrument

The fault measurement instrument, also known as the FDL monitor, is implemented on the Xilinx® VCU128 FPGA board, featuring 4,032 BRAMs, 9,024 DSPs, 2,607,360 FFs, and 1,303,680 LUTs. In this context, BRAM is employed for the retention of substantial data volumes, DSP is harnessed for digital signal processing, FF serves the purpose of state retention and control, and LUT finds application in the implementation of logical functions. These resources collectively constitute the foundational components of an FPGA, empowering it to facilitate a diverse range of applications spanning digital logic and signal processing. The latency and resource consumption of the real-time FDL algorithms are summarized in Table II. The total latency for computing a 500-step instance is less than a sampling time (i.e. 10μs), which means that the real-time FDL is able to be realized with FPGA acceleration. Figure 8 illustrates the top-level algorithmic diagram of the fault measurement system. The system comprises 25 signal inputs, and 69 distinct shapelets are evaluated as worthy features for the classifier. Following signal normalization, Bayesian probability (8) or (9) is obtained based on the calculated distances using (2). These 69 computations are independent, thus permitting concurrent processing, yielding 69 outputs and consuming 54 delays. Thus far, encompassing signal input, distance computation, and result output, the complete Bayesian table process consumes a total of 134 delays. The 69 inputs are directed to the random forest for logical decision-making, consuming 284 delays. Ultimately, an 8-bit binary is outputted to illuminate LEDs, effectuating fault detection.

The ISS emulator and FDL monitor communicate and exchange data through the QSFP interface. Both boards, belonging to the Xilinx® UltraScale™ series, feature identical interfaces, with 2-lane and 4-lane options available respectively, facilitating scalable interconnection. The communication protocol used is the Aurora LogiCORE™ IP core, which facilitates seamless integration of Xilinx® transceivers. The selected Aurora 64B/66B is a lightweight, scalable, and open-access link-layer protocol used for high-speed serial communication, providing a low-cost, high data-rate, and flexible solution for building serial data channels with throughput from 500 Mbps to over 400 Gbps [39].

The top-level interface for HIL testing of FDL is as follows. For the data transmission/acquisition end (VCU118), it consists of two four-channel QSFP interfaces, with each channel wired to a gigabit transceiver (GTY). Each clock rising edge can transmit a maximum of $2 \times 4 \times 64$ bits of data to the receiver. Due to the numerous detection signals (25×32 bits), all signals are written to RAM for storage, and then read and sent via the Aurora IP. For the receiver end (VCU128), it includes four four-channel QSFP interfaces. The data collected through QSFP-GTY is written to the RAM. Once all the required signals are ready within one sampling interval, the deployed algorithm reads the signals to achieve FDL. As for the user observation end, the 8-bit width general purpose input/outputs (GPIO) LED are utilized, theoretically capable of displaying 2^8 different classification results, sufficient to demonstrate the 22 different faults designed in this experiment. The above process and the entire HIL FPGAs integration platform is illustrated in Figure 9. The efficiency (time & resource consumption) of this FDL topology largely depends on the Aurora protocol and the length of user-transmitted data. According to [39], the user data bytes of this design is $100 < 800 < 1000$, with framing efficiency ranging from 96.12% to 99.18%. The resource consumption and latency requirements for Aurora IP are shown in Table II.

VI. Experimental Results and Discussion

The experimental results for the ISS EPDS are elaborated to verify the proposed MTES based FDL method in this section. Training and testing dataset are both collected from the established Matlab model. As listed in Table I, 22 conditions in total are considered to be classified.

