# Fault and Attack Detection and Diagnosis by Analysis of Electrical Waveforms of Power Networks

**Stephen J. Coshatt, Bowen Yang, Jin Ye, WenZhan Song**
**Center for Cyber-physical Systems**
**University of Georgia**
**Athens, GA 30602**
**{stephen.coshatt, bowen.yang, jin.ye, wsong}@uga.edu**

**Feraidoon Zahiri**
**402 CMXG/MXDEO**
**Robins Air Force Base**
**Warner Robins, GA 31098**
**feraidoon.zahiri@us.af.mil**

**James Hill**
**Aging Aircraft Solutions**
**Warner Robins, GA 31098**
**james.hill@agingaircraftconsulting.com**

***Abstract*—In recent years, increasing numbers of electronic control units (ECUs), programmable logic controllers (PLCs), and other types of programmable electronics have been deployed into cyber-physical systems. While such progress increased productivity and product quality, it also introduces vulnerabilities to both hardware and software. This study proposes a data-driven approach to monitoring the electric waveforms of the power network of cyber-physical systems for attack and fault detection and diagnosis. In terms of methodology, most studies focus on classification, which only allows for classification of known attacks or faults. While new attacks could be detected, they cannot be properly diagnosed as a new attack would be forced into one of the existing classifiers, thus leading to an incorrect diagnosis. This study proposes using clustering to detect and diagnosis anomalies. Specifically, it proposes using two dimensional unsupervised shapelets (2D u-shapelets) for clustering. U-shapelets are short time series with discriminatory capabilities that can be automatically extracted from a data set. This study is the first of a two phase study to incorporate dynamic clustering with u-shapelets. The advantage of this long term approach allows a system to notify systems users of a new type of attack or fault, which can later be labelled. Thus, the system can learn to identify new anomalies. Extensive evaluations are conducted to study the algorithm performance,such as the performance metrics vs the number of clusters and anomaly types, and the effectiveness for novel adversarial attacks on such systems.**

## TABLE OF CONTENTS



## 1. INTRODUCTION

As the presence of cyber-physical systems grows, so does the likelihood of cyber-attacks on these systems. The Colonial Pipeline shutdown is just one recent example. Modern cyber-physical systems, often referred to as operational technology (OT) in industry, have external communication capabilities that are increasingly networked to take advantage of advances in Industry 4.0 advanced capabilities including digital engineering, industrial IoT, data analytics, digitization, and integration of the cyber-physical value chain. These communications capabilities increase potential fault and cyber-attack vectors, even in air gapped networks. Industry has growing security concerns that a STUXNET style attack on its cyber-physical systems could degrade or damage their capability to provide services and support. In addition to critical infrastructure, manufacturing systems for the aerospace industry are also an area of concern.

To our knowledge, limited studies have been done on using the information embedded in electrical signals for cyber-threat detection in cyber-physical systems such as manufacturing systems. Some cyber-threats including integrity attacks may not be observed in the cyber-space alone and can only be discovered through inter-dependency analysis of multiple cyber and physical signals. Thus, there is a significant opportunity in exploring physical signals, together with cyber signals, to advance cyberspace security and trustworthy research and design. While there is a plethora of potential cyber-attacks, this study utilizes attacks that directly affect the operation of the electrical machines and its components. In addition to detecting anomalies in a system, diagnosing them is also important. An ideal system should be able to distinguish attacks from faults and should be able to identify the type of each. The attacks considered in this work are data injection attack (DIA), coordinated DIAs, and replay attacks. Additionally, open circuit and short circuit faults are utilized. This study proposes a data-driven approach to monitoring the electric waveforms of the power network of cyber-physical systems for attack and fault detection and diagnosis. In terms of methodology, most studies focus on classification, which only allows for classification of known attacks or faults. While new attacks could be detected, they cannot be properly diagnosed as a new attack would be forced into one of the existing classifiers, thus leading to an incorrect diagnosis. This study proposes using (unsupervised) clustering to detect and diagnosis anomalies. It is the first step in a two phase study that proposes using a dynamic clustering algorithm that can automatically create a new cluster in the diagnosis phase in the case of an attack that does not fit well into a current cluster. The advantage of this approach allows a system to notify systems users of a new type of attack or fault, which can later be labelled. Thus, the system can learn to identify new anomalies. Extensive evaluations are conducted to study the algorithm performance,such as the performance metrics vs the number of clusters and anomaly types, and the

978-1-6654-3760-8/22/$31.00 ©2022 IEEE

effectiveness for novel adversarial attacks on such systems.

## 2. BACKGROUND AND RELATED WORKS

Power electronics converters are becoming more vulnerable to cyber/physical attacks due to their growing penetration in Internet of Things (IoT) enabled applications including the smart manufacturing and smart grids [1]. Due to the lack of cyber awareness in power electronics community [1], it becomes more urgent to develop cyber/physical attack detection and identification strategies for power electronics converters in many safety-critical applications since these malicious attacks can lead to a catastrophic failure and substantial economic loss if not detected in the early stage.

