IoT Load Classification and Anomaly Warning in ELV DC Picogrids Using Hierarchical Extended *k*-Nearest Neighbors

Yang Thee Quek, *Member, IEEE*, Wai Lok Woo, *Senior Member, IEEE*, and Logenthiran Thillainathan, *Senior Member, IEEE*

***Abstract*—The remote monitoring of electrical systems has progressed beyond the need of knowing how much energy is consumed. As the maintenance procedure has evolved from reactive to preventive to predictive, there is a growing demand to know what appliances reside in the circuit (classification) and a need to know whether any appliance requires attention and maintenance (anomaly warning). Targeting at the increasing penetration of dc appliances and equipment in households and offices, the described low-cost solution consists of multiple distributed slave meters with a single master computer for extra low voltage dc picogrids. The slave meter acquires the current and voltage waveform from the cable of interest, conditions the data, and extracts four features per window block that are sent remotely to the master computer. The proposed solution uses a hierarchical extended *k*-nearest neighbors (HE-*k*NNs) technique that exploits the use of distance in *k*NN algorithm and considers a window block instead of individual data point for classification and anomaly warning to trigger the attention of the user. This solution can be used as an *ad hoc* standalone investigation of suspicious circuit or further expanded to several circuits in a building or vicinity to monitor the network. The solution can also be implemented as part of an Internet of Things application. This article presents the successful implementation of the HE-*k*NN technique in three different circuits: 1) lighting; 2) air-conditioning; and 3) multiple load dc picogrids with accuracy of over 93%. Its performance is superior over other anomaly warning techniques with the same set of data.**

***Index Terms*—Anomaly warning, artificial intelligence (AI), dc grid, extra low voltage (ELV), *k*-nearest neighbors (*k*NNs), load classification.**

I. Introduction

THE MODERN power system has gone through many changes over the past few years. Apart from providing power from the source to various loads, its primary goal is to ensure reliability and stability of the power system as power-grid transforms and its loads evolve over the years. The maintenance of the power grid has shifted from the traditional reactive maintenance to the scheduled preventive maintenance to the artificial intelligence (AI)-enabled predictive maintenance. Fault prediction and anomaly warning increases operational reliability and stability, and can help to prevent the unnecessary losses.

The traditional meter only provides data on the energy consumed by the user; this meter requires the service provider and the user to manually record the data. However, recent progress in metering combined different hardware devices and software programs, along with information and communication technology (ICT) and added intelligence. The smart meter can provide useful information wirelessly to the users to optimize consumption efficiency and make informed decisions while performing necessary maintenance or repair. The research development in this area has been greatly encouraged by the recent growth in the Internet of Things (IoT) and smart cities [1], [2]. Yu *et al.* [3] discussed the distributed online energy management for data centers and electric vehicles in smart dc grids. Several low-cost smart electrical meters have also been proposed in [4]–[6]. An anomaly is a deviation from the normal conditions of the various operating modes. Anomaly detection is the process of finding instances in a data set that are different from the majority of the data. There are two basic assumptions for anomaly detection; anomalies only occur very rarely in data and their features do differ from the normal instances significantly [7]. The aim of anomaly detection is to provide some useful information where no information was previously attainable. An anomaly warning can be triggered by predicting potential fault through the analysis of historical data so measures or maintenance can be taken to prevent accidents or to ensure recovery of the system. Constant monitoring of the grid and system with data analytics provides useful information and helps the users to make reasonable decisions to prevent potential faults and reduce adverse effects. The method based on protective relays and circuit breakers addresses the problem after the faults have happened; it cannot do much for predictive maintenance. Consequently, the anomaly detection warning is a significant and valuable research subject.

In the past decade, extra low voltage (ELV) dc loads, such as personal gadgets, household appliances, and office equipment, have been rising in numbers and their power consumption of

Manuscript received October 9, 2018; revised February 8, 2019, May 18, 2019, July 10, 2019, and September 14, 2019; accepted September 25, 2019. Date of publication October 9, 2019; date of current version February 11, 2020. *(Corresponding author: Yang Thee Quek.)*

Y. T. Quek is with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K. (e-mail: q.yang-thee@newcastle.ac.uk).

W. L. Woo is with the Department of Computer and Information Sciences, Northumbria University, Newcastle upon Tyne NE1 8ST, U.K. (e-mail:wai.l.woo@northumbria.ac.uk).

L. Thillainathan is with the School of Engineering and Technology, University of Washington Tacoma, Tacoma, WA 98402 USA (e-mail: loganth@uw.edu).

Digital Object Identifier 10.1109/JIOT.2019.2945425

buildings. There has also been an escalated interest in the use of dc power grids in segregated power systems [8]; with the growing installation of photovoltaic (PV) panels and battery energy storages, the advantage of dc power grids is obvious with a signification reduction in conversion losses since loads, sources, and energy storage can be connected through simpler and more efficient power electronics interfaces [9]. The complex, expensive, and lossy power factor correction (PFC) feature can also be avoided. In addition, a dc power grid, in general, has no worries when it comes to harmonics pollutions. Most of the past studies of power system's state were on ac electrical grids [10], [11]. As the ELV dc loads get smaller, the dc power grids are also shrinking from micro to nano to picogrid [12]. There has not been much attention paid on dc, thus motivating the research on state detection and anomalies warning in dc grids.

Several AI techniques have been proposed and applied in power systems, including the use of long short-term memory (LSTM) and support vector machine (SVM) for trip fault prediction in power systems [13], a Bayesian-based approach for short-term steady-state forecast of a smart grid [14] and k-means clustering with k-nearest neighbors (kNNs) in equipment identification [15]. There are also anomaly detection approaches based on clusters and neighborhoods. Alvarez *et al.* [16] proposed a clustering-based anomaly detection in multiview data, where the anomalies are detected by comparing affinity vectors in multiple views. Kuang and Zulkernine [17] detected intrusion and anomaly by using a combined strangeness and isolation measure KNN (CSI-kNNs) algorithm. These neighborhood-based approaches are not commonly used in dc power systems and they did not take into account of burst error and consecutive error occurrence detection. The burst error and consecutive error occurrence detection are necessary to avoid false-positive prediction in real-life data. Due to the recent growth in dc power systems, there is an increasing interest in AI applications proposed in dc power systems, for example, the use of wavelet and artificial neural networks on medium dc voltage power system [18] and fuzzy control for energy management for a dc microgrid systems [19]. In addition to fault detection, there has also been some interest in load identification and disaggregation as seen in [20] and [21].

This article describes a low-cost smart meter solution for ELV dc picogrid that consists of a low-cost distributed slave meter to acquire and extract features from the current value and a master computer that, using a new hierarchical extended kNN (HE-kNN) technique, performs classification of operating modes or loads and signals a trigger warning if abnormal behavior is detected in the current waveform. The HE-kNN is a semisupervised anomaly detection approach that uses an anomaly free training set only consisting of the normal classes. There is a separated test set that comprises normal records and anomalies. The HE-kNN algorithm exploits the use of distance in the clustering-based approach and differentiates itself from other approaches with the innovative addition of burst error anomaly and consecutive error occurrence detection.

This article is organized as follows. Section II describes the purpose of ELV dc picogrids. Section III presents an overview of the monitoring solution. Section IV discusses the feature extraction from the current waveform and Section V elaborates on the task of the master computer-load classification and anomaly warning using the HE-kNN technique. Section VI discusses the results of the solution in three different types of ELV dc pic-grids. Section VII concludes this article.