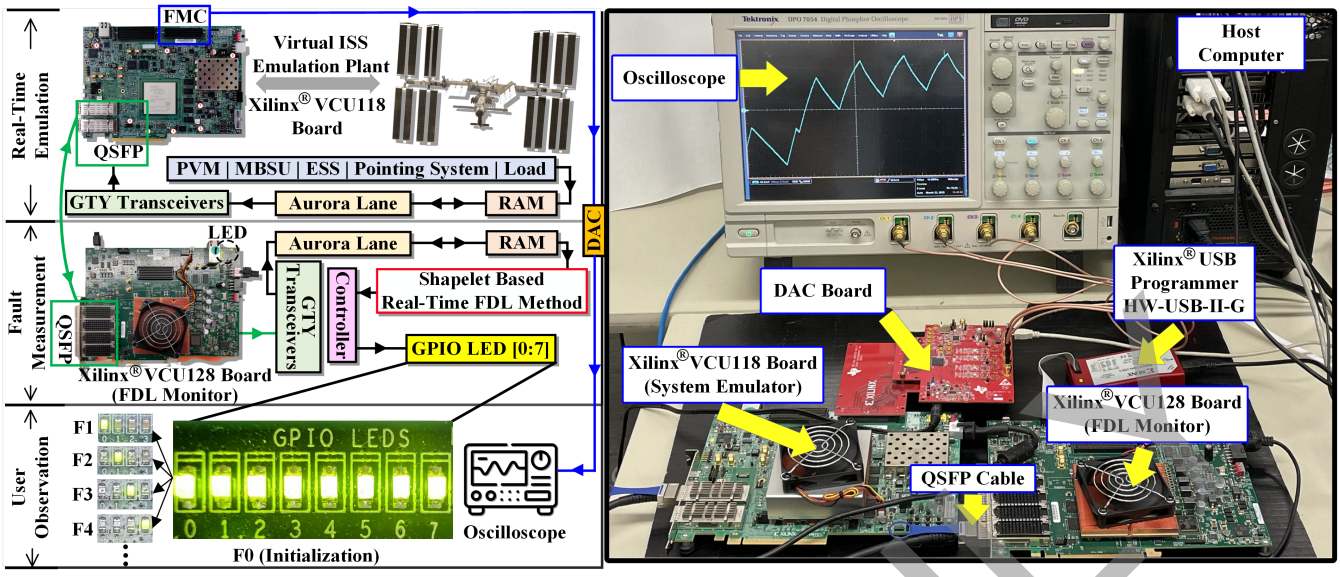

Fig. 9. Hardware-in-the-loop testing for fault measuring design of the ISS system implemented on multiple FPGA-based instruments.

TABLE II
SPECIFICS OF HARDWARE RESOURCE CONSUMPTION AND LATENCY.

	BRAM	DSP	FF	LUT	Latency (µs)
ISS Emulator Xilinx® VCU118 FPGA					
Total	6.22%	40.48%	10.08%	24.52%	10
FDL Monitor Xilinx® VCU128 FPGA					
Bayesian Table	29	4	5,502	80,248	1.34
Random Forest	4	0	2,808	59,388	2.84
Total	33 (0%)	0 (0%)	8,310 (0%)	139,636 (11%)	4.18
Xilinx® Aurora Protocol					
	0	0	2,000	7,000	4

TABLE III
SIGNALS FOR EACH SAMPLE DATA COLLECTED FROM SIMULATED MODEL.

Location	Signals	Tabs
PVM	Output voltage of 1_{st} SAW	u_{w1}
	Output voltage of 2_{nd} SAW	u_{w2}
	Output current of 1_{st} SAW	i_{w1}
	Output current of 2_{nd} SAW	i_{w2}
	Total output current	i_{pvm}
	Output power difference with irradiance input power	p_{diff}
MBSU	Voltage of 120V bus	u_{120}
	Voltage of 160V bus	u_{160}
ESS	Output voltage of six BATS	$u_{b1\sim6}$
	Output current of battery group	i_{bg}
	Branch current	i_{bat}
	Total output voltage	u_{bat}
Pointing system	Output voltage of motor	u_{pmotor}
	Output current of motor	i_{pmotor}
	Branch current	i_{point}
	Sun position tracking difference	x_{diff}
Load	DC branch current	i_{dc}
	DC voltage	u_{dc}
	AC branch current	i_{ac}
	AC voltage	u_{ac}

A. Data Preparation

The model of the ISS system is created in Matlab to simulate different system conditions including 1 normal and 21 faulty ones. The power of ISS strongly relies on photovoltaic power modules. Therefore, the illumination intensity has significant effect on the system. The cycle period of illumination is normalized from 90 minutes to 20s in this model. Data sampling starts after the system reaches the steady-state (20s in this case). The start sampling point are evenly distributed points between 20s to 40s for training, while random points are picked between the same time range for testing. 500 time steps at a sampling frequency of 100kHz are collected for one signal, and faults are injected between 100th to 200th time-step. For each instance, there are 25 voltage and current signals as displayed in Table III. Finally, 100 instances per condition for training and 60 for testing are prepared.

