Anomaly Detection for IoT Time-Series Data: A Survey

Andrew A. Cook , Göksel Mısırlı, and Zhong Fan, *Senior Member, IEEE*

Abstract—Anomaly detection is a problem with applications for a wide variety of domains; it involves the identification of novel or unexpected observations or sequences within the data being captured. The majority of current anomaly detection methods are highly specific to the individual use case, requiring expert knowledge of the method as well as the situation to which it is being applied. The Internet of Things (IoT) as a rapidly expanding field offers many opportunities for this type of data analysis to be implemented, however, due to the nature of the IoT, this may be difficult. This review provides a background on the challenges which may be encountered when applying anomaly detection techniques to IoT data, with examples of applications for the IoT anomaly detection taken from the literature. We discuss a range of approaches that have been developed across a variety of domains, not limited to IoT due to the relative novelty of this application. Finally, we summarize the current challenges being faced in the anomaly detection domain with a view to identifying potential research opportunities for the future.

Index Terms—Anomaly detection, data analysis, Internet of Things (IoT), survey.

I. INTRODUCTION

THE Internet of Things (IoT) is a paradigm within computing related to the enablement of devices, "things," with the ability to communicate data with each other without requiring the direct involvement of human agents [1]. These devices may take the form of sensors, actuators, computers or "smart" objects that are able to observe or interact with their internal and external environments. The growth of IoT has been enabled by the development of a wide range of cost effective sensing and computing solutions able to work in environments that would have previously been unattainable. IoT is currently undergoing rapid expansion with estimates of global economic impact of up to \$11.1 trillion per year by 2025 [2] and up to 20 billion connected devices by 2020 [3].

Within the data analysis performed over IoT data, there is often a need to identify novel or unusual states within a system being monitored by the sensors deployed within the direct environment around that system. This type of analysis has applications within a variety of domains from smart traffic management, remote health care and assisted living, efficient smart energy management, and automated industrial processes. This process is often referred to as novelty detection, anomaly detection, outlier detection or event detection.

Currently, many anomaly detection methods require significant human interaction to enable these systems and subsequently extract and interpret the data generated. It is relatively easier for an expert to look a small subset of data representing the state of a system and manually identify the trends and patterns which are of interest, even if the system is small it may be difficult to identify these trends manually. However, as the number of interconnected devices increases so does the complexity of this data analysis, as such, there is interest in developing automated approaches allowing the experts to only investigate the most important events observed.

In Section II, we define the main types of anomalies which may be encountered in an IoT system. The potential uses of anomaly detection in a variety of IoT settings are discussed in Section III. In Section IV, we discuss the specific challenges which complicate the process of anomaly detection. Section V investigates the range of approaches that have been historically employed as well as those which are under current development. Finally, we will summarize the current research challenges being encountered and potential future directions within the domain in Section VI.

II. DEFINING ANOMALY

There have been a number of attempts to define the nature of anomalous data. Hawkins defines an outlier as: "an observation which deviates so significantly from other observations as to arouse suspicion that it was generated by a different mechanism" [4]. An alternate definition was offered by Barnett and Lewis where: "an outlier is an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data" [5].

A general definition of an anomaly within the context of the IoT is: the measurable consequences of an unexpected change in state of a system which is outside of its local or global norm. This definition comprises of a number of important observations about the nature of the IoT data.

1) The majority of data captured by an IoT system can be considered "normal" in that it represents the usual operating characteristics for that specific system
2) The concept of normal operation of a system can change over time for a variety of reasons.

Fig. 1. Point anomaly (circled in red) in random Gaussian noise.

Fig. 3. Example of a collective anomaly in simulated ECG time series, marked in red adapted from [6].

C. Collective or Pattern Anomalies

A collective anomaly [6] or pattern anomaly [7] is a collection of observations that are anomalous with respect to the rest of the data. Individual observations within a collective anomaly may or may not be anomalous, it is only when they appear as a group that they arouse suspicion (Fig. 3).

III. APPLICATIONS FOR IoT

The IoT approach is being increasingly applied to a variety of domains due to the inexpensive and nonintrusive nature of the devices on the market and in development.

A general application of anomaly detection is the identification of outlier observations which may affect future analytics performed on the data collected within the IoT network.

A. Industrial IoT and Industry 4.0

Anomaly detection methods have been applied to a variety of industrial processes from system health monitoring in large-scale power generation [8], intelligent maintenance scheduling in smaller production plants [9], fault detection in residential heating ventilation and air conditioning (HVAC) systems [10], and quality control techniques in manufacturing [11]. Large and high-value installations can justify the expense of human analysts or specifically tailored solutions, however as the scale and the value of the installation falls the need for more generalized and automated approaches becomes clear.

Anomaly detection is used on sensor readings from engine-based machines in [12], here they use simple machine learning approaches to model the normal behavior based upon a range of parameters with a one-minute resolution. These models are then used to identify specific failure modes when the received data falls outside of the learnt normal regions. In this example, the authors use expert knowledge to define which of the many tracked data streams are most relevant to specific fault modes. They utilize histograms to analyze the relations between these reduced variable combinations to help guide their detection algorithms. This tailored approach assists in specifying the
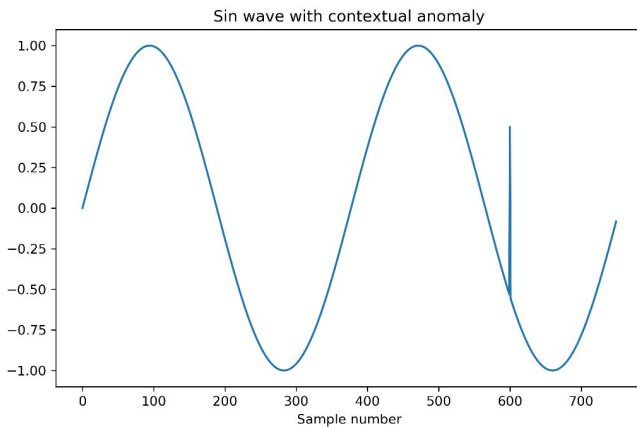

Fig. 2. Example of a contextual anomaly - the anomalous value at 600 is the same as a number of other observations however in context this observation is anomalous. Adapted from [6].

3) The data generated by an IoT deployment represents only a view of the actual processes which govern the system being monitored.

A. Point Anomaly

Point anomalies are the most similar to the definition offered by Hawkins [4]. A key characteristic of these anomaly types (Fig. 1) is the return of the time series to its previous normal state within a very short time period of only a few observations.

These point anomalies may represent statistical noise, they could be produced by faulty sensing equipment or they could represent a significant short period event that is of interest to the operators of the system.

B. Contextual Anomalies

Contextual anomalies [6] are observations or sequences which deviate from the expected patterns within the time series, however, if taken in isolation they may be within the range of values expected for that signal. When taken in the context of the surrounding observations (Fig. 2) a contextual anomaly is a deviation from the norm.
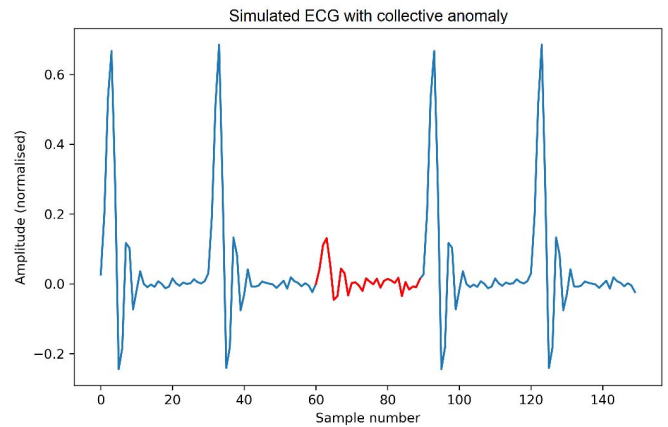
type of detected anomaly and reduces computational complexity in comparison to using all available data streams for each method.