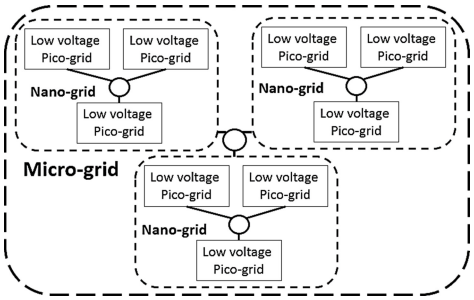

Fig. 1. ELV dc picogrids forming into a microgrid.

II. Extra Low Voltage DC Picogrid

The International Electrotechnical Commission (IEC) and the U.K. Institution of Engineering and Technology (IET) (BS7671:2008) define an ELV device or circuit as one in which the electrical potential between the conductor or the electrical conductor and earth does not exceed 50Vac or 120Vdc (ripple free). In recent years, there has been a significant rise in the number of small appliances in households and offices that are powered by extra-low-dc voltage. Some of these appliances and equipment are low in cost and perform simple tasks, such as turning on and off appliances (for example lights and fans). There is no incentive for them to possess power monitoring features or anomaly detection intelligence as these additional features increase their price significantly. These ELV loads typically share a single power supply, but due to the limited power of the power supply, the number of appliances sharing it is usually very small. The limited power is the fundamental reason for the formation of an ELV dc picogrid. These picogrids can be bundled together to form nanogrids and mini-grids. These nanogrids can then be clustered into bigger microgrids as seen in Fig. 1. This article suggests that the monitoring and management of the grids can be done bottom-up where the monitoring starts at picogrids and progresses up in level and scale.

III. Solution Overview

The technique described in this article focuses on studying application in ELV dc picogrids. The system setup allows the users to perform remote monitoring of the ELV dc picogrid by sensing the main line's current waveform using the slave meter and wirelessly sending the extracted features to the master computer that is at a distance.

The described monitoring system can be set up as a many-to-one system by exploiting the ubiquitous availability of the Wi-Fi network in buildings. The slave meter was designed to be low cost and satisfy the low power requirement; thus, it can be installed in multiple locations in a building and they

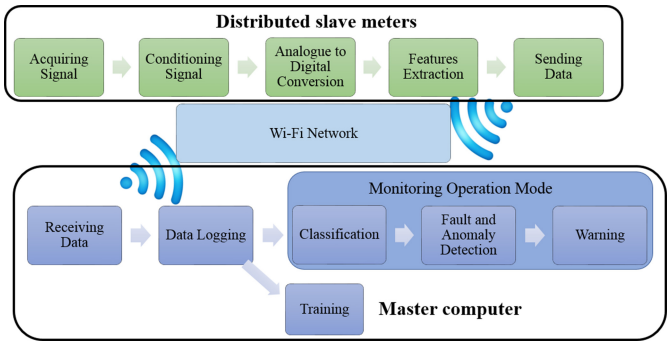

Fig. 2. System overview of the solution.

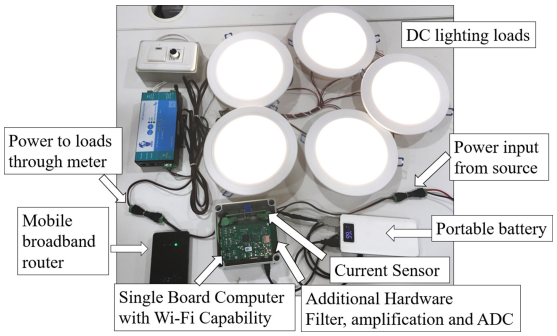

Fig. 3. Example of the slave meter in an ELV dc picogrid of lightings.

can be identified by their IP addresses and communicate with a distant master computer.

The system can also be installed as an *ad hoc* monitoring system for a suspicious ELV dc picogrid when powered with a portable battery bank and has its Wi-Fi local area network (WLAN) wireless network covered by a mobile broadband router. This property eliminates the need to access any power sockets and wireless networks during the investigation of the grid. The following section describes the design of the slave meter, followed by the design of the master computer. Fig. 2 shows the tasks of the slave meter and the master computer in the system overview.

IV. Features Extraction in Distributed Slave Meters

The slave meters designed in this project are meant to be distributed as a functional part of a many slaves to one master system, thus they are to be affordable and portable. Instead of expensive high-end power quality analysis equipment, low-cost single-board small computers are chosen as the platform for the slave meters because they have reasonable computing power, low cost, and have high portability. They consume low power at approximately 240 mA (1.2 W) and can also be powered up with ease as they just require a 5-V portable battery. The duty of the slave meter is to sense the current waveform, condition the signal, extract the features, and send the extracted data wirelessly to the master computer. With the aid of additional low-cost hardware and optimized software, the single-board small computers can perform the above tasks. An example of the slave meter in the monitoring system is shown in Fig. 3.

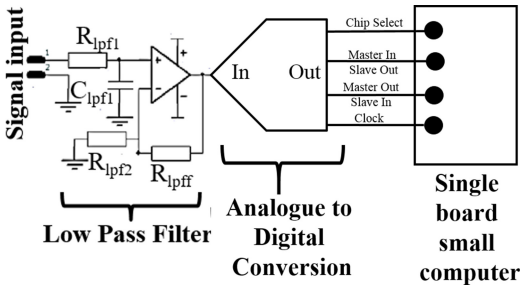

Fig. 4. Hardware system diagram of the slave meter.

Using the closed-loop hall effect technology, the closed-loop current transducers used in this article can measure current over a wide range of frequencies, including the dc current frequency. The transducers provide contact-free coupling to the current that needs to be measured, safe galvanic isolation and high reliability. They can provide fast, accurate, and high-resolution images of the primary current. The selected current transducers can work with a single 5-V power supply with primary nominal current measurement up to 25 A.

Assuming negligible error, the output voltage V_o of the current transducer is related to the primary current linkage Θ_p by sensitivity G

$$V_o = G\Theta_p. \tag{1}$$

The current linkage Θ_p is related to the number of primary turns N_p and the primary current I_p

$$\Theta_p = N_p I_p. \tag{2}$$

Since G and N_p can be extracted from the specification of the transducer, V_o can be determined from I_p

$$V_o = GN_p I_p. \tag{3}$$

Within the range of 0–6 A, the output voltage signal V_o and I_p can be assumed to have a linear relationship.

A low-pass filter is added after the sensing of the current transducer to allow the passing of the lower frequencies up to the cutoff frequency, attenuating the higher frequencies that are above the cutoff frequency f_c. The filter can be implemented using low power single operational amplifier LM321 or low power dual operational amplifier LM358 (4-6). Both LM321 and LM358 require a 5-V dc power supply that can be provided by the portable battery bank. The cutoff frequency f_c selected in this article is approximately 50 Hz, which is 1/2 of the sampling frequency f_s, which is 100 Hz. The operational amplifier is also capable of creating a unity gain follower by setting the R_{lpff} to 0 Ω. Although amplification is not required, the unity gain follower provides the important benefit of isolating the input side of the circuit from the output side of the circuit.