B. Selected Shapelet

At the off-line training stage, best top-k shapelets for each condition of each signal are picked out by the proposed MTES method. Four examples of shapelets are presented in Figure 10, where the shape in red is a shapelet extracted from the first class (or condition) of the signal as shown on top of each sub-figure. Other two are the same signal collected under diverse conditions. It is clear that the shapelet can match the time series belonging to the same class better, which validates that the selected shapelets have the ability of representing the shape feature of the specific class.

This article has been accepted for publication in IEEE Transactions on Instrumentation and Measurement. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TIM.2023.3336447

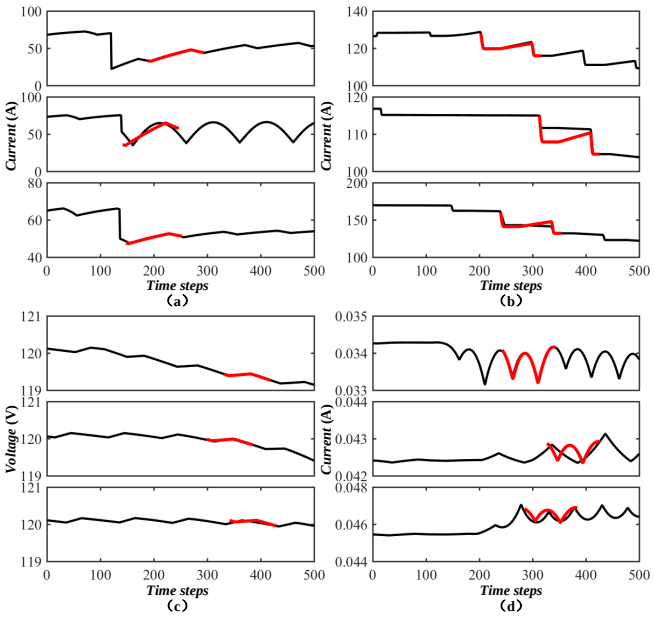

Fig. 10. Examples of the shapelet selection: (a) output current of the 1_{st} SAW (i_{w1}); (b) branch current of ESS (i_{bat}); (c) voltage of 120V bus (u_{120}); (d) branch current of pointing system (i_{point}).

C. FDL Results for the ISS Power System

In the FDL experiment, testing datasets are collected from the emulated model to validate the effectiveness of the proposed method. Finally, the accuracy reaches over 96.29%. Specifically, there are 1271 in 1320 instances that are correctly classified. Major discrepancy happen between Fault 4 and 6 (13 errors) as well as Fault 8, 9 and 10 (7 errors). It is reasonable to expect some misclassification, because under lower illumination intensity, the differentiation between variation trends of time series in different conditions becomes significantly lower. In the EPDS model, the entire period of illumination is emulated and the dataset contains samples collected under diverse illumination including a proportion of low intensity. For instance, (a) and (b) of Figures 11, 12, 13 and 14 are time series of two different conditions under lower illumination. It is obvious that discrimination between (c) and (d), which are samples collected under higher illumination, is much easier than (a) and (b). Therefore, if higher confidence level is given to the evaluation with high illumination, most misclassification can be corrected.

D. Comparison with Other Data-Driven Methods

State-of-the-art data-driven methods, such as recurrent neural network (RNN), gated recurrent unit (GRU) and convolutional neural network (CNN), are prevailing for the solution of time-series classification. Comparative experiments are conducted to demonstrate advantages of the proposed MTES method. Datasets of training and testing for different methods are the same. The architectures of networks are decided by trial and error experiments in consideration of less time consumption with reasonable accuracy and resource utilization. The input of networks is 25 features from a batch