Prediction and diagnosis of faults are performed on a 3 MW wind turbine in [13] where data is collected from the existing supervisory control and data acquisition (SCADA) system avoiding the need for expensive sensor suites to be retrofitted to the device. Operational data was collected at ten-minute intervals with labeled data available for fault states in the working turbine. Using a subset of features they trained a number of classifiers to detect specific fault modes noting difficulties with this training due to the imbalance between normal and anomalous subsequences. They found that they were able to predict faults up to an hour in advance of when the actual anomaly occurred, this may allow automated processes to occur to mitigate the effects of the developing fault thereby allowing for the maintenance to be scheduled to repair the device in question.

Surface mounted audio sensors are retrofitted to industrial machinery in [14]. Audio spectrum data is used to monitor the operation of internal components within the machine. The authors provide a framework for low-cost nonintrusive monitoring of system state which allows for faults to be detected during their early stages before failure occurs, this allows for responsive maintenance to be scheduled thereby reducing unexpected downtime for the machinery they investigated.

IoT sensors were installed in a water treatment facility [15] to aid in the management of chemical and particulate concentrations in storage tanks. The aim of the anomaly detection process would be to automatically identify when the tank entered an unsuitable state and allow for reactive measures to be triggered without human intervention.

As the price for IoT devices falls there is an increasing likelihood of older industrial equipment being brought inline with newer devices by retrofitting monitoring solutions. The general purpose anomaly detection methods may be able to provide deeper insight into the operational state of these devices and thereby improve efficiency and uptime for the processes on which they operate.

B. Smart Energy

The introduction of increased monitoring and sensing within the power network has lead to a change in the way energy is managed. Many countries are introducing "Smart Meters" across their network. These devices are able to monitor power usage at a range of time intervals and automatically report these values to the operator of the network. This provides useful information to both the customer and controller which gives the customer the opportunity to exercise this knowledge to adjust their own behaviors while decreasing the requirement for manual or estimated meter readings on the part of the power company.

One advantage of this near real-time monitoring of power usage is the ability for the energy suppliers to identify faults in the local distribution network as they happen rather than relying upon customers to inform them of outages [16], their approach uses data fusion from multiple customers to identify

faults at the individual or local levels as well as aiding in localization of those detected anomalies. If a number of units all report similar issues at a similar time it is possible to identify the location of a fault as well as potentially which type of fault has been encountered.

In [17], power-line communication signals are used to identify and localize faults in the distribution network, such as electrical faults, impaired cables, and unexpected impedance changes. They utilize a two-part algorithm, the first detects and tracks the evolution of faults over time while the second uses information about the network topology to localize the faults identified by the first algorithm.

Values recorded by micro-synchrophasor units (sensors able to detect voltage and current phase angle and magnitude at GPS accurate time steps) have been shown to be useful in the detection and localization of faults and failures in power networks using more traditional big-data analysis techniques. In [18], the challenge faced in this case is the sampling frequency from each installed unit (around 120 readings/s) which poses a significant computational challenge for any automated detection process which may be deployed.

In addition to the detection of technical losses within power networks, a number of approaches have been suggested for the detection of nontechnical losses (energy theft) using the information provided by commonly installed smart meters [19], [20]. Mashima and Cárdenas [21] approach the problem by assuming a worst case physical attack on a smart meter to show the ability for their method to detect a long-term approach to energy theft.

C. Smart City and Buildings

The IoT paradigm is being extended to management and monitoring of cities and buildings by introducing the networks of sensors to monitor events occurring within their environment. This shift allows for additional data to be collected on the environment in which the network is installed thereby enabling a data-driven analysis of the conditions present.

Within the smart city context, IoT approaches have been demonstrated to identify anomalous road conditions. A number of applications use crowd-sourced data from mobile devices to identify high-congestion locations within their route-finding applications and therefore suggest alternate routes to consumers, often this is enabled by the algorithm detecting and aggregating unusual movement behavior from the positional information reported by user devices [22], this may allow the user to avoid high-congestion areas thereby reducing the impact of their journey. Road surface health monitoring has been suggested by Tai *et al.* [23] and El-Wakeel *et al.* [24] using connected devices to enable monitoring of road conditions, thereby allowing timely maintenance to be performed as to reduce damage to private vehicles and reduce road traffic incidents. Bus trajectory data is used in [25] to map congestion within the urban areas to help guide a data-driven approach to urban management.

Airborne pollution levels within urban environments are another important issue being faced globally. A number of studies have demonstrated the use of networked sensors to

detect and monitor pollution levels in cities [26], [27]. The data collected via these methods enable urban planners to make informed decisions with Health, Traffic, and the Environment in mind.

Smaller IoT networks are increasingly being installed within commercial and residential buildings. Data collected by these networks may be used to analyze and improve energy efficiency within the location [28]–[31]. A number of these approaches introduce additional contextual information into their algorithms to account for variations in usage dependent upon the day of the week or which month is being monitored as well as the current weather conditions.

Smart home data has been suggested for activity monitoring within assisted living situations [32], [33] whereby the normal activities of the individual are learnt and significant deviations may be raised as anomalous thereby giving increased awareness to carers or health services. An example provided in [34] discusses the event that a monitored individual is found to be on the floor in a kitchen for an extended period of time, this behavior would be unusual and may be suggestive of a fall or collapse and therefore require assistance from emergency services.

IV. CHALLENGES FACED IN ANOMALY DETECTION

At a basic level anomaly detection is the identification of patterns which do not conform to the expected norm for the system [6]. There are a number of elements which make this basic interpretation very challenging.

A. Elements of IoT Data

IoT data may present similarly to the data collected from other domains however there are a number of aspects to the structure of the time series as well as the environment in which the data are being produced and analyzed which could affect the success of an anomaly detection algorithm.

1) Contextual Information: With a variety of sensors distributed around the environment of the system being monitored there is the opportunity to include contextual information into the anomaly detection process [35], the inclusion of this information offers the chance to improve the abilities of the analytic framework but similarly introduces a number of challenges which must be overcome.

Temporal Context: As the majority of IoT data is generated in the form of time-series data [36] (whether sampled at set intervals or via irregular sampling) there is some implication of temporal correlation between observations—that is the reading at time t is in some-way related to the observations at times $[t_{-1} \rightarrow t_{-n}]$.

Spatial Context: Similarly when multiple sensors are deployed monitoring the same system there is some implicit spatial context to be managed [36], this becomes more difficult to handle as the spatial context is increased in size or when the sensors themselves are made mobile via some mechanism. An example would be sensors mounted on a platform such as a train. Observations which may be normal on flat ground in a city may be anomalous when observed as the train is climbing an incline in a rural area. This may be mitigated by the

addition of other sensors to the network if this is thought to be important, for instance, an accelerometer measuring the current angle of the engine could provide vital information when monitoring the system health.

External Context: A subset of spatial context would be the external conditions around the system being monitored. For example, if an IoT system were monitoring power usage in a building with relation to internal temperature, for instance, to ensure the heating network was performing optimally, it would be important to know the weather conditions outside of the building being monitored. This additional contextual information could be gathered by external sensors mounted on the roof (however their readings may be inaccurate due to weather conditions and heat loss through the roof) or by using third-party weather information such as forecasts or local weather station data.

The introduction of contextual information can enrich the ability for an anomaly detection algorithm to correctly identify those observations or sequences which do not conform to the expected behavior, however, it does increase the complexity of the process and therefore it is important to select the correct contextual information when choosing the anomaly detection process as well as when designing the initial sensor network.

2) Dimensionality: Dimensionality describes the number of separate data attributes captured in each observation [6], the dimensionality of the data affects the choice of method used as certain approaches are unsuitable for higher dimensional data. Additionally, the computational cost of higher dimension data may be more than that of lower dimensional data.