The single-board small computer does not possess the ability to receive analog input thus an analog to digital conversion (ADC) is required. MCP3008 is a 10-bit ADC that operates over a broad range of voltages. It can communicate with the single-board small computer using the serial peripheral interface (SPI) protocol. MCP3008 is used in the slave meter to receive the filtered analog signal from the operational

amplifiers and communicate it in digital form to the single-board small computer. Fig. 4 shows the system drawing of the slave meter

$$f_c = \frac{1}{2\pi R_{lpf1} C_{lpf1}} \tag{4}$$

$$A_{lpf} = \left(1 + \frac{R_{lpff}}{R_{lpf2}}\right) \tag{5}$$

$$A_v = \frac{V_{o,lp}}{V_i} = \frac{A_{lpf}}{\sqrt{1 + \left(\frac{f}{f_c}\right)^2}}. \tag{6}$$

The hardware saves the small computer much calculation resources thus allowing it to focus on the extraction of the features of the signal and on sending it to the master computer. The software in the slave meter converts the acquired signal from SPI into electrical current readings by using the least significant bit (LSB) value of the ADC MCP3008 as the value of each digital step

$$\text{Digital step value} = \text{LSB} = \frac{\text{Range}}{2^n - 1}. \tag{7}$$

Range is the difference between the maximum and minimum values of the acquired data and n is the number of bits.

The slave meter reads the signal at 100 Hz and extracts four features [mean μ, variance σ^2, largest gradient $\max(\bar{\imath})$, and range R] from each 1-s block of 100 data points. The data samples of the raw current waveform are placed in a 1-D matrix $\boldsymbol{I}$ as seen below, where m is the total number of samples logged

$$\boldsymbol{I} = [i_1, i_2, i_3, \ldots, i_m]. \tag{8}$$

Given i_n is the element in the input signal $\boldsymbol{I}$ and M is the number of points in the window subset. μ in

$$\mu = \frac{1}{M} \sum_{j=n-M+1}^{n} i_j \tag{9}$$

is the computed mean and σ^2 in

$$\sigma 2 = \frac{\sum_{n=1}^{m} (i_n - \mu)^2}{m - 1} \tag{10}$$

is the computed variance. The largest gradient $i_{\text{grad,max}}$ is computed by finding the largest difference between subsequent data points $\bar{\imath}$ in

$$\bar{\imath} = \frac{\Delta i}{\Delta t} = \frac{i_n - i_{n-1}}{T} \tag{11}$$

and range R is the difference between the maximum and minimum values of the window

$$R = [\max(I) - \min(I)]. \tag{12}$$

These data are packaged into a packet and transmitted to the master computer once every second via Wi-Fi for higher level processing. The operation of the slave meter is the same for both training mode and routine operation mode. This similarity reduces the manhandling of the slave meters.

V. Classification and Anomaly Warnings in Master Computer

The master computer can be a laptop or workstation depending on the resource requirement of the task. A workstation is necessary to monitor multiple meters. A laptop will be sufficient for *ad hoc* single point monitoring of suspicious dc picogrid. The software used in this article is MATLAB R2017B. There are two modes in the master computer, namely, the training mode and the monitoring operation mode.

A. Training Mode

During the training mode, the user starts the training software that receives data from the slave meter and capture the 4-feature data point of the various normal operating modes of the interested dc picogrid. These data points form the elements in the overall kNN training set $X = [x_1, x_2, \ldots, x_N] \in \Re$ where N is the number of training samples, $x_i = [x_{1,i}, x_{2,i}, \ldots, x_{s,i}] \in \Re$, is a vector that represents the ith training sample and s is the number of extracted feature from the data set. As this is a supervised machine learning technique, the users need to indicate the operating mode of each data point, cluster them and the training software then label them with classes $C = [c_1, c_2, \ldots, c_M] \in \Re$, where M is the number of cluster or class. The users need to ensure no anomaly occurs during the training phases and the features need to be captured for all normal operating modes including the no load situation. The extracted features should be informative, nonredundant, and a good descriptive of the data set. In the event that the user finds that the extracted features are not a good representation of the class or cluster, the training phase can be reperformed. It is faster and resource efficient to use extracted features instead of raw data points.

B. Monitoring Operation Mode

After the training phase, the user needs to switch to the monitoring operation mode for routine operation to monitor and inspect the remote dc picogrid. The classification of loads in the master computer is performed using kNN. kNN is a supervised machine learning algorithm that is instance-based—it is based on the computation of the k nearest training elements in the overall training set and on the election of the class through majority voting on the labels of the nearest elements. The training phase took care of the clustering and labeling of the training elements.

kNN is a distance-based algorithm and this article uses Euclidean distance to determine the closeness between the elements. Given that all training samples are stored in M clusters, the number of training samples in the $m(1 \leq m \leq M)$ cluster is C_m. For the i training sample x_i in the m cluster, the kNN rule performed in this cluster using the following equation for distance $d_{i,j}$:

$$d_{i,j} = \|x_i - x_j\| = \sqrt{\sum_{s=1}^{S} x_{i,s}^2 \mp x_{j,s}^2}, \quad j = 1, 2, \ldots, C_m; j \neq i. \tag{13}$$

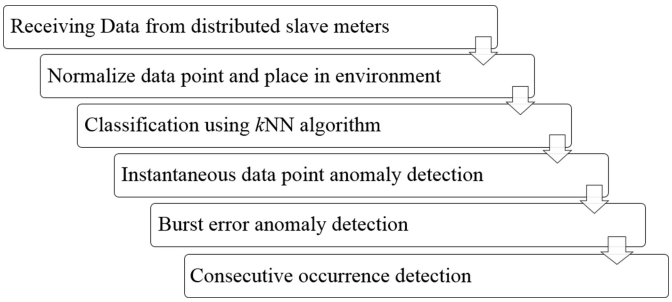

Fig. 5. Hierarchy processes in HE-*k*NN technique.

As the features were of different scale, normalization of the features' values was performed to avoid over reliance on any dimension. The normalization can be performed as follows:

$$x_{\text{new}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{14}$$

The *k*NN algorithm is nonparametric as it does not make any explicit assumptions on the model or function. However, it requires potentially large data set thus possibly requiring higher computational cost and resources. This requirement is also one of the reasons why it is advisable to perform the computation using the more powerful master computer. It is a robust and versatile machine learning technique that is commonly used in solving classification problems. The function of the traditional *k*NN algorithm is extended enhanced in this project to perform anomaly warning detection in addition to its usual classification. The next section describes the hierarchical extended *k*NNs (HE-*k*NNs) technique.

C. Hierarchical Extended k-Nearest Neighbors

This article exploits the use of distance in the *k*NN algorithm for anomaly warning and fault detection. In addition to the usual classification process of the *k*NN algorithm, several additional steps are added to enhance the algorithm to perform meaningful anomaly detection, thus forming a hierarchy process as seen in Fig. 5.

During the monitoring operation mode, the first layer of the HE-*k*NN receives the data point from the slave and convert the four features into a vector as the test object. The second layer is to normalize the test object and places it into the test environment with the trained and labeled elements. Depending on the k value indicated by the user, the third layer of the technique performs the classification by the majority vote of the *k*NNs. The selection of k value is important. It is recommended to run through a range of possible k values and find the most appropriate k value with high accuracies in classification, taking into consideration that a small value of k means that noise will have significant influence on the result and large value of k makes it computationally expensive. The fourth to sixth layers of the HE-*k*NN perform anomalies detection and trigger warnings to the user.

Two enhancements were applied in the fourth layer of the HE-*k*NN technique for instantaneous error detection. The first enhancement technique defines the centroids of each identified cluster and set an acceptable boundary around the centroid

$$\text{Centroid}_m = \left[\bar{s}_{\text{mean},m}, \bar{s}_{\text{var},m}, \bar{s}_{\text{range},m}, \bar{s}_{\text{grad},m}\right]$$
$$\text{where } \overline{s_m} = \frac{\sum_{i=1}^{N_m} s_{i,m}}{\text{size}(m|}. \tag{15}$$

After the test object is labeled with a cluster's class, the distance between the test object and that cluster's centroid d_{testc}, is calculated as

$$d_{\text{testc}} = \left\| x_{\text{test},s} - x_{\text{centroid},m,s} \right\|. \tag{16}$$

If the test object falls out of the boundary of its labeled centroid, it is flagged as an instantaneous error.