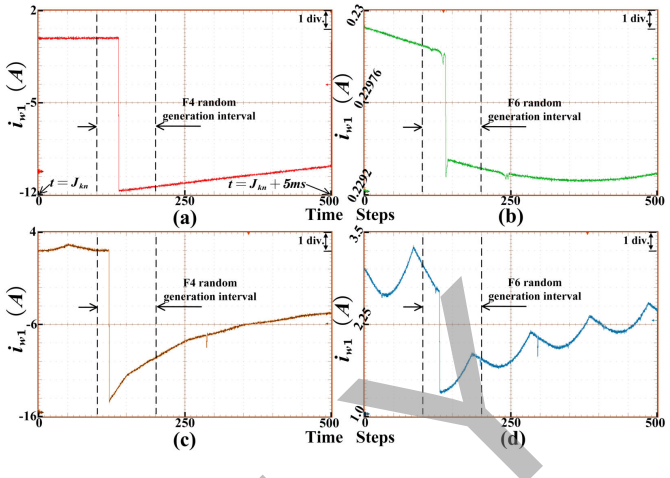

Fig. 11. Example of emulated signal of i_{w1}: (a) Fault 4 under low illumination; (b) Fault 6 under low illumination; (c) Fault 4 under high illumination; (d) Fault 6 under high illumination.

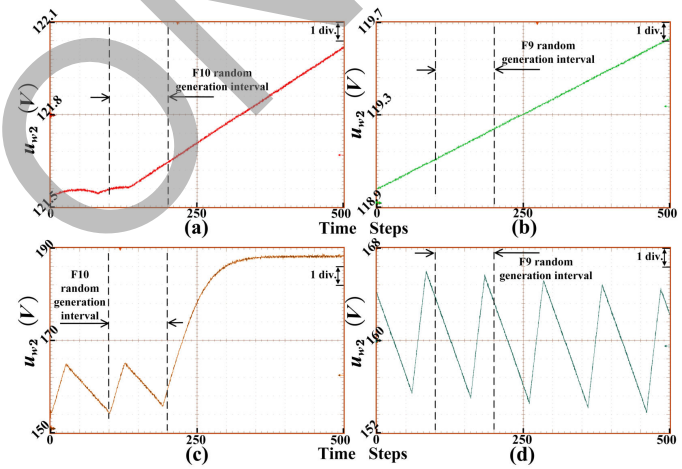

Fig. 12. Example of emulated signal of u_{w2}: (a) Fault 10 under low illumination; (b) Fault 9 under low illumination; (c) Fault 10 under high illumination; (d) Fault 9 under high illumination.

size of samples. RNN and GRU have one RNN/GRU layer with 32 hidden units. While a kernel size of 2 is applied to each of the 8 filters for CNN1D layer which is connected to a maxpool layer sliding a window of height 5 across the output of CNN1D layer. Then one fully connected layer with 16 hidden units filtered by ReLU activation function is linked. The output layer is a fully connected layer cascaded to a softmax classifier having 22 classes. The batch size for training is 64, and the optimizer is Adam with the learning rate of 0.001. The hardware resource consumption, latency and accuracy of the three methods are shown in Table IV. Compared with Table II, the algorithm of MTES method is simpler to implement on FPGA since it consumes much less execution time and hardware resources than others. More importantly, the accuracy for FDL using MTES is obviously higher.

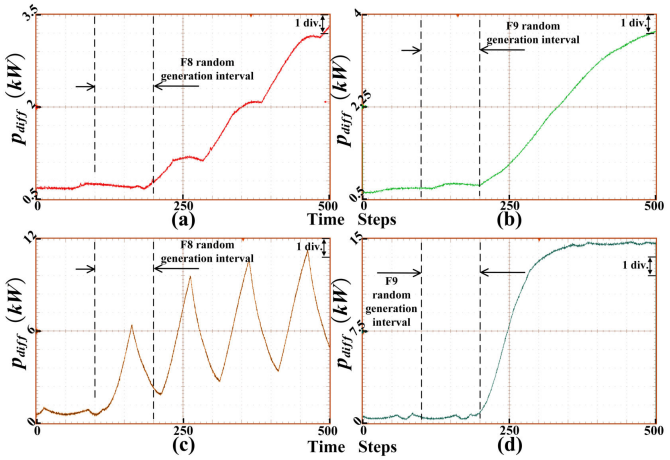

Fig. 13. Example of emulated signal of p_{diff}: (a) Fault 8 under low illumination; (b) Fault 9 under low illumination; (c) Fault 8 under high illumination; (d) Fault 9 under high illumination.