Attacks are studied in applications which are intensively dependent on power electronics converters, including power grids with voltage support devices [2], distribution systems with solar farms [3], with power electronics driven HVAC (Heating, ventilation, and air conditioning) systems [4], and microgrids [5, 6]. However, they mostly focus on either analyzing or detecting cyber attacks affecting grid level stability, functionality and operational costs. In [7], a model-based method was developed to detect data integrity attacks on automation generation control of transmission systems. In [3], a physical-law based detection was developed to detect false data attacks which attempt to reduce the output power of solar energy in distribution systems. In [4], a secure information flow framework was developed for 118-bus distribution network with power electronics driven HVAC system. In [8], a physics-based, cooperative mechanism was developed to detect stealthy attacks in DC microgrids with a number of DC-DC converters, which can bypass most of observer-based detection methods. In [9], a physics-based framework to detect false-data injection attacks in DC microgrids with a number of DC-DC converters. While power electronics converters are included in their cyber security monitoring frameworks, they are designed to detect one particular type of grid-level cyber attacks but those on the devices (power electronics converters) are not studied. Thus, their cyber security framework is not applied to (1) cyber attack detection on power electronics converters, which might affect the performance of power electronics converters; and (2) the root cause identification when a variety of attacks occur.

Data-driven approaches are gaining increased attention in recent years due to the advancements in sensing and computing technologies [10–12]. They show great potentials in detecting and identifying complicated cyber and physical attacks. Shapelets are a concept that was introduced by Koegh and Ye in [13]. "Time series shapelets are small, local patterns in a time series that are highly predictive of a class" [14]. Shapelets have primarily been used in time series classification [14]. The concept of u-shapelets were introduce by Zakira et. al. in [14]. This study utilizes the concept of "good enough" shapelets that was introduced by Ulanova et. al. in [15]. The u-shapelet extraction algorithm used in this study is a modified version of the Random Local Search (RLS) algorithm developed by Meng and Pu in [16]. While u-shapelets have been demonstrated to be effective with time series, to our knowledge, they have only been used with univariate time series. The short term goal of this study is to demonstrate the utility of shapelets [13] in the detection and diagnosis of cyber-attacks and faults in the analysis of electrical waveforms. In addition, we will attempt to do so with multivariate time series using a modification to u-shapelets call 2D u-shapelets. These 2d u-shapelets are incorporated into a modified version of the RLS algorithm. Long term, the goal is to combine u-shapelets with dynamic clustering and to do so in a streaming/online fashion. By dynamic clustering, we mean clustering that does not have a fixed number of clusters. At present, there is a dynamic clustering algorithm called Dynamic clustering for tracking evolving environments (DyClee) [17], with some modifications, that fits our long term goals and approach.

U-shapelets have demonstrated features that make them useful for clustering. First, they can ignore irrelevant data [15] [16]. Second, u-shapelets are designed to work well with time series in which the objects of interest in the data are of different lengths [15] [16]. Third, u-shapelets can separate data that belongs two one class from all other data that does not belong in that class [14] [15]. A "good enough" shapelet concept developed by [14]. Lastly, u-shapelets can be selected from time series data without human intervention [14].

**Definition 1:** A time series $T$ is a finite sequence of real-valued numbers $t_i$ on $i = 1, 2, \ldots, n$. This number $n$ is the length of $T$. [16]

**Definition 2:** Subsequences $S_{i,l}$ denote segments $t_i, t_{i+1}, \ldots, t_{i+l-1}$ of $T$ starting at position $i$ with length $l$, for $1 \leq i \leq n$ and $1 \leq l \leq n - i + 1$. [16]

**Definition 3:** An unsupervised shapelet (u-shapelet) $\acute{S}$ is a subsequence that can divide a dataset $D$ into two groups, $D_A$ and $D_B$. The distance between $\acute{S}$ and any time series in group $D_A$ is much smaller than $sdist$ between $\acute{S}$ and any time series in $D_B$. [16] [needs modification for this project].

**Definition 4:** An orderline is a vector of subsequence distances $sdist(\acute{S}, T_i)$ between a u-shapelet candidate $\acute{S}$ and all time series $T_i$ in the dataset. [15]

**Definition 5:** A gap score is a separation measure between $D_A$ and $D_B$ on the orderline where [14]:

$$gap = \mu B – \sigma B – (\mu A + \sigma A) \tag{1}$$

A "good enough" shapelet concept developed by [15] is described as follows:

**Definition** 6: Let the best u-shapelet in the dataset have a gap score of $n_{best}$ and the left part of its orderline contain a set of time series $D_{A_best}$. We call a u-shapelet having a gap score $n_{good}$ and containing the same set of time series on the left part of its orderline $D_{A_good} = D_{A_best}$ as a good enough u-shapelet if it has the following property: there is no u-shapelet candidate with a gap score of $n_{any} > n_{good}$ and left part of its orderline $D_{A_any}$ such that ($D_{A_any} \neq D_{A_best}$) [14].

**Definition 7**. The distance map [14] $DIS$ is a matrix containing the distance between all u-shapelets and all time series within $D$. [16]

Note that the distance map is used as input to a clustering algorithm.

## 3. ALGORITHM AND SYSTEM DESIGN

In this work, we study both detection and diagnosis by analyzing electrical waveforms. This study uses u-shapelets and a simple anomaly detection algorithm for detection and u-shapelets with clustering for diagnosis. Once the detection

algorithm detects an anomaly, data is then sent to the clustering algorithm for diagnosis.