The boundary of the cluster is set to be the sum of the mean distance, $\overline{d_m}$, and three times the standard deviations σ_m of all the training elements with the centroid in the cluster. This limit is with reference to the statistical process control (SPC) that considers a process to be in control and stable if the measured value is within the control limits of mean±3 standard deviations from the mean and [22], [23]

$$d_{\text{limit},m} = \overline{d_m} + 3\sigma_m \tag{17}$$
$$\text{Anomaly warning} = \begin{cases} \text{YES}, & \text{if } d_{\text{testc}} > d_{\text{limit},m} \\ \text{NO}, & \text{otherwise.} \end{cases} \tag{18}$$

The second enhancement technique uses the average distance between the trained elements in the same cluster as the constraint for abnormal operation. This technique assumes the closeness of the test object to its INNs, D_{test}, as an indication of whether the test object is normal

$$D_{\text{test}} = \frac{1}{k} \sum_{j=1}^{k} d_j. \tag{19}$$

For every element in the cluster, the average distance between it and its neighbors are calculated to find the average of all the distances $\overline{D}_m$

$$\overline{D}_m = \frac{1}{\text{size}(m)k} \sum_{j=1}^{\text{size}(m)} \sum_{j=1}^{k} d_j. \tag{20}$$

Referring to the SPC, the instantaneous error criterion $\dot{D}_{\text{limit},m}$ is set the sum of the average distance $\overline{D}_m$ and three times the standard deviation σ_m

$$\dot{D}_{\text{limit},m} = \overline{D_m} + 3\sigma_m \tag{21}$$
$$\text{Anomaly warning} = \begin{cases} \text{YES}, & \text{if } D_{\text{test}} > \dot{D}_{\text{limit},m} \\ \text{NO}, & \text{otherwise.} \end{cases} \tag{22}$$

Fig. 6 illustrates the examples for both anomaly warning conditions.

Although the two methods presented can flag instantaneous errors, an instantaneous error is not a good indicator of an anomaly as it can be caused by noise or may occur when there is a change in the operation mode. Further processes are required to ensure that the anomalies flagged by the technique are meaningful and significant enough to trigger warnings.

The fifth layer in the HE-*k*NN is to employ the burst error anomalies detection. In this step, the user can enter the window block size w_b and the threshold number of error

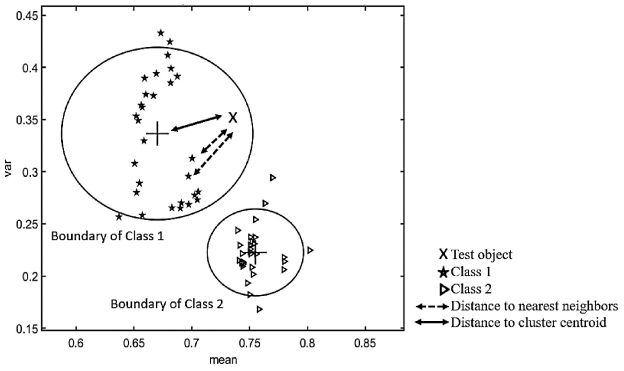

Fig. 6. Anomaly warning criteria in HE-*k*NN.

in the window block or the burst error limit, e_b, to be considered as an anomaly. The error may not occur continuously. For example, let us consider that the user entered $w_b = 20$ (in this case, it is a 20-s block), and $e_b = 5$. An anomaly indicator is recorded if the number of instantaneous errors that occurs in the 20-s block is 5 or more, else it is considered as normal operation with the mode previously classified.

In addition, the steps in detecting burst error anomalies also consider the frequency interchanging of operation modes as an error since there should not be frequent changing of modes in practical scenarios of a system. For example, it is rare to have the cooling function of an air-conditioner turning on and off at a continuous interval of 2–3 s. These errors should indicate to the user the need to take a closer look at the air-conditioner if they persist over a long time. Using the previous example of $w_b = 20$ and $e_b = 5$, an anomaly indicator would be recorded if there are five or more changes in the operation mode in the 20-s block.

The final layer in the HE-*k*NN technique is to consider the consecutive occurrence of the anomalies identified in the previous layer c_e. The user decides the desired number of consecutive occurrences of anomalies count to be considered as a meaningful situation for the user to take note and consider inspecting the ELV dc picogrid manually. This layer can lead to predictive maintenance before the dc picogrid actually breaks down for reactive action and maintenance.

The next section discusses the application of HE-*k*NN solution and its results on various ELV dc picogrids.

VI. Results and Discussion

The HE-*k*NN solution was tested in the remote monitoring of several ELV dc picogrids. The slave meter that was attached to the ELV dc picogrids extracted four features from the 1-s block of 100-Hz signals and they were sent via Wi-Fi to the master computer which would perform load classification and anomaly warning using the HE-*k*NN technique. For the purposes of this article, this section describes the application of the proposed setup in 3 ELV dc picogrids that highlight the capabilities of the HE-*k*NN technique in dc environments for monitoring and diagnosing. Each of these grids was injected with faults and anomalies to be detected and triggered as a warning for predictive maintenance.

TABLE I
Class Allocation for the ELV DC Lighting Grid

Class	1	2	3	4	5	6	7
Loads	No load	1 light	2 lights	3 lights	4 lights	5 lights	6 lights

**Anomalies are indicated as class -1*

The three ELV dc picogrids are as follows.

1) A dc lighting grid consisting of five LED downlights.
2) A dc single split air-conditioner with a wall mount indoor unit and an outdoor unit.
3) A dc grid with three different loads of phone, LCD TV, and laptop.

A. Application of HE-kNN on ELV DC Lightings Grid of Five LED Downlights

The recent advancement in LED technology has spurred the lighting market turning it from ac to dc lighting [24]. Lighting is one of the most common building loads. It has always been a challenge to monitor and manage the lighting system in a building. The common methods to identify spoiled or degraded lights are either through the complaints of tenants or through manual walk-pass inspection by technicians or security guards. These are especially tedious and unnecessary for far locations with low traffic. Relamping is usually necessary for the following two common issues with lights.

1) Degrading light output.
2) Flickering light.

This section describes both scenarios and the application of the setup and HE-*k*NN to detect the anomalies and trigger warnings. Fig. 7(a) shows the current signal waveform acquired by the slave meter. The five lights in the circuit were switched off one by one. The two anomalies were injected in the system; the degrading of one lamp's light output occurred at around 72–97 s and flickering of another lamp occurred at around 206–220 s. Fig. 7(b) shows the extracted mean current data received by the master computer from the slave meter. These anomalies are boxed out in the figures.

The selected k value was 5, the window block size w_b was set at 15, the burst error limit e_b was set at 5, and consecutive error warning c_e was set at 3. The class allocation of the ELV dc lighting grid is shown in Table I. The detected anomaly warnings are marked with class -1.

Fig. 7(c) shows the result after performing classification using traditional *k*NN algorithm. Each data point was labeled via the majority of votes from the *k*NNs. There was no fault detection or anomaly warning features. The period of degraded light output was classified alternatively between class 4 and 5. The flickering period was classified under class 5.