IoT data is produced in two broad categories.

Univariate data consists of a sequence of observations taken by a single sensor. These data streams are most often in the form of a key–value pair where the key is the timestamp of the observation with the value being a scalar, nominal or ordinal reading of the environment being monitored, x^t. These may also be aggregated data from multiple sensors which has been combined into a single value during a preprocessing stage.

Multivariate data consists of a sequence of observations taken by multiple sensors. These data streams are most often in the form of a key–vector pair with a number of observations taken at the same timestamp each associated with a different sensor or actuator monitoring a single system, $x^t = [x_1^t, \ldots, x_n^t]$. These can be thought of as being a collection of temporally correlated univariate data streams that provide a more complete view of the system being monitored.

Anomaly detection over univariate streams relies upon the comparison of the current observation against the local or global history of the time-series being analyzed. This is contrasted with multivariate streams where not only is the history of the stream important to the detection task but also the relationship between each of the measurements which combine to form the observation at a given time-step.

3) Noise: Noise is inherent in real-world systems. Noisy data represents fluctuations in the reported values which is not significant to the overall structure of the data as a whole and may be caused by minor variations in the sensitivity of the detector, unrelated events occurring within the vicinity of the

sensor or transmission-based errors in the data management system.

In an IoT environment where a large number of low cost, resource-constrained sensors are deployed the data quality is often affected by significant noise, inconsistencies and missing or duplicated data. Where the sensors are powered by battery these challenges are often amplified as the available charge decreases [37], it is often possible to aggregate data from multiple similar sensors into a single observation to reduce the environmental noise.

In some cases a change in the quality, pattern, or distribution of the noise may represent a significant event within the system, therefore it is important to understand the nature and causes of the noise and as such it may not be suitable to apply traditional noise reduction techniques to the raw data before the anomaly detection stage.

4) Stationarity: A stationary time series is the one where the mean, variance, and autocorrelation does not vary with time. There are a number of ways in which a real-world time series can display nonstationarity and it is these elements which make many approaches unsuitable for IoT anomaly detection.

Concept drift is the change in statistical distribution of a data stream over time [7], [38], [39].

Seasonality refers to a special case of concept drift where cyclical changes occur over varying time scales of much higher period than the sampling resolution [40].

Change points are locally or globally permanent changes in the normal state of a system being monitored [41]. These changes are generally more abrupt than those seen in concept drift and represent the rapid adoption of a new state within a system. Change points may be expected such as when upgrading a component within a machine, or unexpected such as a sudden increase in usage of a particular stretch of road.

The ability for an anomaly detection method to adapt to changes in the structure of the data is important for longer term deployments as data points which may have represented anomalies at some point in the history of the system may now no longer be seen as anomalous given the current state of the system.

B. Prior Knowledge

When deploying an anomaly detection method into a new or poorly known system it is often impossible to provide a sufficient historic data set to be able to correctly define both the normal operating state of the system as well as any or all potential anomaly types which may be seen [42].

This lack of prior knowledge of the data as well as the relative scarcity of anomalous sequences of observations in any data which may have already been collected causes difficulty when applying traditional supervised machine learning approaches. There are a number of methods by which an imbalanced data set may be manipulated to allow for supervised learning approaches. Resampling of initial training data, i.e., the reduction of normal instances or introduction copies of known anomalies, can lead to significant under- or over-fitting of the final model. Possibly the most applicable approach for

the IoT domain would be the introduction of synthetic anomalies into the training set based upon known anomaly modes using tools such as PARANOM [43]. All of these approaches may, however, damage the temporal context of those anomalies as there may be important prior trends which are not as visible to the analyst designing the training data. Additionally, the use of a supervised classifier for anomaly detection, while useful for identification of known anomaly modes, may subsequently have difficulty with identifying novel anomaly modes upon which it has not been trained.

There are some situations where *a priori* knowledge can be transferred from similar systems, this is the case for many network intrusion and security tasks such as the detection of distributed denial-of-service attacks [44], or in Industrial IoT systems where the same machine has been deployed in multiple locations with known failure modes.

For the majority of cases, however, this corpus of historic data representing both the normal and "anomalous" instances is simply not available and as such a basic assumption must be made. The majority of observations made about the system are within the bounds of normal operation.

There are machine learning techniques which provide an opportunity to combat this lack of knowledge by utilizing unsupervised or semisupervised processes. In these approaches, the system will be trained using the normal data collected about the state of the system and therefore when data falls outside of some boundary condition it is reported as anomalous. This approach allows for the discovery of novel anomaly modes or application to new or unique environments at the cost of detailed information about the specific anomalies identified.

As the corpus of knowledge increases, more normal data is collected as well as various data relating to anomalies it may be possible to begin to include classifications of these anomalous states into the analytics pipeline.

C. Time and Resource Constraints

Within an IoT deployment the majority of devices will be of low power with limited computational resources, as such the current model is generally to collect and process the data at some centralized location, usually using cloud or datacenter computing technologies. This model allows for greater resources to be leveraged for the analytic process however this also introduces some level of latency to the system due to round-trip delays as well as resource scheduling [45].

In some cases this is allowable as it is not important to act quickly upon the knowledge gathered from the data, however, when looking toward automation of connected resources it may be a requirement for the data to be processed quickly and therefore reports be generated as soon as possible after the data are generated [46]. The use of Edge/Fog devices offers an opportunity for this to occur closer to the location the data is generated, however, these devices are inherently lower powered than cloud services and therefore it is important to understand the computational cost of any analytic task being performed on that data.

Wireless IoT devices send small bursts of data, generally using a low-powered or long-range communication protocol. This allows for devices to be deployed in remote locations still be able to communicate with central systems. The limits on the quantity of data which can be sent via these protocols is significantly limited. Repeated long range communications also incur a significant cost in terms of battery usage where the devices are self powered, as such if limited processing can be performed closer to the device and aggregated information sent at a lower frequency it may be possible to increase the lifetime of any such device [47].

Management and storage of data also becomes a concern when large numbers of sensors are deployed [46], it may be impractical to store the entire data set collected by an IoT network in a format which can easily be accessed by the anomaly detector, rather initial analysis can be performed before the data are further transformed and archived. This requires the use of techniques which do not require the presence of the entire data set.

Sliding windows offer an opportunity to reduce the storage requirements on the devices tasked with processing the data by only retaining recent observations. However, there may be features that can be missed by only performing the analysis on these data windows, therefore, the anomaly detection model would require some way of "remembering" past trends and patterns without necessarily requiring access to the entire historical data set.

Incremental processing is the extreme limit of the windowed approach, in this case, only the most recent observation is processed and as such, each data point is analyzed exactly once by the anomaly detection method. Historic trends and patterns must therefore be retained entirely within the model being used for the task.

D. Reporting Method

There are two primary ways in which anomalous data may be reported [6], [48].

1) Anomaly Score: An anomaly score is a value representing the degree to which a given observation deviates from the expected value as defined by the anomaly detection model being used. There are a variety of methods of generating anomaly scores unique to each algorithm. This method is often of use when performing later analysis of the collected data as the analyst may choose to investigate only the top-n anomalies within a given time period.

This scoring may also be useful in the identification and management of outliers when performing associated analytic tasks such as predictive analytics.

2) Labels: A binary label may be applied to each observation noting whether the detection algorithm has identified the observation as normal or "anomalous". Some algorithms may directly report this binary classification however often this is calculated using some threshold over the initial score generated by the detection algorithm. A basic method would be to assume some distribution of scores over the time series and utilize a user-defined threshold or deviation beyond which an anomaly is flagged.

This approach may see the greatest utilization where immediate reporting is required such as in the identification of failures in the system being monitored where the operator or owner of the system requires near real-time notification of anomalous states.