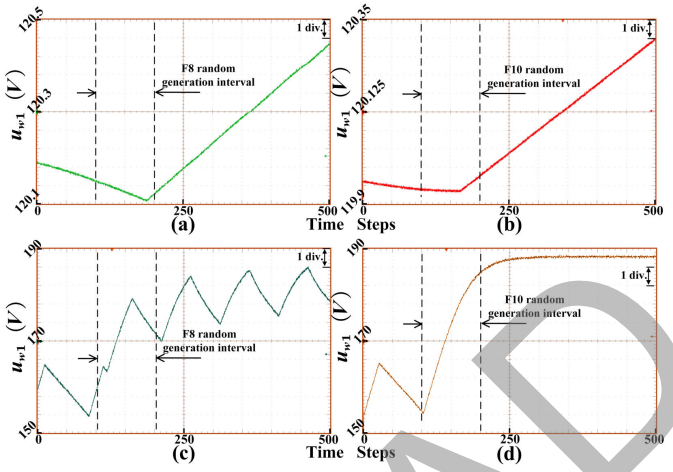

Fig. 14. Example of emulated signal of u_{w1}: (a) Fault 8 under low illumination; (b) Fault 10 under low illumination; (c) Fault 8 under high illumination; (d) Fault 10 under high illumination.

VII. Conclusion

A shapelet-based multi-dimensional time-series classification method for real-time FDL on the ISS EPDS is proposed in this work. A multi-trend extracted shapelet discovery approach was created to fast extract shape features from raw time-series signals. Then best top-k shapelets were further selected from the shapelet candidate set using randomization and information gain. The selected shapelets manifested that they had the ability of representing the shape feature of the specific class. Finally, faults were detected and localized by a designed random forest classifier, which learned from a Bayesian probability table combining information of all multi-dimensional time series. Besides, shapelets for testing in real time were further shrunk by the correlation-based feature subset selection. The experiment showed that over 96% accuracy was achieved. With the help of the FPGA hardware implementation, system conditions can be estimated in less than 10µs to realize the real-time FDL, which was verified through the HIL real-time emulation testing. Comparing with other state-of-the-art data-driven methods, MTES method reached better classification accuracy with much less time and resource consumption on hardware.

TABLE IV
Specifics of hardware resource consumption, latency, and accuracy of other methods.

	BRAM	DSP	FF	LUT	Latency (ms)	Accuracy
RNN	457	3,038	380,026	601,996	0.248	67.95%
GRU	457	6,112	544,075	1,103,740	0.566	89.84%
CNN	112	1,438	124,043	234,833	1.087	88.11%

The application of MTES method could be extended to any TSC tasks having clear clue of signal variation trend. Furthermore, the classification accuracy should be guaranteed as long as the trend in a class is definite and consistent, even without tremendous data support. On the contrary, misclassification may be made if the trend is variable in one class or not obviously different between classes. This also was the reason that misclassifications were made occasionally in this study. The signal trends between two faults became similar on the condition of low illumination intensity, which could be avoided by increasing the confidence level of the evaluation with higher illumination for the FDL task of the ISS.

References

[1] R. C. Dempsey, D. E. Contella, D. H. Korth, M. L. Lammers, C. R. McMillan, E. Nel-son, R. J. Renfrew, B. T. Smith, S. A. Stover, and E. A. V. Cise, "The international space station: operating an outpost in the new frontier." https://www.nasa.gov/feature/new-nasa-e-book-offers-inside-look-at-space-station-flight-controllers, 2018. [Online].

[2] NASA, "Reference guide to the international space station." https://www.nasa.gov/pdf/508318main_ISS_ref_guide_nov2010.pdf, 2010. [Online].

[3] J. Zhenhua, L. Shengyi, and R. A. Dougal, "Design and testing of spacecraft power systems using VTB," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 3, pp. 976–989, 2003.

[4] Y. Sun, L. Guo, Y. Wang, Z. Ma, and Y. Niu, "Fault diagnosis for space utilisation," *J. Eng.*, vol. 2019, no. 23, pp. 8770–8775.

[5] K. Kim, M. Capell, A. Lebedev, G. Viertel, and J. Yang, "A power distribution system for the ams experiment on the International Space Station," *J. Instrum.*, vol. 1, no. 12, p. T12001, 2006.

[6] M. Boucher, "Critical space station spacewalk a success." http://spaceref.com/international-space-station/critical-space-station-spacewalk-a-success.html, 2012. [Online].