*Threat Model*

Power converters in the cyber-physical system are connected to the power network and the cyber network to simulate an integrated system. Given this setup, it is possible for malicious actors to compromise the integrity of the converter controllers via the communications network. This set up is also potentially vulnerable to an insider threat that has direct access to the controllers. Two simple attacks are performed on the PV converter sensors, a data injection attack (DIA) and a replay attack. For a coordinated DIA, multiple converter sensors are targeted simultaneously. There has been a multitude of studies on these types of integrity attacks on cyber-physcial system with some of the most recent being [18] [19] [9] [20] [21] [22]. For these attacks, the vector feedback signal is denoted as $Y$ and the compromised sensor measurements as $\widetilde{Y}$. For the PV converter, there is $Y = [u_{pv}, i_{pv}, u_{dc}, i_f, u_c]^T$ [23]. The attack duration is denoted as $T_a = [t_s, t_e]$, where $t_s$ and $t_e$ represent the start and end time of attack, respectively. Then, the DIA can be expressed as

$$\widetilde{Y}(t) = \begin{cases} Y(t),\ t \notin T_a \\ \nu * Y(t) + \epsilon,\ t \in T_a \end{cases} \quad (2)$$

where $v$ and $\epsilon$ are unknown attack signals. Replay attacks are expressed as $\widetilde{Y}(t) = \mathbf{Y}$ during $T_a$, where $\mathbf{Y}$ represents a prerecorded set of the past sensor signals.

**Table 1**: Definition of cyber-physical threats

| Intrusion type | Location | Name |
|---|---|---|
| Single DIA | $i_f$ (3-phase) | A1 ... A8 |
| | $i_f$ (1-phase and 2-phase) | A9 ... A16 |
| Coordinated DIA | $\{i_f, u_c, i_g\}$, $\{i_f, u_c\}$ | A17 ... A24 |
| | $\{i_{pv}, u_{pv}\}$ | A25 ... A28 |
| Replay attack | $i_f$ (3-phase) | A29, A30 |
| Short-circuit fault | High Voltage Line Phase A | FS1, FS2 |
| | HV Line Phase AB | FS3 ... FS7 |
| | HV Line Phase AC | FS8 ... FS12 |
| | HV Line Phase ABC | FS13 ... FS15 |
| Open-circuit fault | PMW IGBT | FO1 ... FO4 |

In addition to the cyber-attacks previously mentioned, open circuit and short circuit faults were included in the model.

*Problem Setup*

The aim of the study to use data-driven anomaly detection to identify the cyber-attacks and physical faults utilizing the electrical waveform data. An assertion is that if we have sequential observations at the $k_{th}$ time instant, as follows:

$$\mathbf{x(k)} = [\mathbf{U_a(k)}, \mathbf{U_b(k)}, \mathbf{U_c(k)}, \mathbf{I_a(k)}, \mathbf{I_b(k)}, \mathbf{I_c(k)}]^{\mathbf{T}} \quad (3)$$

where $U_a, U_b, U_c$ and $I_a, I_b, I_c$ are the 3-phase voltage [V] and current [A] in the PCC node. A time-series data set can be formulated as

$$\mathbf{X} = [\mathbf{x(k-N_t+1)}, \mathbf{x(k-N_t+2)}, ..., \mathbf{x(k)}],\ \mathbf{N_t} \in \mathbb{N}. \quad (4)$$

Then, the problem of data-driven anomaly detection at $k_{th}$ time instant is to detect and cluster cyber-attacks, short-circuit faults, and open-circuit faults by using $\mathbf{X}$, as shown in Fig. 2.

Our study uses clustering for diagnosis and separates anomaly detection as a separate algorithm. Data is fed to the anomaly detection algorithm first, and once it detects an anomaly, that data is sent to a clustering algorithm.

The study uses the raw sensor data and the Fast Fourier Transform to extract the frequency, magnitude, phase angle, and the total harmonic distortion (THD). These in turn are used to extract frequency domain residuals. In total, 15 features were extracted.

Based on the magnitudes of the three phase current and the three phase current, a residual is calculated for each. This residual for the current is $\bar{R}_{m1}$ and this residual for voltage is $\bar{R}_{m2}$.

The frequency residuals were calculated for each of the three phase currents and each three phase voltage. These residuals were calculated from the distance between the fundamental frequency and the observed frequency. A tolerance of 0.5 Hz for current and 0.2 Hz for voltage was set in accordance with industry standards. The frequency residuals for current are represented as $\bar{R}_{f,C1}$, $\bar{R}_{f,C2}$, $\bar{R}_{f,C3}$ and the frequency residuals for voltage are represented as $\bar{R}_{f,V1}$, $\bar{R}_{f,V2}$, and $\bar{R}_{f,V3}$.

The THD residuals were also calculated for each of the three phase currents and each three phase voltage. The maximum THD allowed was set to 5% in accordance with industry standards. The frequency residuals for current are represented as $\bar{T}_{f,C1}$, $\bar{T}_{C2}$, $\bar{T}_{C3}$ and the frequency residuals for voltage are represented as $\bar{T}_{V1}$, $\bar{T}_{V2}$, and $\bar{T}_{V3}$.

Lastly, a time series feature, Mean Current Vector (MCV), was extracted from the three phase current. It is represented as $P_{mcv}$.

*2D U-Shapelet Extraction*

In this section, we will discuss our contribution, which is the 2D u-shapelet concept. Additionally, this section will describe the primary u-shapelet algorithms that we chose to build on and improve.