For more complex anomaly detection systems, in particular, those utilizing a supervised learning approach, there may be multiple anomaly classes each with their own label allowing the option of different notifications to be triggered based upon the assigned label.

V. CURRENT METHODS

As the growth of the IoT technology is relatively recent there are only a small number of approaches presented which operate in purely IoT environments, there is however a long history of anomaly detection over time series and nontemporal data from a broad variety of domains from which techniques may be investigated for IoT specific applications.

There are a number of surveys which have investigated the problem of anomaly detection in general, often with short sections discussing the detection of anomalies in time-series data. Early works include Hawkins [4] and Abraham and Chang [49]. Markou and Singh [50], [51] provided a comprehensive two-part survey investigating both statistical and neural network approaches up to 2003. Chandola *et al.* [6] provided a deep investigation of the methods available in 2009, with Zhang *et al.* [52] discussing the approaches applicable to the early IoT. More recently Chalapathy and Chawla [42] investigated the application of deep learning approaches to the broader field of anomaly detection with some space given to time-series situations, such as IoT and Industrial IoT. This review pursues a narrower scope investigating only those techniques most applicable to the types and structures of data expected within the IoT time-series domain.

There are a broad range of algorithms and approaches presented for the purpose of detecting anomalies in time-series data. While some techniques may combine elements from multiple approaches, the general methods can be divided into the following groups.

1) *Statistical and Probabilistic:* These methods utilize historical data to model the expected behavior of a system. When a new observation is received it is compared against the current model for that system and if it does not fit within that model it is registered as an anomaly [51].

2) *Pattern Matching:* This method uses direct modeling of the time series. In a supervised setting with known characteristics for expected anomalous subsequences, the detector will compare each new observation against a database of labeled anomaly events and flag those which are most similar. In the case where there is a lack of prior labeled anomalies, the detector may learn the most common historic patterns within the normal data and flag those novel subsequences which do not match the historic corpus as anomalies.

3) *Distance-Based:* A distance metric is defined such that a newly received observation can be compared against

those preceding it with the assumption that a lower distance would most likely occur from similar mechanisms and therefore would be flagged as normal. Conversely, a larger distance would indicate the observation as having been generated by a different mechanism and as such would be flagged as anomalous [6].

4) *Clustering:* This approach projects the data into a multidimensional space and utilizes the density of the resulting clusters. Those observations which present close and within dense clusters are indicated as normal observations while those which present further away from, or do not belong to, these clusters are reported as anomalous [6].

5) *Predictive:* A regression model is generated based upon the recent and longer term trends of the system predicting the expected value at some future time. When a new observation is received it is compared against these predicted values and an assessment is made of how accurate that prediction was, where the observed value and the predicted value vary greatly that observation is flagged as anomalous [15].

6) *Ensemble:* The ensemble approach uses a number of different algorithms to observe each data point and some form of the voting mechanism is employed over the outputs from each method. An ensemble can be constructed from a group of similar detectors, such as a range of predictive models, or from a collection of dissimilar detectors, such as the combination of probabilistic, clustering, and statistical detectors. Often the use of ensemble techniques can improve the overall success of a detection suite at the potential expense of increased setup complexity and computational time.

The choice of approach is strongly dependent upon a number of factors within the data being monitored as well as the environment in which the anomaly detector will be deployed.

A. Anomaly Detection on Univariate Time-Series Data

Univariate time series represent the data output from a single source linked with the time of the observation. This may be a current trading price for a stock or share, the electrical signal from a single trace in an electroencephalogram (EEG), total network traffic at a specific time step or the value produced by a single IoT sensor. The structure of the underlying system being monitored is of high importance to the accuracy of any univariate anomaly detection method.

1) Nonregressive Approaches: For a stationary time series, the simplest detection method is to manually set high and low thresholds such that when an observation is received outside of these bounds an anomaly is reported.

A more advanced method is to produce a mean and variance for the historic data and with a threshold defined based upon these measures to report anomalies that fall outside of this range [53]. Similar to this is the box-plot approach where the distribution of the data is split into a range of smaller categories and new observations are compared against these ranges [54], this may be extended with a larger number of splits which leads to a histogram approach. These

techniques are very computationally efficient requiring only a small footprint both in terms of processor time and memory requirements, however, these approaches do not work for a majority of time series as they mostly ignore the temporal aspects of the data and treat it as a simple distribution over univariate data, they are therefore unable to detect a majority of contextual and collective anomalies.

Artificial neural networks have also been applied to the problem. Autoencoder neural networks work by taking the values presented in the input layer and passing them into a number of hidden layers with fewer neurons before symmetrically expanding that network toward the output layer. The ability for a trained autoencoder to reconstruct any given input vector gives some insight into how normal that input vector is. A higher reconstruction error suggests that there is some information within the input data which is not expected given the data used to train that network. Autoencoders are placed onto the resource-constrained sensor devices in [55], each device is responsible for collecting sequential data over a period of time and detecting anomalies based upon the reconstruction error produced by its shallow autoencoder network. Training is performed in a daily batch method in a central cloud location using the reported input and output vectors generated by each sensor. This relocates the expensive training requirement away from the constrained device and into a more suitable location while also reducing power requirements caused by multiple communications per day.

Recurrent neural networks (RNNs) utilize feedback loops within the hidden layers in a neural network to allow certain neurons to be affected by outputs from previous time steps thereby providing some level of memory within the network itself. This allows the network to capture relationships between observations over a period of time. Early RNNs suffered from vanishing gradients, that is difficulty in training over large data sets, however with the development of new arrangements of gates such as long short-term memory (LSTM) and gated recurrent units (GRUs) this problem was mitigated.

An LSTM-based encoder–decoder neural network is employed in [56] on a variety of univariate time series where the reconstruction error of the autoencoder is used to identify anomalous sequences within the data, their method is a semisupervised approach in that the initial network is trained only with normal data. They provide a thresholding mechanism over their computed anomaly score to allow for tuning of the system within a supervised or human-in-the-loop setting based upon maximizing the f_β score.

2) Regression-Based Approaches: Another popular approach to identifying outliers is to apply some form of predictive modeling of the time series. The newly received observation is compared against the predicted value and an assessment is made based upon the difference between the predicted and actual values [15].

There are a variety of methods for which can be used for the predictive Portion of This Approach. Autoregressive moving average (ARMA) [57] builds a parametric model of the time series. ARMA has seen widespread usage in a number of fields however this approach has difficulty with nonstationary data sets, in particular, those which display significant

seasonality or mean shift. Autoregressive integrated moving average (ARIMA) allows for the management of nonstationarity by adding a number of differencing steps during the processing phase to move the data toward a more stationary distribution [58]–[63]. Seasonal ARMA (SARMA) approaches account for differing levels of seasonality within the data by generating multiple models across the different seasonal time lags and apply the same techniques [64].

Another approach to the predictive method is to use artificial neural networks to capture the dynamics of a time series, early multilayer perceptron (MLP) approaches showed similar predictive abilities to those demonstrated in ARMA derivative models [65] for stationary and nonseasonal time series.

ARIMA models are combined with MLPs in [29] for predictive analysis, with a simple 2σ thresholding over the error value to identify anomalous observations. They demonstrated their method using electricity consumption data gathered each minute from a university office situation. A very large window size was utilized to generate their models (4 weeks and 8 weeks) however full week ahead predictions were made based upon these data. They note that this method was very sensitive to certain occasional use situations such as when a printer was in use which automatically exceeded the 2σ threshold they had selected and therefore they introduced some additional rules into the detection engine to compensate for these activities.

With the development of RNNs [66], [67] such as LSTM and GRU the ability for the neural network approach to better model the variability present in complex univariate systems has been demonstrated in [68].