[7] S. J. Lu, P. Siqueira, V. Vijayendra, H. Chandrikakutty, and R. Tessier, "Real-time differential signal phase estimation for space-based systems using FPGAs," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 1192–1209, 2013.

[8] P. P. Lin and K. Jules, "An intelligent system for monitoring the microgravity environment quality on-board the international space station," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 5, pp. 1002–1009, 2002.

[9] M. Khanafer and S. Shirmohammadi, "Applied AI in instrumentation and measurement: the deep learning revolution," *IEEE Instrum. Meas. Mag.*, vol. 23, no. 3, pp. 10–17, 2020.

[10] I. M. Karmacharya and R. Gokaraju, "Fault location in ungrounded photovoltaic system using wavelets and ANN," *IEEE Trans. on Power Del.*, vol. 33, no. 2, pp. 549–559, 2018.

[11] F. Aziz, A. Ul Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A novel convolutional neural network-based approach for fault classification in photovoltaic arrays," *IEEE Access*, vol. 8, pp. 41889–41904, 2020.

[12] C. Chen, N. Lu, B. Jiang, Y. Xing, and Z. Zhu, "Prediction interval estimation of aeroengine remaining useful life based on bidirectional long short-term memory network," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[13] B. Ma, Z. Liu, Q. Dang, W. Zhao, J. Wang, Y. Cheng, and Z. Yuan, "Deep reinforcement learning of UAV tracking control under wind disturbances environments," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[14] Q. Liu, T. Liang, Z. Huang, and V. Dinavahi, "Real-time FPGA-based hardware neural network for fault detection and isolation in more electric aircraft," *IEEE Access*, vol. 7, pp. 159831 – 159841, 2019.

[15] J. J. Q. Yu, Y. Hou, A. Y. S. Lam, and V. O. K. Li, "Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks," *IEEE Trans. on Smart Grid*, vol. 10, no. 2, pp. 1694–1703, 2019.

[16] D. K. Asl, A. Hamedi, M. Shadaei, H. Samet, and T. Ghanbari, "A non-iterative method based on fast fourier transform and least square for fault locating in DC microgrids," in *EEEIC / I CPS Europe*, pp. 1–5, 2020.

[17] K. Thattai, A. Sahoo, and J. Ravishankar, "On-line and off-line fault detection techniques for inverter based islanded microgrid," in *CPE-POWERENG*, pp. 1–6, 2018.

[18] L. Ye and E. Keogh, "Time series shapelets: A new primitive for data mining," in *Proc. 15th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, p. 947–956, 2009.

[19] Z. Li, R. Outbib, S. Giurgea, and D. Hissel, "Fault diagnosis for PEMFC systems in consideration of dynamic behaviors and spatial inhomogeneity," *IEEE Trans. Energy Convers.*, vol. 34, no. 1, pp. 3–11, 2019.

[20] X. Yang, Y. Zheng, Y. Zhang, D. S.-H. Wong, and W. Yang, "Bearing remaining useful life prediction based on regression shapalet and graph neural network," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[21] T. Rakthanmanon and E. Keogh, "Fast shapelets: A scalable algorithm for discovering time series shapelets," in *Proc. 2013 SIAM Int. Conf. Data Mining*, pp. 668–676, 2013.

[22] H. Zhao, Z. Pan, and W. Tao, "Regularized shapelet learning for scalable time series classification," *Computer Netw.*, vol. 173, pp. 107–171, 2019.

[23] J. Grabocka, N. Schilling, M. Wistuba, and L. Schmidt-Thieme, "Learning time-series shapelets," in *Proc. 20th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, p. 392–401, 2014.

[24] X. Renard, M. Rifqi, W. Erray, and M. Detyniecki, "Random-shapelet: An algorithm for fast shapelet discovery," in *IEEE Int. Conf. DSAA*, pp. 1–10, 2015.

[25] Q. Meng and P. Pu, "RLS: An efficient time series clustering method based on u-shapelets," *Intell. Data Anal.*, vol. 22, no. 4, pp. 767–785, 2018.

[26] M. A. Hall, *Correlation-Based Feature Selection for Machine Learning*. PhD thesis, The University of Waikato, Hamilton, New Zealand, 1998.