Fig. 7(d) shows the calculated distance of each data point with reference to the labeled cluster's centroid and that information can be used to decide whether an anomaly has occurred. Fig. 7(e) shows that anomaly warnings were triggered when the distance of the datapoint from the centroid exceed the boundary. This is an instantaneous check of individual datapoint without consideration of previous or

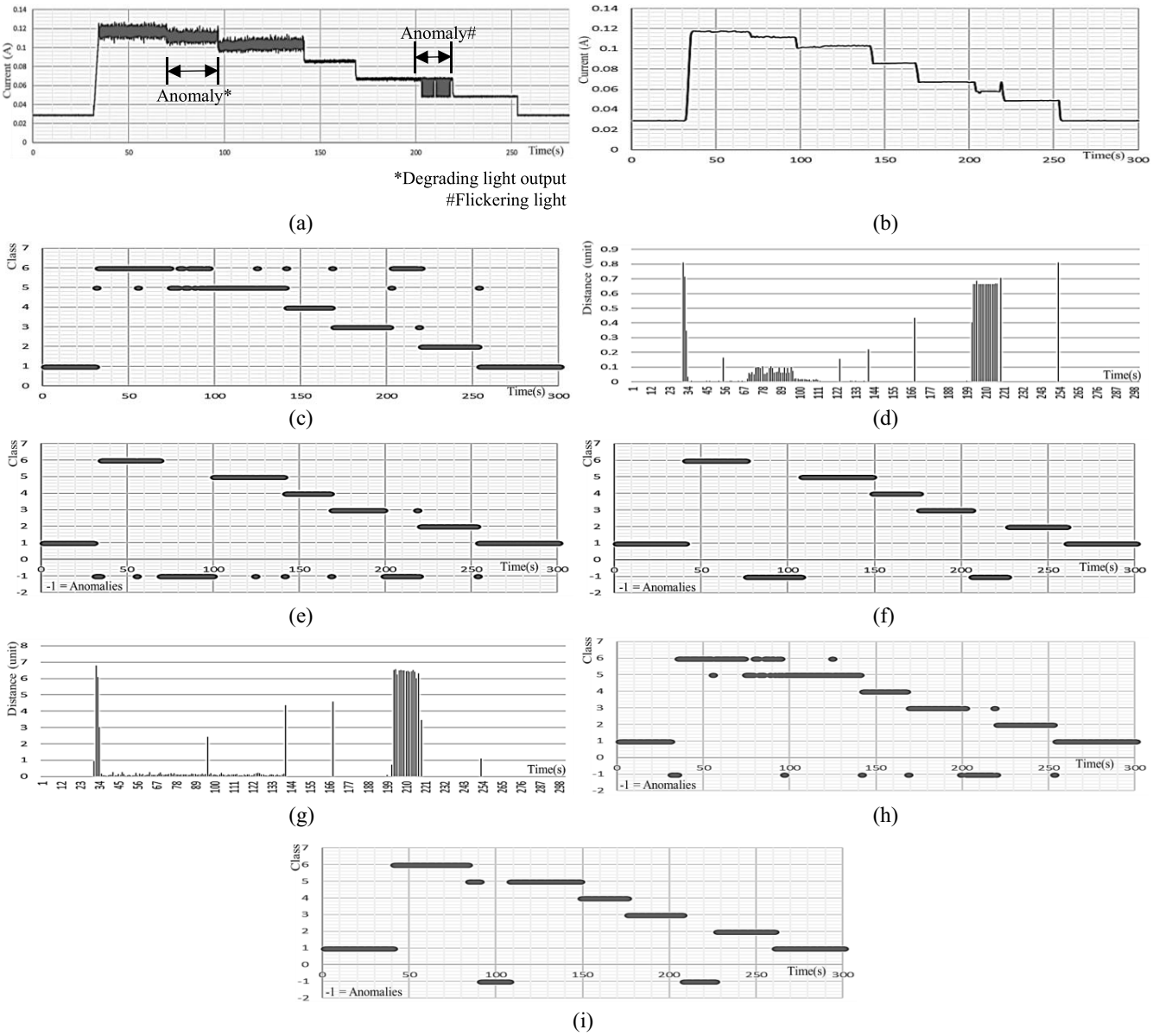

Fig. 7. Waveforms and results from the five LED downlights grid. (a) Current signal sensed by the slave meter. (b) Current signal received by the master computer. (c) Classification by *k*NN. (d) Distance of data points from the labeled cluster's centroid. (e) Instantaneous classification and anomaly warning by the distance from the centroid. (f) HE-*k*NN classification and anomaly warning by the distance from the centroid. (g) Distance of data points from the *k*NNs. (h) Instantaneous classification and anomaly warning by distance from the nearest neighbors. (i) HE-*k*NN classification and anomaly warning by distance from the nearest neighbors.

group datapoints. The results showed a good result except when the individual lights were being switched on and off. There were also some incorrect warnings triggered due to stray readings. Fig. 7(f) shows the results with the implementation of HE-*k*NN technique. As HE-*k*NN considers a window block of data instead of a single individual data point, it removed the stray anomaly warnings and also removed the warnings between the changing of stages. The tradeoff observed was a slight delay in the triggering of warning.

Fig. 7(g) shows the average distance of each data point received by the master computer with reference to its *k*NNs. Fig. 7(h) shows the instantaneous check of individual datapoint where an anomaly is triggered if the average distance of that particular data point is more than the boundary set by the average distance between the elements plus three times its standard deviations. Although the flickering light was detected, the degraded light period was indicated as changing between 5 lights and 4 lights instead of being detected as anomalies. This classification is because the instantaneous data points of the degraded light were within the acceptable boundaries of either class 6 (5 lights) or class 5 (4 lights). There were also stray warnings and switching on and off of lights were also triggered as warnings. In Fig. 7(i), applying HE-*k*NN while considering the distance between *k* neighbors gave better result as compared to the instantaneous anomaly detection. Applying HE-*k*NN triggered warning for the degraded light and flickering light. The stray warning and interchanging between stages

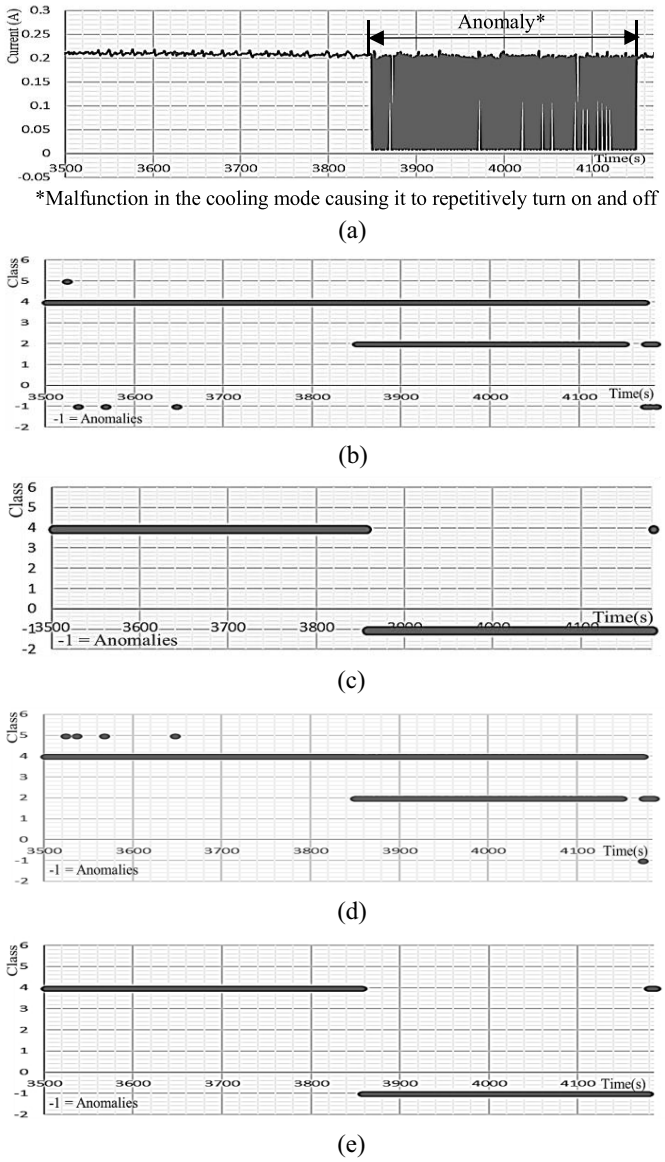

Fig. 8. Waveforms and results from the ELV dc air conditioner on a malfunction of cooling mode. (a) Current signal received by the master computer. (b) Instantaneous classification and anomaly warning by the distance from the centroid. (c) HE-*k*NN classification and anomaly warning by the distance from the centroid. (d) Instantaneous classification and anomaly warning by the distance from the nearest neighbors. (e) HE-*k*NN classification and anomaly warning by the distance from the nearest neighbors.