An online time-series prediction approach is presented by [69] whereby the online updating of their LSTM-based neural network is weighted by the loss value from each new data point where this loss is significant the algorithm reduces its effect on the updation of the network thereby minimizing the effect of point anomalies on the predictive capacity of the network while allowing for change points to be gracefully handled by the network. While it is not mentioned in this article there is the opportunity for a pipeline to be developed to allow for these anomalies to be reported to the system operators.

Attention-based RNNs are employed within an autoencoder in [70] to more accurately predict complex long term patterns within data.

Malhotra et al. [71] presented two approaches using stacked layers of recurrent sigmoid units (RSUs) and LSTMs to capture long term dynamics in a variety of univariate systems. Their networks predict the expected values for a number of time steps ahead and the resulting error values are used to calculate a probability score that the observation at that later time is within the expected normal range, a threshold value is computed for this probability score and those observations falling below this level are reported as anomalous. They note that for systems with long term temporal dependencies the LSTM approach significantly outperforms the RSU approach. A similar deep LSTM network has been applied to ECG signals in [72] to identify a variety of different anomalous signals, again using the multiple time-step ahead probabilistic error

measure. These approaches both use offline training with a semisupervised approach.

In [73] RNNs are used for regression and two approaches are taken to converting between the raw output and a binary label. Their first method uses a thresholding method before being passed into an accumulator which counts up each time an observation is deemed to be anomalous and counts down by a larger factor each time a normal observation is taken, thereby detecting collective anomalies due to their longer presentation period. Their second uses a probabilistic approach to calculate the anomaly likelihood in the most recent observations.

Online time-series anomaly detection using deep RNNs is performed in [38] alongside local normalization of the incoming data and incremental retraining of the neural network to allow the network to adapt to the concept drift across a variety of data sets showing the applicability of the approach to a variety of domains. Their approach uses the predictive error of the network over a number of time steps to quantify the presence of anomalous observations in a scoring style manner.

While RNNs have shown promise for the prediction of time series the detection and reporting of anomalous observations based upon these predictions is still somewhat of a challenge. Xie et al. [74] presented a method of analyzing the prediction errors using a Gaussian Naive–Bayes model to process output of an RNN-based model.

The Greenhouse method [75] computes a vector for each observation using a multistep ahead predictive RNN. Their approach uses a three-phase training method, the initial phase fits the RNN to normal data in a typical semisupervised approach, the second phase fits the error vectors generated to distribution and the final phase calculates Mahalanobis-distances between these error vectors to produce a scoring method to identify outliers according to a user supplied threshold. When presented with a new time series the algorithm can therefore label each new observation as normal or anomalous based upon the post-processed error vector. This approach is currently an offline method and therefore is susceptible to changes in the distributions of the input data.

The RNN model presented by Bontemps et al. [76] focused on detecting collective anomalies by defining a minimum period for a collective anomaly and calculating error measurements over time, where the average error is above a given threshold for a period of time an anomaly is identified.

Bayesian Neural Networks are investigated in [77], by using an LSTM-based auto-encoder to perform prediction for a number of steps ahead, followed by an MLP to perform the final prediction steps, this construction provides not only a prediction for later values but a level of certainty in that prediction, when a new observation is made which falls outside of a defined predictive interval it would be flagged as anomalous.

A recent development within the artificial neural network domain is a process described as hierarchical temporal memory (HTM) [78], this process is a bio-inspired model for processing time series based upon the behaviors of the neocortex. This method is applied to sequential streamed univariate data in [79] and [80] and compared against a range of predictive models for time-series modeling. The technique is

further applied to the anomaly detection problem in [81]–[84] of note is the noise resistance of the approach as well as the ability for continual online learning allowing for the method to adjust to changes in data distribution over time without extensive offline retraining.

Simple online regression anomaly detector (SORAD) is presented in [84] which uses an initial unsupervised off-line training phase to learn the key features of the presented time series and then employs an online learning method to recompute mean and variance values at later time steps with stronger weighting to newer values. The algorithm has a built in approach to converting from predictive error to anomaly labels whereby an online method of calculating the series mean and variance is used and thresholding subsequently applied. The authors intended for their method to act as a baseline for comparing benchmarking methods however they identified that it performed well compared to the contemporary methods they compared it against, including an offline variant of the same algorithm. They do however note the importance of online adaptation in anomaly detection methods.

B. Anomaly Detection on Multivariate Time-Series Data

It is rare for a single sensor to be able to completely capture the complex nature of environments such as those monitored by IoT networks, therefore it is important to investigate approaches which combine information from multiple sources. This may allow the anomaly detection method to build a more accurate model of the hidden processes behind the data it receives by utilizing the additional contextual information granted by multiple monitoring approaches [35]. Additionally, the combination of multiple spatially related data streams measuring the same environmental variable provides additional noise tolerance characteristics over the single-sensor usecase.

Similarly to the univariate case, the choice of model for a multivariate system is highly dependent upon the nature of the data being produced and the functionality of the system being modeled, this is exacerbated by the variability on relations between each of the measured time-series as well as their temporal characteristics.

1) Dimensionality Reduction: When multiple sensors are monitoring a single system there is often a relationship between the values generated by each sensor, the interrelation between these values can be used to provide insights into the current state of the system. Dimensionality reduction seeks to identify and abstract the key relationships between these attributes. By modeling the normal operation of the system it is possible to identify irregularities in the input data by the effect it has on these reduced representations, this lowers the quantity of variables that must be handled by an anomaly detection algorithm and can provide insights into the hidden states within the data.

A common method for approaching multivariate systems where there is unknown but likely co-dependence between variables is to employ principal component analysis (PCA). This approach decomposes a multivariate system into a reduced set of independent variables, thereby reducing the overall size of the system to be investigated. PCA was applied to network traffic anomalies in [85] here the authors note that the method works as intended however it faces a wide number of limitations inherent with their PCA method such as large temporal window sizes causing difficulty with pinpointing the origin of the anomaly, difficulty with tuning the PCA model to a given data distribution, as well as the opportunity for a sufficiently abnormal anomaly to contaminate the "idea" of normality within the PCA approach. A recursive PCA with the clustering-based detection method was applied to an IoT sensor environment in [86] which showed the ability to correctly identify anomalous sequences, again however they note the computational complexity of PCA as a limitation for edge sensor implementations.

Projection pursuit provides another method of reducing the dimensionality of a multivariate system. In [87] the approach is applied to outlier detection by reducing the dimensionality of a complex system to one in which univariate methods may be applied. Similarly to PCA projection pursuit incurs a significant expense in the form of computational time.

Due to the reduced length of the hidden layers with an autoencoder these methods can be utilized in a similar manner to PCA. Reconstruction error is used to detect anomalies using a range of autoencoder methods over satellite telemetry and artificial data in [88]. The authors note that the hidden representation of the input data deviates significantly in anomalous observations when compared to the normal training data. Computational cost is compared against PCA methods and found to be significantly lower in the autoencoder approaches.

An ensemble of autoencoders and convolutional autoencoders are employed on building energy data in [89] to highlight anomalies as well as inefficiencies in control strategies with anomalies scored based upon reconstruction errors and the addition of date-time and other contextual data is shown to improve functionality. This method is suited to offline analysis as the authors sort observations by anomaly score and select the top-N for further investigation.

Surface mounted audio sensors are combined with convolutional autoencoders in [14] to detect faults in the industrial machinery again using a semisupervised approach. They again use the accuracy of the reconstruction of the input data to provide a measure of the normality of the input data. A threshold value α is learnt above which the observation window is highlighted as anomalous. A second approach is demonstrated where boundaries are learnt around the hidden representation in the central layer of the auto-encoder. As this is trained based on normal operation when a value presents outside of these boundaries an anomaly is identified.

Convolutional variational autoencoders (CNN-VAE) are utilized in an IoT inspired environment in [9], here the authors demonstrate a method of reducing the size, complexity, and training cost of the autoencoder without damaging its ability to identify anomalous instances. This makes the squeezed CNN-VAE (SCVAE) more suitable for deployment in edge devices within an IoT network.