[27] V. Dinavahi and N. Lin, *Real-Time Electromagnetic Transient Simulation of AC-DC Networks*. Piscataway, NJ, USA: Wiley, 2021.

[28] A. M. McNelis, R. F. Beach, J. F. Soeder, N. B. McNelis, R. May, T. P. Dever, and L. Trase, "Simulation and control lab development for power and energy management for NASA manned deep space missions," in *12th Int. Energy Convers. Eng. Conf.*, pp. 1–8, 2014.

[29] W. Chen, S. Zhang, and V. Dinavahi, "Real-time ML-assisted hardware-in-the-loop electro-thermal emulation of LVDC microgrid on the international space station," *IEEE Open J. Power Electron.*, vol. 3, pp. 168–181, 2022.

[30] I. Al-Zyoud and K. Khorasani, "Neural network-based actuator fault diagnosis for attitude control subsystem of a satellite," in *World Autom. Congr.*, pp. 1–6, 2006.

[31] M. S. Arani and M. A. Hejazi, "The comprehensive study of electrical faults in PV arrays," *Journal of Electrical and Computer Engineering*, vol. 2016, pp. 1–10, 2016.

[32] E. Aghabarari and C. J. Adams, "Off-nominal conditions and caution and warning techniques for the electric power system for the international space station," in *IECEC-97 Proc. 32nd Intersociety Energy Conv. Eng. Conf.*, vol. 4, pp. 2246–2251, 1997.

[33] A. Mehmood, H. A. Sher, A. F. Murtaza, and K. Al-Haddad, "A diode-based fault detection, classification, and localization method for photovoltaic array," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[34] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A comprehensive review of catastrophic faults in PV arrays: Types, detection, and mitigation techniques," *IEEE J. Photovolt.*, vol. 5, no. 3, p. 982–997, 2015.

[35] T. Kwa-Sur, Y. Lifeng, and N. Dravid, "Modeling the protection system components of the space station electric power system," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 800–808, 1994.

[36] G. Di Leo and F. Sardanelli, "Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach," *European Radiology Experimental*, vol. 4, no. 18, pp. 1–8, 2020.

[37] Xilinx, "VCU118 evaluation board user guide." https://www.xilinx.com/support/documentation/boards_and_kits/vcu118/ug1224-vcu118-eval-bd.pdf, 2018. [Online].

[38] Xilinx, "VCU128 evaluation board user guide." https://www.xilinx.com/support/documentation/boards_and_kits/vcu128/ug1302-vcu128-eval-bd.pdf, 2018. [Online].

[39] *Aurora 64B/66B LogiCORE IP Product Guide, PG074(V11.2).* San Jose, CA, USA: Xilinx Inc., 2017.

Qin Liu (Student Member, IEEE) received the B.Eng. degree in measurement and control technology and instrument from Beijing Jiaotong University, Beijing, China, in 2013, and the Ph.D. degree in mechanical engineering from Beijing Jiaotong University in 2021. She was a visiting Ph.D. student in Electrical and Computer Engineering at the University of Alberta, Edmonton, Alberta, Canada, from 2018 to 2020. Her research interests include real-time simulation and control of power systems, hardware acceleration, and machine learning-based fault diagnosis.

Weiran Chen (Student Member, IEEE) received the B.Eng. degree in electrical engineering from Harbin Engineering University, Harbin, Heilongjiang, China, in 2018. He is currently pursuing the Ph.D. degree in electrical and computer engineering at the University of Alberta, Edmonton, AB, Canada. His research interests include real-time simulation of power systems, power electronic systems, and field programmable gate arrays.

Venkata Dinavahi (Fellow, IEEE) received the B.Eng. degree in electrical engineering from Visvesvaraya National Institute of Technology (VNIT), Nagpur, India, in 1993, the M.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT) Kanpur, India, in 1996, and the Ph.D. degree in electrical and computer engineering from the University of Toronto, Ontario, Canada, in 2000. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada. He is a Fellow of the Engineering Institute of Canada (EIC) and a Fellow of the Asia-Pacific Artificial Intelligence Association (AAIA). His research interests include real-time simulation of power systems and power electronic systems, electromagnetic transients, device-level modeling, artificial intelligence machine learning, large-scale systems, and parallel and distributed computing.