were also removed because of the additional steps of burst error anomaly detection and consecutive occurrence detection. In a normal operation of a system, there should not be numerous changing of modes within a short period of time, it should be picked up as an anomaly. However, there were delays in the warning and classification.

B. Application of HE-kNN on ELV DC Air-Conditioning

Another major load in a building is heat ventilation and air-conditioning (HVAC). This section describes the application of HE-*k*NN in a 48-V dc-powered single split air-conditioning system. Table II shows the class allocation of the ELV air-conditioning. The two scenarios described here were as follows.

TABLE II
CLASS ALLOCATION FOR THE ELV DC AIR-CONDITIONING

Class	1	2	3	4	5
Loads	No load	Low fan	High fan	Low fan with cooling	High fan with cooling

**Anomalies are indicated as class -1*

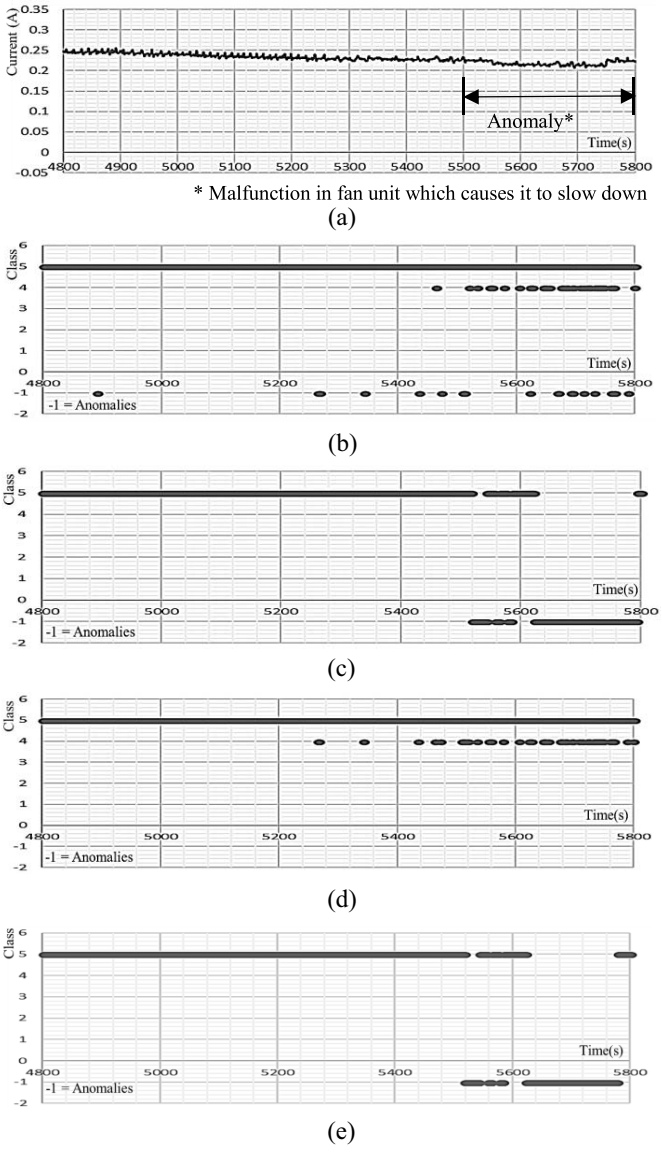

Fig. 9. Waveforms and results from the ELV dc air conditioner on a malfunction in the fan. (a) Current signal received by the master computer. (b) Instantaneous classification and anomaly warning by the distance from the centroid. (c) HE-*k*NN classification and anomaly warning by the distance from the centroid. (d) Instantaneous classification and anomaly warning by the distance from the nearest neighbors. (e) HE-*k*NN classification and anomaly warning by the distance from the nearest neighbors.

1) Malfunction in the cooling mode causing it to repetitively turn on and off.
2) Malfunction in the fan unit that causes it to slow down.

The k value selected was 5, window size w_b was set at 30, burst error limit e_b was set at 5, and consecutive error warning c_e was set at 3.

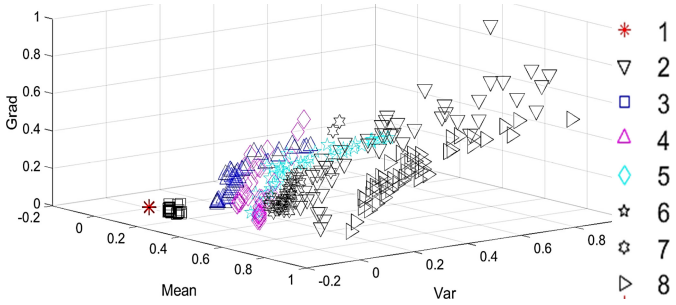

Fig. 10. Clusters for the ELV 3 loads dc grid.

TABLE III
CLASS ALLOCATION FOR THE ELV DC GRID OF THE THREE LOADS

Class	1	2	3	4	5	6	7	8
Phone	OFF	ON	OFF	OFF	ON	ON	OFF	ON
LED TV	OFF	OFF	ON	OFF	ON	OFF	ON	ON
Laptop	OFF	OFF	OFF	ON	OFF	ON	ON	ON

**Anomalies are indicated as class -1*

Fig. 8(a) shows the current waveform for a period of the dc air condition where there was a repetitive decrease and increase of the current consumption. This repetition was caused by the mode of the dc air conditioner being switched between low fan mode (class 2) and low fan with cooling mode (class 4). This was an anomaly in the operation of the dc air conditioner, but it was not picked up as one when the distance of individual data points was considered in the algorithms, as shown in Fig. 8(b) and (d). The individual data points were within either the acceptable boundary of class 2 or 4 thus no anomaly was triggered. However, the HE-*k*NN algorithm, as seen in Fig. 8(c) and (e), was able to correctly identify the period as an anomaly by considering additional steps of burst error anomaly detection and consecutive occurrence detection in both distance from the centroid and distance from *k*NNs. Stray warnings were also removed.

The second scenario of the malfunction in the dc air-conditioner was the reduction in fan speed. This malfunction occurred in the high fan with cooling mode (class 5). As seen in Fig. 9(a), there was a gradual decrease in the current consumption of the system as the fan slowed down. Both techniques using instantaneous classification and anomaly warnings were not able to identify it as an anomaly. The results also show that the technique using distance from the centroid [Fig. 9(b)] as compared to using the distance from nearest neighbors [Fig. 9(d)] produced slightly better results as it identified the anomalies intermediately.