Variational autoencoders are again employed in [90] where they are combined with GRUs to learn temporal and relational characteristics of multivariate time-series, by applying

a threshold to the reconstruction probability reported by the VAE phase anomalies can be detected within the system. Kieu *et al.* [91] presented LSTM-autoencoder and convolutional autoencoder approach which uses data enrichment during the preprocessing phase, this allows the autoencoder a larger feature space from which to identify the most representative features.

An unsupervised generative adversarial network is presented in [92] where LSTM is used to capture the temporal nature of the system in both the generator and discriminator portions of the network. This is used to train the discriminator the characteristics of the normal input data and therefore it can directly report anomalous observations when they are encountered. The presented use case is a cyber–physical system with both sensor and actuator data present. PCA is employed to reduce the dimensionality of the input multivariate data.

2) Clustering: Lui *et al.* [93], [94] presented an ensemble method utilizing isolation trees (termed isolation forests) to perform the anomaly detection, this is compared to a number of other contemporary methods, including ORCA [95], one-class SVM [96], LOF [97], and random forests [98]. They note that due to the computational efficiency of their approach it could be applied to streaming data.

Multiple kernel anomaly detection (MKAD) [99] is applied to aerospace data. MKAD uses kernel functions to learn similarity measures between variables within the data stream with a one-class SVM being applied to perform the classification task. They compared their results against ORCA [95] and sequenceminer [100], noting significant gains in detection ability across both discrete and continuous streams.

3) Other Methods: A range of methods has been presented using RNNs to capture the temporal nature of multivariate systems [101]–[103]. LSTM and GRU-based neural networks are applied to aircraft flight data in [104] where their approach is compared to the results produced by the MKAD method achieving greater success at identifying a range of anomaly types within those data.

LSTM-based detectors using offline training methods are applied to space-craft telemetry in [105] where the authors describe near real-time performance over 700 telemetry channels. They utilize the prediction errors from prior batches as well as including knowledge from domain experts to calculate threshold values for detection in later batches thereby adapting for changes in the data they are receiving overtime and to account for rare or occasional expected processes.

CNNs are combined with a trainable wavelet transform layer in [106] for the detection of change points in synthetic multivariate data. This method is able to identify gradual concept drifts and changes in the distribution of the input data over time and may provide a method of detecting anomalies before their main presentation.

A supervised approach to anomaly detection is presented in [44], here the authors utilize the class labels from the training data to provide additional information for the system operator during the decision making process. This method displays a potential direction toward which a semisupervised or unsupervised anomaly detection method could be taken as operator knowledge is combined with the raw anomaly data.

Dynamic Bayesian networks (DBNs) were applied to both univariate and multivariate environmental data in [107] where they present a number of methods attempting to perform real-time anomaly detection over sensor data. They note that their DBN approach improved in detection ability when multiple related sensor streams were combined to add context to the detection process.

A graph-based method is employed in [108] to learn dependencies between variables. Nodes within the graph represent individual observations or sequences of observations, where the weighting between nodes is low (representing a low dependency on other nodes) that node is flagged as anomalous.

A variation on self organising maps (SOMs) are utilized in [109] where the authors demonstrate the ability for their method to capture the seasonal temporal characteristics of multivariate data in an unsupervised manner. The authors also show utility for their method when applied to the univariate systems.

VI. Research Challenges and Future Directions

While anomaly detection has existed in literature and practice for a long time [110], there are still a number of problems that must be overcome in order to allow broad implementation. Currently, there is no single best approach to the problem, rather a number of approaches which may be more applicable to certain domains. Below we present a summary of the major challenges within the field which need to be investigated to allow for increased utility.

1) *Real-Time Processing:* As discussed in Section IV-C, for a majority of use cases where data is being used to aid in short term decision making or automated decision making (such as IIoT, smart traffic or smart energy) the ability for an anomaly detection method to operate in real time or near real time is important. If a detector takes longer to process an observation or set of observations than the time between measurements eventually the computational resources supplied to the detector will be exceeded and the system will fail.

2) *Window or Incremental Approaches:* Due to the volume of data being produced it would be costly to hold the entire data set available for analysis especially when the analysis is performed on resource-constrained devices. Therefore, as described in Section IV-C a sliding window or incremental approach will reduce the memory and storage requirements for the processing platform.

3) *Online Adaptive Learning:* The nonstationarity of IoT time series described in Section IV-A4 leads to a need for adaptive approaches to anomaly detection. Therefore, while offline methods may be of use for the initial deployment there should be some method for the candidate system to improve its model over time to adapt to foreseen and unforeseen changes in the data distributions without requiring extensive retraining of the system.

4) *Semisupervised or Unsupervised:* In real-world use cases there will often be a severe lack of available labeled anomaly data and it can be assumed that those data which are available will not fully represent the range

of anomalies that could occur which we describe in Section IV-B. Similarly due to the imbalance between the normal data and anomalous data classical multiclass machine learning approaches would be insufficient to capture the nature of the data stream. Therefore, an approach that trains the candidate algorithm only on normal data with anomalies being reported when they fall outside of some region around this normal data would be the most viable [9].

5) *Multivariate Data:* As shown by a number of methods discussed in Section V-A the addition of contextual information can improve the suitability of a given detector this may be in the form of temporal information, environmental information or additional sensor streams. As such, there is the need for anomaly detectors to operate successfully within a multivariate setting.

6) *Generalized Approach:* While it is likely that no single approach will be the best for every possible scenario, the development of algorithms that can be applied to multiple domains will aid in the reusability of techniques and the ease of deployment of anomaly detection methods to a variety of tasks.

VII. Conclusion

The IoT approach offers significant opportunity for the application of a number of analytical techniques in order to extract useful knowledge from the large volume of data being collected. In most cases, manual analysis of these data streams is impractical or financially infeasible and therefore automated methods must be developed to convert from the raw data being collected into actionable information.

In this survey, we have discussed the definition of what an anomaly is within the domain of time series and IoT data. We have described the use of anomaly detection as a data analysis tool within a number of IoT specific use cases outlining the aims and results for those approaches as well as some of the benefits which may be derived from these applications.

We describe the major challenges (Section IV) faced while developing an anomaly detection solution given the dynamic and novel systems being monitored by IoT deployments and discuss some methods which may be used to mitigate these challenges. While there has been a historic focus on detection of anomalous observations in univariate data (Section V-A) the complexity of the systems being monitored by typical IoT deployments will generally require the processing of multiple data streams and a multivariate approach (Section V-B) to detecting changes in the relationships between those variables. This direction does however bring additional challenges due to the increased complexity and computation required to manage the larger number of dimensions within the data. Machine learning offers some solutions to the problems encountered however the low availability of prelabeled data continues to offer challenges to these methods. We finish by suggesting a range of research challenges which may be faced when developing a novel anomaly detection systems for both case specific and more general approaches.

As more IoT applications are developed and deployed across the growing sectors of smart cities, the energy sector and a variety of vertical industries we would expect for anomaly detection play an increasingly important role in the processing and analysis of the data being collected.

References

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.

[2] J. Manyika *et al.*, "The Internet of Things: Mapping the value beyond the hype," McKinsey Glob. Inst., New York, NY, USA, Rep., Jun. 2015. [Online]. Available: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/the-internet-of-things-the-value-of-digitizing-the-physical-world

[3] M. Hung, "Leading the IoT," Gartner Res., Stamford, CT, USA, Rep., 2017. [Online]. Available: https://www.gartner.com/imagesrv/books/iot/iotEbook_digital.pdf

[4] D. M. Hawkins, *Identification of Outliers*, vol. 11. London, U.K.: Chapman & Hall, 1980.

[5] V. Barnett and T. Lewis, *Outliers in Statistical Data*. Chichester, U.K.: Wiley, 1974.