In previous research, u-shapelets have been utilized only for univariate time series data and are represented as a 1D vector. Our research is concerned with multivariate time series data and unsupervised clustering. There has been some research with shapelets and multivariate time series with classification [24] [25] [26]. One method considered for u-shapelet extraction of multivariate data is to simply utilize one of the existing algorithms for each feature in time series. So for a dataset with 10 features, said algorithm would have to be run 10 times, once for each feature. For large datasets with lots of features, u-shapelet extraction would be very time consuming. The idea behind 2D u-shapelets is to consider them as a stack of 1D vectors. Subsequences of a time series would be treated in the same manner. In this approach, matrix comparisons are used instead of distance measures to determine "likeness" or "distance" between a 2D u-shapelet and 2D subsequence of a time series. In this study, the Frobenius norm was chosen due to its ease of computation and frequency of use. The equation below is used in place of our chosen algorithms

distance measure, the Random Local Search Algorithm (RLS) [16]. For our variation or RLS, $s$ represents a shapelet and $T_s$ represents a subequence of $T$ with a length equal to $s$. The Frobenius norm of matrix $A$ is represented $\| A \|_F$

$$\mathbf{d} = \frac{\| \mathbf{T_s} \|_\mathbf{F} - \| \mathbf{s} \|_\mathbf{F}}{\| \mathbf{s} \|_\mathbf{F}} \tag{5}$$

This approach should be effective where the features are correlated, since a change in state should affect all features. So in the case of 3-phase voltage and 3-phase current (and features derived from them), all features should be correlated. Figure 1 demonstrates the concept.

| | Time 0 | Time 1 | Time 2 | Time 3 | Time 4 | Time 5 |
|---|---|---|---|---|---|---|
| feature 1 | 1 | 4 | 3 | 3 | 4 | 0 |
| feature 2 | 4 | 3 | 1 | 3 | 0 | 3 |
| feature 3 | 1 | 2 | 4 | 2 | 1 | 2 |
| feature 4 | 2 | 0 | 3 | 1 | 4 | 4 |

**Figure 1**: Example of a simple 2D u-shapelet

In considering options for a u-shapelet algorithm modify for 2D shapelets, the following criteria were used:

- Simple algorithms preferred over complex ones.
- Time to extract u-shapelets. The shorter the better.
- A straight forward way incorporate the 2D u-shapelet concept.

The first study on u-shapelets was presented by Zakira et. al. [14]. The study used a brute force approach in finding the best u-shapelets. The algorithm essentially considers every subsequence of every possible size in a time series as a potential candidate and u-shapelet, although a range of subsequence lengths can be specified. Once a valid candidate is found, the u-shapelet and all subsequences that are similar are dropped from the time series and the process repeats until no more time series remain. This approach, however was very time consuming and only proved effective with small data sets [15] [16] [27] [28]. Ulanova et. al. introduced a u-shapelet extraction algorithm, called Scalable U-Shapelet (SUSh), that was up to two orders of magnitude faster than the brute force one [15] [16]. This study also introduced the concept of "good enough" shapelets. The authors assert that in any time series, regardless of size, there exists a number of shapelets that vary in their u-shapelet score only slightly but are "good enough" to provide good clustering. Thus SUSh attempts to find a number of these good enough shapelets and then stop searching for more. This algorithm converts the subsequences into Symbolic Aggregate Approximation (SAX) representation. A random mask is applied to candidate shapelets prior creating the orderline and gap scores to determine if it is a good enough shapelet. The random masking is done so that similar subsequences will collide with the candidate shapelet. Too many or not enough collisions mark a candidate as a bad candidate. While this algorithm is much faster than its predecessor, it does have some disadvantages [16]. First, it only searches for u-shapelets of a fixed length [16] [15]. Second, the SAX and random masking process is quite complicated [16].

Unsupervised Salient Subsequence Learning (USSL), an algorithm created by Zhang et. al. [28], introduced the concept of learned unsupervised shapelets (lu-shapelets). This algorithm is an improvement of the Unsupervised Shapelet Learning Model (USML) by the same authors in a previous study [29]. It is based on an optimization model that integrates shapelet learning, spectral analysis, pseudo-class labels, and least-squares minimization [28]. The authors demonstrated the effectiveness of their algorithm against a wide variety of unsupervised time series clustering. USSL was compared against the brute-force clustering. While SUSh was referenced, it was not compared to USSL. The RSL algorithm, discussed in the next section, was neither referenced nor compared to USSL. However, since the brute-force method finds the best u-shapelets versus the good enough u-shapelets, the brute force method should cluster better than either RSL or SUSh. According to the USSL study, it out performed the brute force method by and average of %20 [28]. The time to discover u-shapelets is not easy to compare based on these studies.

The Random Local Search algorithm developed by Meng and Pu in [16] is an attempt to make improvements over the SUSh algorithm. Like SUSh, RLS looks for "good enough" u-shapelets. It does so by using a random search. RLS randomly searches the dataset until it finds a specified number of candidates. It then performs local search of nearby candidates (neighborhood) to see if a better one exists. If so, the initial candidate is the replaced by the best candidate in the neighborhood. Once this is complete for all initial candidates, the resulting candidates are put in ascending order based on their gap score. The top $k$ candidates are then used to create a distance map. The distance map is simply the distance of each time series subsequence in the data set from each u-shapelet. In the case of RLS, the length normalized euclidean distance is used. The distance map is then used as an input into a clustering algorithm.

The primary candidates were the SUSh, RLS, and the USSL. Incorporating 2D shapelets was a rather simple modification for both SUSh and RLS, however this was not so for USSL. In fact, we attempted a version with SUSh in addition to RLS. The SAX and random masking process caused issues with collisions as a means to test the discriminatory ability of candidate u-shapelets. In a 2D matrix, collisions were very rare. So rare that the algorithm rarely found a good candidate. Increasing the number of masks did not show much improvement until the number of mask reached roughly half the number of values in the matrix. At his point, there were concerns that any candidate shapelets would not have any discriminatory ability. Thus the SUSh approach with 2D u-shapelets was dropped and we continued with the RLS only.

Our algorithm is a modification of the RLS, where 1D shapelets on univariate data are replaced with 2D shapelets of multivariate data. Additionally, the length normalized Euclidean distance metric used in the original RLS is replaces by our Frobenius norm comparison metric discussed above.