The HE-*k*NN technique was able to trigger a warning for anomalies in both applications of distance from the centroid [Fig. 9(c)] and the distance from the nearest neighbors [Fig. 9(e)].

C. Application of HE-kNN on ELV DC Picogrid With Three Loads

This section describes the application of the HE-*k*NN anomaly technique on an ELV dc picogrid with three dc loads, namely, mobile phone (5 V), LCD TV (12 V), and laptop

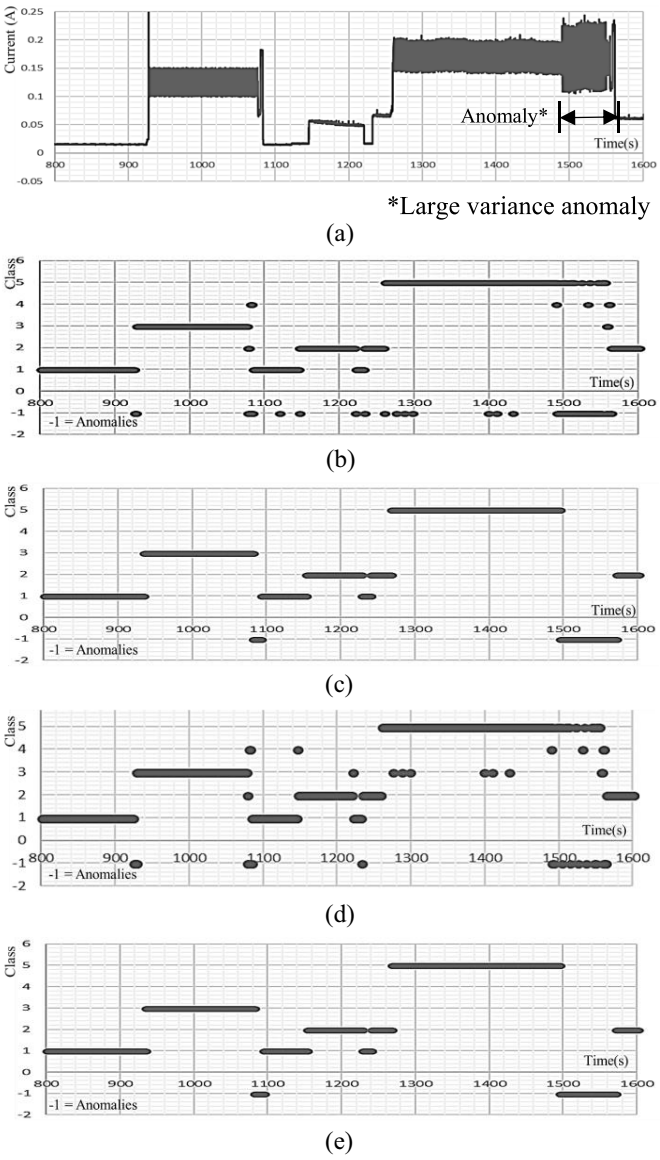

Fig. 11. Waveforms and results from the ELV dc picogrid of 3 loads. (a) Current signal sensed by the slave meter. (b) Instantaneous classification and anomaly warning by the distance from the centroid. (c) HE-*k*NN classification and anomaly warning by the distance from the centroid. (d) Instantaneous classification and anomaly warning by the distance from the nearest neighbors. (e) HE-*k*NN classification and anomaly warning by the distance from the nearest neighbors.

(19 V). Four features were extracted from the received current waveform, they are μ (mean), σ^2 (variance), $i_{\text{grad,max}}$ (largest gradient), and R (range which is the difference between the maximum and minimum value of the window). Table III shows the classification allocation, and Fig. 10 gives a clear 3-D representation of the clusters for the different classes.

Noise was added into the grid and caused the current signal to have larger variance while the mean current value remained unchanged. This anomaly is difficult to detect using a multimeter since a low range multimeter only provides the average reading of the system, which in this case, would indicate no changes. This anomaly is shown in Fig. 11(a) from around 1490 to 1550 s. Fig. 11(b) and (d) shows that both

TABLE IV
COMPARISON TABLES OF PERFORMANCE

Technique	Accuracy	F1-score
3 loads dc pico-grid (1000 data points)		
Baseline random generator	0.4865	0.2165
SPC on single parameter	0.9015	0.5800
Instantaneous detection	0.9111	0.6277
HE-*k*NN	0.9371	0.8025
5 LED lights dc pico-grid (300 data points)		
Baseline random generator	0.5033	0.2513
SPC on single parameter	0.7867	0.5294
Instantaneous detection	0.9366	0.8191
HE-*k*NN	0.9667	0.8936
Multi-mode air-conditioner dc pico-grid (5000 data points)		
Baseline random generator	0.4971	0.2016
SPC on single parameter	0.8928	0.3296
Instantaneous detection	0.8810	0.2199
HE-*k*NN	0.9717	0.8992

techniques that use instantaneous classification and anomaly without hierarchical enhancement produced several false-positive predictions, whereas, HE-*k*NN was able to identify the anomaly and trigger warnings using both the distance from the centroid and distance from the nearest neighbors techniques [see Fig. 11(c) and (e)].

D. Comparison of Results

Table IV shows the comparison table of performance using various types of anomaly warning techniques. The HE-*k*NN technique is placed in comparison with a random generation baseline, Standard Process Control that uses the sum of mean current waveform ± 3 times its standard variations as normal operation conditions and instantaneous error detection using the boundary around the cluster centroid in *k*NN as given in (15).

The results were compared using accuracy and $F1$-score. An accuracy is the most straightforward measure of performance. It is simply the ratio of the sum of true positive and true negative over the total population

$$\text{Accuracy} = \frac{\sum \text{True Positive} + \sum \text{True Negative}}{\text{Total Population}} \quad (23)$$

$F1$-score is the weighted average of Precision and Recall. It considers both false positives and false negatives (24)

$$F1 - \text{score} = \frac{2 * \sum \text{True Positive}}{2 * \sum \text{True Positive} + \sum \text{False Positive} + \sum \text{False Negative}}. \quad (24)$$

The proposed HE-*k*NN technique produced the best results for all three ELV dc picogrids. In terms of accuracy, the proposed HE-*k*NN technique produced the best result for all three dc picogrids at 0.9371, 0.9667, and 0.9717 for three loads dc picogrid, 5LED lights dc picogrid and multimode air-conditioner dc picogrid, respectively. The comparison of the $F1$-scores sets HE-*k*NN technique apart from the other techniques. HE-*k*NN achieved high $F1$-score in all three dc picogrids while other techniques performed badly especially in the multimode air-conditioner dc picogrid where HE-*k*NN achieved 0.8992 as compared to the next best $F1$-score of 0.3296 using SPC on a single parameter. This shows that the HE-*k*NN technique will also perform well in uneven class distribution such as a large number of actual negatives.

VII. CONCLUSION

This article introduced a low-cost monitoring system with classification features and anomaly detection and warning for ELV dc picogrids. The system consisted of multiple distributed inexpensive smart slave meters to a master computer. The innovative brain of the master computer is the proposed HE-*k*NN technique. The master computer received the four extracted features from the slave meter every second over the Wi-Fi network. By using the HE-*k*NN technique that considers the distance of test data to the centroid or its nearest neighbors and considering the burst error window and consecutive errors, the proposed HE-*k*NN technique produced very good results as shown through the experiments in the three ELV dc picogrids mentioned in this article. The proposed HE-*k*NN is a semisupervised anomaly detection algorithm that requires users' intervention in setting some of the parameters, while future improvement includes making the process automated. An example is to automate the process of running through a range of k value to select the appropriate k value based on the accuracy and amount of computational resource. This cost-efficient monitoring system is currently operating using the ubiquitous Wi-Fi network in buildings but it can be further developed to cater to other communication technologies or expand to a larger scale in a building or estate and be implemented as part of an IoT solution.