[6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surveys*, vol. 41, no. 3, pp. 1–58, Jul. 2009. [Online]. Available: http://doi.acm.org/10.1145/1541880.1541882

[7] D. Choudhary, A. Kejariwal, and F. Orsini, "On the runtime-efficacy trade-off of anomaly detection techniques for real-time streaming data," 2017. [Online]. Available: arXiv:1710.04735 [stat.ML].

[8] A. Ajami and M. Daneshvar, "Data driven approach for fault detection and diagnosis of turbine in thermal power plant using independent component analysis (ICA)," *Int. J. Elect. Power Energy Syst.*, vol. 43, no. 1, pp. 728–735, 2012.

[9] D. Kim *et al.*, "Squeezed convolutional variational auto encoder for unsupervised anomaly detection in edge device industrial Internet of Things," in *Proc. Int. Conf. Inf. Comput. Technol. (ICICT)*, Dec. 2018, pp. 67–71.

[10] A. Beghi, R. Brignoli, L. Cecchinato, G. Menegazzo, M. Rampazzo, and F. Simmini, "Data-driven fault detection and diagnosis for HVAC water chillers," *Control Eng. Pract.*, vol. 53, pp. 79–91, Aug. 2016.

[11] A. Kanawaday and A. Sane, "Machine learning for predictive maintenance of industrial machines using IoT sensor data," in *Proc. 8th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, 2017, pp. 87–90.

[12] G. Shah and A. Tiwari, "Anomaly detection in IIoT: A case study using machine learning," in *Proc. ACM India Joint Int. Conf. Data Sci. Manag. Data*, 2018, pp. 295–300.

[13] K. Leahy, R. L. Hu, I. C. Konstantakopoulos, C. J. Spanos, and A. M. Agogino, "Diagnosing wind turbine faults using machine learning techniques applied to operational data," in *Proc. IEEE Int. Conf. Prognostics Health Manag. (ICPHM)*, 2016, pp. 1–8.

[14] D. Y. Oh and I. D. Yun, "Residual error based anomaly detection using auto-encoder in SMD machine sound," *Sensors*, vol. 18, no. 5, p. 1308, Apr. 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5982511/

[15] F. Giannoni, M. Mancini, and F. Marinelli, "Anomaly detection mod-1154els for IoT time series data," 2018. [Online]. Available: arXiv:1812.00890 [eess.SP].

[16] R. Moghaddass and J. Wang, "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5820–5830, Nov. 2018.

[17] F. Passerini and A. M. Tonello, "Smart grid monitoring using power line modems: Anomaly detection and localization," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6178–6186, Nov. 2019.

[18] M. Farajollahi, A. Shahsavari, and H. Mohsenian-Rad, "Location identification of distribution network events using synchrophasor data," in *Proc. IEEE North Amer. Power Symp. (NAPS)*, 2017, pp. 1–6.

[19] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, and X. S. Shen, "Energy-theft detection issues for advanced metering infrastructure in smart grid," *Tsinghua Sci. Technol.*, vol. 19, no. 2, pp. 105–120, Apr. 2014.

[20] S.-C. Yip, W.-N. Tan, C. Tan, M.-T. Gan, and K. Wong, "An anomaly detection framework for identifying energy theft and defective meters in smart grids," *Int. J. Elect. Power Energy Syst.*, vol. 101, pp. 189–203, Oct. 2018.

[21] D. Mashima and A. A. Cárdenas, "Evaluating electricity theft detectors in smart grid networks," in *Proc. Int. Workshop Recent Adv. Intrusion Detect.*. 2012, pp. 210–229.

[22] E. D'Andrea and F. Marcelloni, "Detection of traffic congestion and incidents from GPS trace analysis," *Expert Syst. Appl.*, vol. 73, pp. 43–56, May 2017.

[23] Y.-C. Tai, C.-W. Chan, and J. Y.-J. Hsu, "Automatic road anomaly detection using smart mobile device," in *Proc. Conf. Technol. Appl. Artif. Intell*, Hsinchu, Taiwan, 2010. [Online]. Available: https://scholar.google.com/scholar?cites=8949933769078213843

[24] A. S. El-Wakeel, J. Li, M. T. Rahman, A. Noureldin, and H. S. Hassanein, "Monitoring road surface anomalies towards dynamic road mapping for future smart cities," in *Proc. IEEE Glob. Conf. Signal Inf. Process. (GlobalSIP)*, 2017, pp. 828–832.

[25] X. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2018.

[26] R. Jain and H. Shah, "An anomaly detection in smart cities modeled as wireless sensor network," in *Proc. IEEE Int. Conf. Signal Inf. Process. (IConSIP)*, 2016, pp. 1–5.

[27] L.-J. Chen, Y.-H. Ho, H.-H. Hsieh, S.-T. Huang, H.-C. Lee, and S. Mahajan, "ADF: An anomaly detection framework for large-scale PM2.5 sensing systems," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 559–570, Apr. 2018.

[28] D. Wijayasekara, O. Linda, M. Manic, and C. Rieger, "Mining building energy management system data using fuzzy anomaly detection and linguistic descriptions," *IEEE Trans. Ind. Informat.*, vol. 10, no. 3, pp. 1829–1840, Aug. 2014.

[29] J.-S. Chou and A. S. Telaga, "Real-time detection of anomalous power consumption," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 400–411, May 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032114001142

[30] D. B. Araya, K. Grolinger, H. F. ElYamany, M. A. M. Capretz, and G. T. Bitsuamlak, "Collective contextual anomaly detection framework for smart buildings," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, 2016, pp. 511–518.

[31] M. Peña, F. Biscarri, J. I. Guerrero, I. Monedero, and C. León, "Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach," *Expert Syst. Appl.*, vol. 56, pp. 242–255, Sep. 2016.

[32] V. Jakkula and D. J. Cook, "Anomaly detection using temporal data mining in a smart home environment," *Methods Inf. Med.*, vol. 47, no. 1, pp. 70–75, 2008.

[33] U. Bakar, H. Ghayvat, S. F. Hasanm, and S. C. Mukhopadhyay, "Activity and anomaly detection in smart home: A survey," in *Next Generation Sensors and Systems*. Cham, Switzerland: Springer, 2016, pp. 191–220.

[34] C. Zhu, W. Sheng, and M. Liu, "Wearable sensor-based behavioral anomaly detection in smart assisted living systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1225–1234, Oct. 2015.

[35] M. A. Hayes and M. A. M. Capretz, "Contextual anomaly detection in big sensor data," in *Proc. IEEE Int. Congr. Big Data*, Jun. 2014, pp. 64–71.

[36] O. B. Sezer, E. Dogdu, and A. M. Ozbayoglu, "Context-aware computing, learning, and big data in Internet of Things: A survey," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 1–27, Feb. 2018.

[37] Y. Zhang, N. Meratnia, and P. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 2, pp. 159–170, 2nd Quart., 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5451757/

[38] S. Saurav *et al.*, "Online anomaly detection with concept drift adaptation using recurrent neural networks," in *Proc. ACM India Joint Int. Conf. Data Sci. Manag. Data (CoDS-COMAD)*, 2018, pp. 78–87. [Online]. Available: http://doi.acm.org/10.1145/3152494.3152501

[39] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Comput. Surveys*, vol. 46, no. 4, pp. 1–37, Mar. 2014. [Online]. Available: http://doi.acm.org/10.1145/2523813

[40] P. H. Franses, "Seasonality, nonstationarity and the forecasting of monthly time series," *Int. J. Forecast.*, vol. 7, no. 2, pp. 199–208, Jan. 1991. [Online]. Available: https://repub.eur.nl/pub/2067/

[41] M. Basseville *et al.*, *Detection of Abrupt Changes: Theory and Application*, vol. 104. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[42] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019. [Online]. Available: arXiv:1901.03407 [cs, stat].