*Anomaly Detection and Clustering with 2D U-Shapelet*

To demonstrate the efficacy of 2D u-shapelets, it was decided to use a simple algorithm. A basic K Nearest Neighbors (KNN) was used in this study.

For clustering, K-Means was chosen because it was commonly used in other u-shapelet studies for clustering and because it is a rather simple and well known.

The main purpose of this work is to demonstrates that our proposed 2D u-shapelets are effective approach to detect anomalies and separate them into different clusters, even with the simple KNN and K-means machine learning methods. Advanced methods can further improve the accuracy and performance.

## 4. Experiments and Evaluations

*Experiment setup*

The model and data used in this study is based on a testbed model co-developed by the Intelligent Power Electronics Electric Machine Lab and the Sensorweb Research Lab at the University of Georgia (UGA) for generating electric waveform data. Data from this testbed is reffered to in this study as the UGA dataset. The model consists of seven solar panels, seven first stage DC/DC converters, seven second stage DC/AC converters and a transformer. The model is depicted in Fig. 1. An OPAL-RT testbed connected to an IEEE 37-bus power grid is used to create a real-time model. OPAL-RT and an embedded field-programmable gate array (eFPGA) were used to create the seven sets of PV converters. Each PV converter grid voltage is set to 1500V and the grid side LCL rated voltage is set to 480V. Notice that, the power electronics networks of manufacturing system have similar characteristics.

*Data Set for Training, and Testing*

The dataset contains 49 files, each with 30 seconds of normal data and 10 seconds of anomaly data. The data is the raw sensor data composed of 3-phase current and 3-phase voltage. Each set has 800,000 samples total (20,000 samples per second). Only one anomaly per file. This data was appended together in two sets, one for training and one for testing. See Table 2 for the breakdown of each dataset. Thus a very large dataset of 1,881,600,000 samples was created. This sets up a situation where the random search has a 75% chance to jump into a section of normal data. To reduce the time of find u-shapelets and to increase the chance of the random search finding shapelets within anomaly data, a condensed dataset used only for shapelet extraction was created. We selected 1.75 seconds of data from at least one file of each anomaly which contained 0.25 seconds of normal data and 1.5 seconds of the anomaly.

From these sets, the raw data is down-sampled to 2,000 samples per second. Then 15 features are extracted and this extracted dataset is referred to as the 15-Features set. These features and 2D u-shapelets extracted from these features were then used as inputs into the selected models for comparison.

**Table 2**: Number of Each Type of Anomaly in Dataset

| Anomaly Type | # Training | # Testing |
|---|---|---|
| Single DIA | 12 | 4 |
| Coordinated DIA | 9 | 3 |
| Replay attack | 1 | 1 |
| Short-circuit fault | 11 | 4 |
| Open-circuit fault | 3 | 1 |
| **Set Totals** | **36** | **13** |

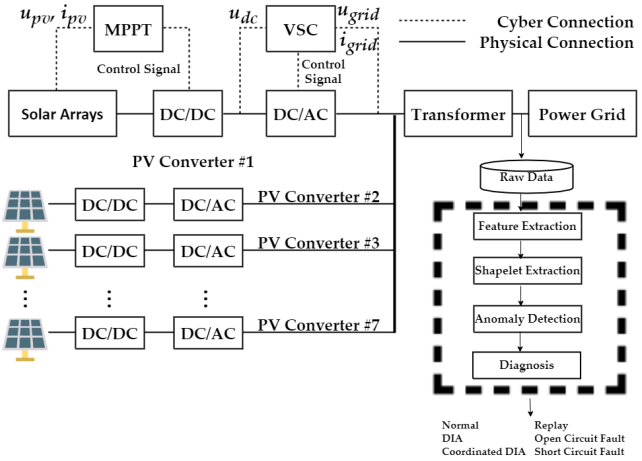


**Figure 2**: Schematic diagram of the PEC-enabled PV farm.

For evaluation, the u-shapelets performance in diagnosis and detection were compared to the 15-Features used in the same algorithms.

The algorithms involved in this study utilize some hyper-parameters and will be discussed briefly here. The RLS algorithm has the following: the number of u-shapelets to evaluate ($r$), the number of u-shapelets to use in clustering ($k$), radius of the neighbourhood search ($R$), the minimum length of a u-shapelet ($minlen$), and the maximum length of a u-shapelet ($maxlen$). After multiple runs of various values of $k$ from 2 to 6, $k = 2$ generally provided better results in clustering. For the neighborhood search, $R$ was set to 20, an approximation of the RLS's neighborhood calculation. The minimum and maximum lenghths were set to 5 and 70 respectively. Multiple runs showed that shapelets smaller in length produce better results in clustering. The best shapelets were between 5 and 16 in length. For the KNN, the contamination was set to $0.243$, which was calculated from the truth labels. For K-means, the number of clusters, which is known in this case, was set to 6. Since K-means relies on randomization of the initial centroids, the $n\_init$ parameter was set to 50. The maximum number of iterations for a single run ($max\_init$) was set to 500.

*Results*

Receiver operating characteristic (ROC) and Precision were chosen for comparing the KNN detection algorithms with u-shapelets and 15-Features. The results are in Table 3. Both performed very well and scored closely on each metric.

**Table 3**: KNN Diagnosis Comparisons

| Metric (average) | # 15-Features | # RLS |
|---|---|---|
| ROC Training | 0.980 | 0.975 |
| ROC Testing | 0.975 | 0.970 |
| Precision Training | 0.985 | 0.974 |
| Precision Testing | 0.978 | 0.965 |

Adjusted Rand Index (ARI) and Normalised Mutual Information (NMI) were chosen for comparing the K-means clustering algorithms with u-shapelets and 15-Features. The results are in Table 4. Neither performed well, but the use of 15-features fared better than the u-shapelets.