REFERENCES

[1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014.
[2] E. Spanò, S. D. Pascoli, and G. Iannaccone, "Internet-of-Things infrastructure as a platform for distributed measurement applications," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Pisa, Italy, 2015, pp. 1927–1932.
[3] L. Yu, T. Jiang, and Y. Zou, "Distributed online energy management for data centers and electric vehicles in smart grid," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1373–1384, Dec. 2016.
[4] G. C. Koutitas and L. Tassiulas, "Low cost disaggregation of smart meter sensor data," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1665–1673, Mar. 2016.
[5] G. Aurilio, D. Gallo, C. Landi, M. Luiso, and G. Graditi, "A low cost smart meter network for a smart utility," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Montevideo, Uruguay, 2014, pp. 380–385.
[6] Y. T. Quek, W. L. Woo, and T. Logenthiran, "A low cost master and slave distributed intelligent meter for non-intrusive load classification and anomaly warning," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Houston, TX, USA, 2018, pp. 1–6.
[7] M. Goldstein and A. Dengel, "Histogram-based outlier score (HBOS): A fast unsupervised anomaly detection algorithm," in *Proc. Annu. German Conf. Artif. Intell.*, 2012, pp. 59–63.
[8] B. Nordman and K. Christensen, "DC local power distribution: Technology, deployment, and pathways to success," *IEEE Electrific. Mag.*, vol. 4, no. 2, pp. 29–36, Jun. 2016.
[9] G. Madingou, M. Zarghami, and M. Vaziri, "Fault detection and isolation in a DC microgrid using a central processing unit," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Washington, DC, USA, 2015, pp. 1–5.
[10] D. G. Ece and O. N. Gerek, "Power quality event detection using joint 2-D-wavelet subspaces," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 4, pp. 1040–1046, Aug. 2004.
[11] H. Li, A. Monti, F. Ponci, W. Li, M. Luo, and G. D'Antona, "Voltage sensor validation for decentralized power system monitor using polynomial chaos theory," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 5, pp. 1633–1643, May 2011.
[12] Y. T. Quek, W. L. Woo, and T. Logenthiran, "Smart sensing of loads in an extra low voltage DC Picogrid using machine learning techniques," *IEEE Sensors J.*, vol. 17, no. 23, pp. 7775–7783, Dec. 2017.

[13] S. Zhang, Y. Wang, M. Liu, and Z. Bao, "Data-based line trip fault prediction in power systems using LSTM networks and SVM," *IEEE Access*, vol. 6, pp. 7675–7686, 2017.
[14] A. Bracale, P. Caramia, G. Carpinelli, A. R. D. Fazio, and P. Varilone, "A Bayesian-based approach for a short-term steady-state forecast of a smart grid," *IEEE Trans. Smart Grid*, vol. 4, no. 4, pp. 1760–1771, Dec. 2013.
[15] Y. T. Quek, W. L. Woo, and T. Logenthiran, "DC equipment identification using K-means clustering and kNN classification techniques," in *Proc. IEEE Region 10 Conf. (TENCON)*, Singapore, 2016, pp. 777–780.
[16] A. M. Alvarez, M. Yamada, A. Kimura, and T. Iwata, "Clustering-based anomaly detection in multi-view data," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manag.*, San Francisco, CA, USA, 2013, pp. 1545–1548.
[17] L. Kuang and M. Zulkernine, "An anomaly intrusion detection method using the CSI-KNN algorithm," in *Proc. ACM Symp. Appl. Comput.*, Fortaleza, Brazil, 2008, pp. 921–926.
[18] W. Li, A. Monti, and F. Ponci, "Fault detection and classification in medium voltage DC shipboard power systems with wavelets and artificial neural networks," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2651–2665, Nov. 2014.
[19] Y.-K. Chen, Y.-C. Wu, C.-C. Song, and Y.-S. Chen, "Design and implementation of energy management system with fuzzy control for DC microgrid systems," *IEEE Trans. Power Electron.*, vol. 28, no. 4, pp. 1563–1570, Apr. 2013.
[20] M. Šíra and V. N. Zachovalová, "System for calibration of nonintrusive load meters with load identification ability," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 6, pp. 1350–1354, Jun. 2015.
[21] M. Pöchacker, D. Egarter, and W. Elmenreich, "Proficiency of power values for load disaggregation," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 46–55, Jan. 2016.
[22] A. Hossain, Z. A. Choudhury, and S. Suyut, "Statistical process control of an industrial process in real time," *IEEE Trans. Ind. Appl.*, vol. 32, no. 2, pp. 243–249, Mar./Apr. 1996.
[23] C. J. Spanos, H.-F. Guo, A. Miller, and J. Levine-Parrill, "Real-time statistical process control using tool data (semiconductor manufacturing)," *IEEE Trans. Semicond. Manuf.*, vol. 5, no. 4, pp. 308–318, Nov. 1992.
[24] L. Kukacka, P. Dupuis, G. Zissis, J. Kraus, and M. Kolar, "Extra low voltage DC grid LED lighting systems: Photometric flicker analysis," in *Proc. IEEE Int. Workshop Electron. Control Meas. Signals Appl. Mechatronics (ECMSM)*, Liberec, Czech Republic, 2015, pp. 1–6.

Yang Thee Quek (M'99) received the B.Eng. (Hons.) and M.Sc. degrees from the National University of Singapore, Singapore, and the Ph.D. degree in electrical and electronic engineering from the Newcastle University, Newcastle upon Tyne, U.K.

His current research interests include smart dc grid, renewable energies, and applications of computational intelligence techniques in power systems.

Wai Lok Woo (M'11–SM'12) received the B.Eng. degree in electrical and electronics engineering, and the M.Sc. and Ph.D. degrees from Newcastle University, Newcastle upon Tyne, U.K., in 1993, 1995, and 1998, respectively.

He is currently a Professor of machine learning with Northumbria University, Newcastle upon Tyne. He has published over 400 papers on these topics on various journals and international conference proceedings. His current research interests include the mathematical theory and algorithms of machine learning and applications to deep learning, information processing, anomaly detection, digital health, and digital sustainability.

Prof. Woo was a recipient of the IEE Prize and the British Commonwealth Scholarship. He serves as an Associate Editor for several international journals, including sensors, IET *Signal Processing*, *Journal of Computers*, and *Journal of Electrical and Computer Engineering*. He is also a member of the Institution Engineering Technology.

Logenthiran Thillainathan (M'07–SM'16) received the B.Sc. degree in electrical and electronic engineering from the University of Peradeniya, Peradeniya, Sri Lanka, and the Ph.D. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2005 and 2012, respectively.

He is currently a Faculty with the School of Engineering and Technology, University of Washington, Seattle, WA, USA. His current research interests include smart grid, micro grid, renewable energy resources, and the applications of computational intelligent techniques and intelligent multiagent system for the development of smart grid and modern power systems.

Dr. Logenthiran was a recipient of several awards and scholarships, including the first prize award in Siemens Smart Grid Innovation Contest and Newcastle Teaching Award. He is a member of the Institution of Engineering and Technology and the Associate Fellow of the Higher Education Academy, U.K. He is a Chairman of the IEEE PES Singapore chapter for year 2018.