**Table 4**: K-means Diagnosis Comparisons

| Metric (average) | # 15-Features | # RLS |
|---|---|---|
| ARI Training | 0.565 | 0.555 |
| ARI Testing | 0.975 | 0.970 |
| NMI Training | 0.413 | 0.315 |
| NMI Testing | 0.410 | 0.314 |

*Analysis*

Diagnosis performed well with RLS, 2D u-shapelets, and KNN. This lends some hope for their viability with monitoring of electric waveforms for diagnosis and detection. As to the poor performance of clustering, we believe this is due to monitoring the combined signals of seven converters at the PCC node. At the PCC, all of the anomalies are much different than the normal, thus they can be easily detected. However, the differences between each type of anomaly are not as drastic, and the differences are likely muted in the combined signal. The fact that the features used in this study also did not perform well with clustering lends credibility to this assessment.

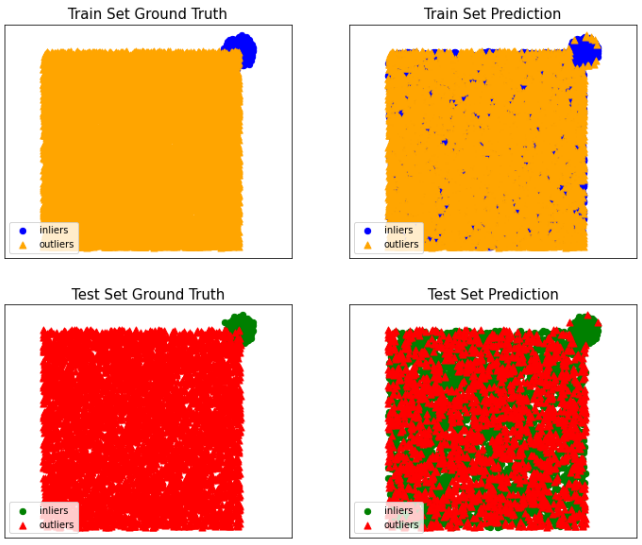


**Figure 3**: KNN plot with 15-Features

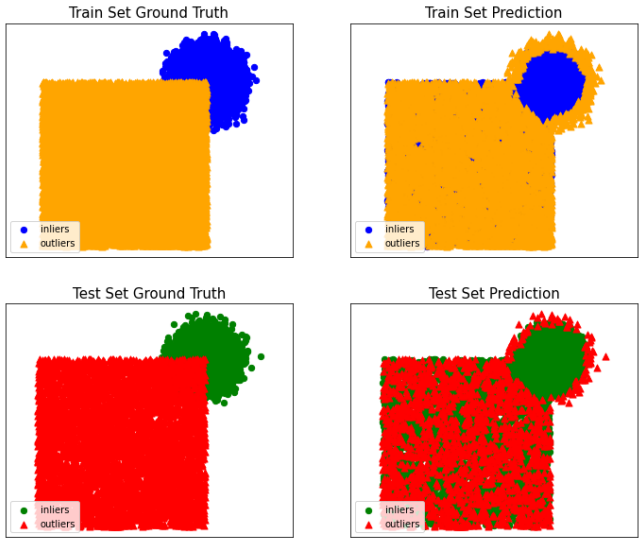


**Figure 4**: KNN plot with 2D u-shapelets

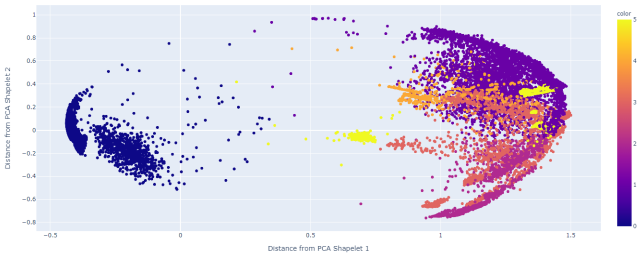


**Figure 5**: K-means Training plot with 15-Features

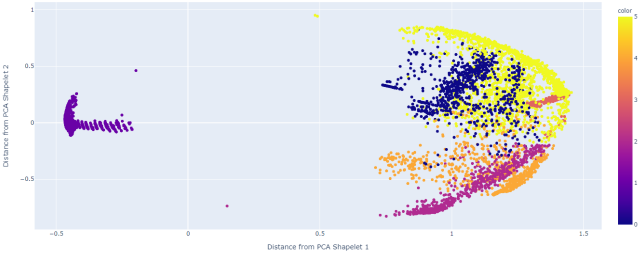


**Figure 6**: K-means Testing plot with 15-Features

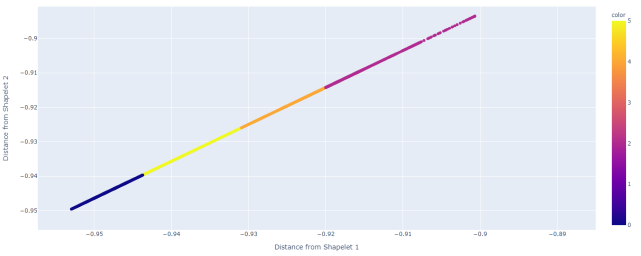


**Figure 7**: K-means Training plot with 2D u-shapelet

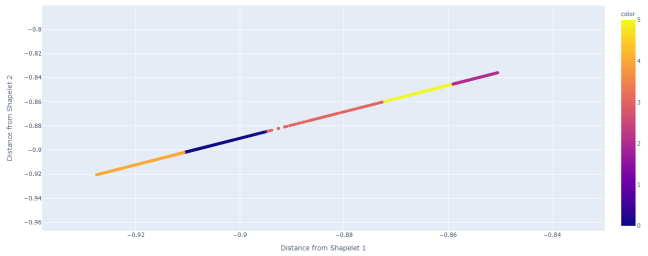


**Figure 8**: K-means Testing plot with 2D u-shapelets

*Weakness and Issues*

The 2D-shapelets did not perform as well as hoped. A better starting point for this study probably would have been analyzing attacks and faults at a single PV converter instead of a combined seven. Another issues that requires further investigation is to research more alternatives for comparing a 2D u-shaplet to a subsequence (i.e. comparing matrices). Initial choices were based on ease of computation, but this may not be the best means. Also, further exploration of adapting the USSL algorithm to learn 2D u-shapelets might improve both the quality of the shapelets and decrease the time to extract them.

## 5. Conclusions

This paper has presented a new concept of 2D u-shapelets and attempted to demonstrate their utility with multivariate time series data and the detection and diagnosis of cyber-attacks and faults in electrical machine networks. Initial results are promising for detection, but more work is required for diagnosis. This study performed well using a simple clustering method, KNN, using 2D u-shapelets. However, it did not perform well in clustering with k-means. The study also demonstrates that data-driven methods with 2D u-shapelets have utility in anomaly detection. However, more research is required for the utility of 2D u-shapelets with clustering.

## Acknowledgments

Our research is partially supported by DOE-EE0009026, NSF-SaTC-2019311, DOD-FA8571-21-C-0020 and Aging Aircraft Solutions.

## References

[1] J. C. Balda, A. Mantooth, R. Blum, and P. Tenti, "Cybersecurity and power electronics: Addressing the security vulnerabilities of the internet of things," *IEEE Power Electronics Magazine*, vol. 4, no. 4, pp. 37–43, 2017.

[2] B. Chen, S. Mashayekh, K. L. Butler-Purry, and D. Kundur, "Impact of cyber attacks on transient stability of smart grids with voltage support devices," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[3] Y. Isozaki, S. Yoshizawa, Y. Fujimoto, H. Ishii, I. Ono, T. Onoda, and Y. Hayashi, "Detection of cyber attacks against voltage control in distribution power grids with pvs," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1824–1835, 2016.

[4] Y. Cao, K. Davis, and S. Zonouz, "A framework of smart and secure power electronics driven hvac thermal inertia in distributed power systems," in *2018 IEEE Green Technologies Conference (GreenTech)*. IEEE, 2018, pp. 127–132.

[5] X. Liu, M. Shahidehpour, Y. Cao, L. Wu, W. Wei, and X. Liu, "Microgrid risk analysis considering the impact of cyber attacks on solar pv and ess control systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1330–1339, 2017.

[6] H. Zhang, W. Meng, J. Qi, X. Wang, and W. X. Zheng, "Distributed load sharing under false data injection attack in an inverter-based microgrid," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1543–1551, 2019.

[7] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.

[8] S. Sahoo, S. Mishra, J. C.-H. Peng, and T. Dragicevic, "A stealth cyber attack detection strategy for dc microgrids," *IEEE Transactions on Power Electronics*, 2018.

[9] O. A. Beg, T. T. Johnson, and A. Davoudi, "Detection of false-data injection attacks in cyber-physical dc microgrids," *IEEE Transactions on industrial informatics*, vol. 13, no. 5, pp. 2693–2703, 2017.

[10] H. Liu, F. Hussain, Y. Shen, S. Arif, A. Nazir, and M. Abubakar, "Complex power quality disturbances classification via curvelet transform and deep learning," *Electric Power Systems Research*, vol. 163, pp. 1–9, 2018.

[11] D. D. Ferreira, J. M. de Seixas, A. S. Cerqueira, C. A. Duque, M. H. J. Bollen, and P. F. Ribeiro, "A new power quality deviation index based on principal curves," *Electric Power Systems Research*, vol. 125, pp. 8–14, 2015.

[12] O. P. Mahela, A. G. Shaik, and N. Gupta, "A critical review of detection and classification of power quality events," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 495–505, 2015.

[13] L. Ye and E. Keogh, "Time series shapelets: a new primitive for data mining," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 947–956.

[14] J. Zakaria, A. Mueen, and E. Keogh, "Clustering time series using unsupervised-shapelets," in *2012 IEEE 12th International Conference on Data Mining*. IEEE, 2012, pp. 785–794.

[15] L. Ulanova, N. Begum, and E. Keogh, "Scalable clustering of time series with u-shapelets," in *Proceedings of the 2015 SIAM international conference on data mining*. SIAM, 2015, pp. 900–908.

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

[17] N. B. Roa, L. Travé-Massuyès, and V. H. Grisales-Palacio, "Dyclee: Dynamic clustering for tracking evolving environments," *Pattern Recognition*, vol. 94, pp. 162–186, 2019.

[18] B. Yang, L. Guo, F. Li, J. Ye, and W. Song, "Vulnerability assessments of electric drive systems due to sensor data integrity attacks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3301–3310, 2019.

[19] ——, "Impact analysis of data integrity attacks on power electronics and electric drives," in *IEEE Transportation Electrification Conference and Expo (ITEC)*, 2019, pp. 1–6.

[20] R. D. Trevizan, C. Ruben, K. Nagaraj, L. L. Ibukun, A. C. Starke, A. S. Bretas, J. McNair, and A. Zare, "Data-driven physics-based solution for false data injection diagnosis in smart grids," in *2019 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2019, pp. 1–5.

[21] K. G. Lore, D. M. Shila, and L. Ren, "Detecting data integrity attacks on correlated solar farms using multi-layer data driven algorithm," in *2018 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2018, pp. 1–9.

[22] M. R. Habibi, H. R. Baghaee, T. Dragi ˇcevi, F. Blaabjerg *et al.*, "Detection of false data injection cyber-attacks in dc microgrids based on recurrent neural networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2020.

[23] F. Li, J. Clemente, and W. Song, "Non-intrusive and non-contact sleep monitoring with seismometer," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 449–453.

[24] S. Roychoudhury, F. Zhou, and Z. Obradovic, "Leveraging subsequence-orders for univariate and multivariate time-series classification," in *Proceedings of the 2019 SIAM International Conference on Data Mining*. SIAM, 2019, pp. 495–503.

[25] S. Stelter, G. Bartels, and M. Beetz, "Multidimensional time-series shapelets reliably detect and classify contact events in force measurements of wiping actions," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 320–327, 2017.

[26] J. Grabocka, M. Wistuba, and L. Schmidt-Thieme, "Fast classification of univariate and multivariate time series through shapelet discovery," *Knowledge and information systems*, vol. 49, no. 2, pp. 429–454, 2016.

[27] J. Zakaria, A. Mueen, E. Keogh, and N. Young, "Accelerating the discovery of unsupervised-shapelets," *Data mining and knowledge discovery*, vol. 30, no. 1, pp. 243–281, 2016.

[28] Q. Zhang, J. Wu, P. Zhang, G. Long, and C. Zhang, "Salient subsequence learning for time series clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 9, pp. 2193–2207, 2018.

[29] Q. Zhang, J. Wu, H. Yang, Y. Tian, and C. Zhang, "Unsupervised feature learning from time series." in *IJCAI*. New York, USA, 2016, pp. 2322–2328.

## BIOGRAPHY

**Stephen Coshatt** received his B.S. and M.S. degrees in computer engineering from Mercer University. He also received a MS in Management Information Systems from Georgia College State University. He is currently a civil service employee with Robins AFB. He is attending UGA to pursue a PhD in Engineering as a DoD SMART Scholar. Before his current role as a student, he spent over 10 years as an engineer with Robins AFB. He has worked in software engineering, process management, and cyber-security positions. He currently holds CISSP and CISM certifications in cyber-security.

**James Hill** is President at Aging Aircraft Solutions, a small business based in Georgia. He has over 20 years of experience managing aerospace engineering and research development DoD programs of over $50 million in total value. He has been the principal investigator on an Air Force SBIR and STTR topics and successfully achieved a phase III awards. For previous companies he has managed multiple research and development projects sponsored by agencies such as DARPA and the Office of Naval Research. Mr. Hill has a B.Sc. in Biochemistry from Georgetown University and a Master's in Business Administration (MBA) from Emory University's Goizueta Business School. Mr. Hill is a proven innovator on research and development programs claiming lead inventor on two issued U.S. Patents.

**WenZhan Song** is Georgia Power Mickey A. Brown Professor of Engineering and Founding Director of Center for Cyber-Physical Systems at the University of Georgia. Dr. Song's research expertise is on networked sensing, computing and security and their applications in health, energy and environment monitoring. He received numerous awards from his university and professional society, such as NSF CAREER Award, Outstanding Research Contribution Award, Chancellor Research Excellence Award, Mark Weiser Best Paper Award. Dr. Song serves many premium IEEE conferences (such as IEEE INFOCOM, IEEE GLOBECOM) and journals (such as IEEE Internet of Things, ACM Transaction on Sensor Networks) as editor, chair or TPC member.

**Jin Ye** (IEEE S'13-M'14-SM'16) received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2008 and 2011, respectively. She also received her Ph.D. degree in electrical engineering from McMaster University, Hamilton, Ontario, Canada in 2014. She is currently an assistant professor of electrical engineering and the director of the intelligent power electronics and electric machines laboratory at the University of Georgia. She is a general chair of 2019 IEEE Transportation Electrification Conference and Expo (ITEC), a publication chair and women in engineering chair of 2019/2020 IEEE Energy Conversion Congress and Expo (ECCE). She is an associate editor for *IEEE Transactions on Power Electronics, IEEE Open Journal of Power Electronics, IEEE Transactions on Transportation Electrification and IEEE Transactions on Vehicular Technology*. Her main research areas include power electronics, electric machines, energy management systems, smart grids, electrified transportation, and cyber-physical systems.

**Bowen Yang** received the B.S. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently pursuing the Ph.D. degree with the University of Georgia, Athens, GA, USA. He is also a Research Assistant with the University of Georgia, USA. His current research interests include advanced control for power electronics and electric machines, energy management system, and cyber-physical security for intelligent electric drives.

**Feraidoon Zahiri** is an engineer at the US Air Force, received his Master of Science degree in Engineering Mechanics from Penn State University. He has more than 20 years of experience in technology research, development, and transition in numerous scientific and engineering disciplines. He has overseen large-scale decision-support projects incorporating analytics, simulation, and visualization for resource allocation, system troubleshooting and repair, and data-driven predictive maintenance. He was selected as a winner of the Small Business Technology Council's (SBTC) 2017 "Champion of Small Business Technology Commercialization" Award, for

his role in filling a critical technology gap in the Air Force Sustainment community. He has played a critical role in providing opportunities for small business to promote the technology across all DoD services, boosting their technology transition even further.