[43] J. Gottschlich, "Paranom: A parallel anomaly dataset generator," 2018. [Online]. Available: arXiv: 1801.03164 [cs.LG].

[44] K. Amarasinghe, K. Kenney, and M. Manic, "Toward explainable deep neural network based anomaly detection," in *Proc. 11th Int. Conf. Human Syst. Interact. (HSI)*, Jul. 2018, pp. 311–317.

[45] S. K. Bose, B. Kar, M. Roy, P. K. Gopalakrishnan, and A. Basu, "ADEPOS: Anomaly detection based power saving for predictive maintenance using edge computing," in *Proc. ACM 24th Asia South Pac. Design Autom. Conf.*, 2019, pp. 597–602.

[46] M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke, "Middleware for Internet of Things: A survey," *IEEE Internet Things J.*, vol. 3, no. 1, pp. 70–95, Feb. 2016.

[47] M. S. Mahmoud and A. A. Mohamad, "A study of efficient power consumption wireless communication techniques/modules for Internet of Things (IoT) applications," *Adv. Internet Things*, vol. 6, no. 2, pp. 19–29, 2016.

[48] M. Goldstein and S. Uchida, "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data," *PLoS ONE*, vol. 11, no. 4, 2016, Art. no. e0152173.

[49] B. Abraham and A. Chuang, "Outlier detection and time series modeling," *Technometrics*, vol. 31, no. 2, pp. 241–248, May 1989. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00401706.1989.10488517

[50] M. Markou and S. Singh, "Novelty detection: A review—Part 2: Neural network based approaches," *Signal Process.*, vol. 83, no. 12, pp. 2499–2521, Dec. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168403002032

[51] M. Markou and S. Singh, "Novelty detection: A review—Part 1: Statistical approaches," *Signal Process.*, vol. 83, no. 12, pp. 2481–2497, Dec. 2003. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0165168403002020

[52] Y. Zhang, N. Meratnia, and P. J. Havinga, "Outlier detection techniques for wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 2, pp. 159–170, 2nd Quart., 2010.

[53] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[54] J. Laurikkala, M. Juhola, E. Kentala, N. Lavrac, S. Miksch, and B. Kavsek, "Informal identification of outliers in medical data," in *Proc. 5th Int. Workshop Intell. Data Anal. Med. Pharmacol.*, vol. 1, 2000, pp. 20–24.

[55] T. Luo and S. G. Nagarajan, "Distributed anomaly detection using autoencoder neural networks in WSN for IoT," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[56] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder-decoder for multi-sensor anomaly detection," 2016. [Online]. Available: arXiv:1607.00148 [cs.AI].

[57] H. Wold, *A Study in the Analysis of Stationary Time Series*. Stockholm, Sweden: Almqvist & Wiksell, 1938.

[58] D. Ventura *et al.*, "ARIIMA: A real IoT implementation of a machine-learning architecture for reducing energy consumption," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.*, 2014, pp. 444–451.

[59] B. Zhu and S. Sastry, "Revisit dynamic ARIMA based anomaly detection," in *Proc. IEEE 3rd Int. Conf. Privacy Security Risk Trust IEEE 3rd Int. Conf. Soc. Comput.*, Oct. 2011, pp. 1263–1268.

[60] H. Z. Moayedi and M. A. Masnadi-Shirazi, "ARIMA model for network traffic prediction and anomaly detection," in *Proc. Int. Symp. Inf. Technol.*, vol. 4, Aug. 2008, pp. 1–6.

[61] A. H. Yaacob, I. K. T. Tan, S. F. Chien, and H. K. Tan, "ARIMA based network anomaly detection," in *Proc. 2nd Int. Conf. Commun. Softw. Netw.*, Feb. 2010, pp. 205–209.

[62] F. Knorn and D. J. Leith, "Adaptive Kalman filtering for anomaly detection in software appliances," in *Proc. IEEE INFOCOM Workshops*, Apr. 2008, pp. 1–6.

[63] A. M. Bianco, M. G. Ben, E. J. Martínez, and V. J. Yohai, "Outlier detection in regression models with ARIMA errors using robust estimates," *J. Forecast.*, vol. 20, no. 8, pp. 565–579, 2001. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/for.768

[64] F. Kadri, F. Harrou, S. Chaabane, Y. Sun, and C. Tahon, "Seasonal ARMA-based SPC charts for anomaly detection: Application to emergency department systems," *Neurocomputing*, vol. 173, pp. 2102–2114, Jan. 2016.

[65] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *Eur. J. Oper. Res.*, vol. 160, no. 2, pp. 501–514, 2005.

[66] S. L. Ho, M. Xie, and T. N. Goh, "A comparative study of neural network and Box–Jenkins ARIMA modeling in time series prediction," *Comput. Ind. Eng.*, vol. 42, nos. 2–4, pp. 371–375, 2002.

[67] M. Ghiassi, H. Saidane, and D. K. Zimbra, "A dynamic artificial neural network model for forecasting time series events," *Int. J. Forecast.*, vol. 21, no. 2, pp. 341–362, 2005.

[68] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in *Proc. IEEE Youth Acad. Annu. Conf. Chin. Assoc. Autom. (YAC)*, 2016, pp. 324–328.

[69] T. Guo, Z. Xu, X. Yao, H. Chen, K. Aberer, and K. Funaya, "Robust online time series prediction with recurrent neural networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Montreal, QC, Canada, Oct. 2016, pp. 816–825. [Online]. Available: http://ieeexplore.ieee.org/document/7796970/

[70] Y. Qin, D. Song, H. Cheng, W. Cheng, G. Jiang, and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2017, pp. 2627–2633. [Online]. Available: http://arxiv.org/abs/1704.02971

[71] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proc. Comput. Intell.*, 2015, p. 7.

[72] S. Chauhan and L. Vig, "Anomaly detection in ECG time signals via deep long short-term memory networks," in *Proc. IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA)*, Oct. 2015, pp. 1–7.

[73] D. T. Shipmon, J. M. Gurevitch, P. M. Piselli, and S. T. Edwards, "Time series anomaly detection; detection of anomalous drops with limited features and sparse examples in noisy highly periodic data," 2017. [Online]. Available: arXiv:1708.03665 [cs, stat].

[74] X. Xie, D. Wu, S. Liu, and R. Li, "IoT data analytics using deep learning," 2017. [Online]. Available: arXiv:1708.03854 [cs.NI].

[75] T. J. Lee, J. Gottschlich, N. Tatbul, E. Metcalf, and S. Zdonik, "GreenHouse: A zero-positive machine learning system for time-series anomaly detection," 2018. [Online]. Available: arXiv:1801.03168 [cs.AI].

[76] L. Bontemps, V. L. Cao, J. McDermott, and N.-A. Le-Khac, "Collective anomaly detection based on long short term memory recurrent neural network," in *Proc. Int. Conf. Future Data Security Eng.*, 2016, pp. 141–152.

[77] L. Zhu and N. Laptev, "Deep and confident prediction for time series at Uber," in *Proc. IEEE Int. Conf. Data Min. Workshops (ICDMW)*, Nov. 2017, pp. 103–110.

[78] Y. Cui, S. Ahmad, and J. Hawkins, "Continuous online sequence learning with an unsupervised neural network model," *Neural Comput.*, vol. 28, no. 11, pp. 2474–2504, 2016. [Online]. Available: http://arxiv.org/abs/1512.05463

[79] Y. Cui, C. Surpur, S. Ahmad, and J. Hawkins, "A comparative study of HTM and other neural network models for online sequence learning with streaming data," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 1530–1538.

[80] E. N. Osegi, "Using the hierarchical temporal memory spatial Pooler for short-term forecasting of electrical load time series," *Appl. Comput. Informat.*, to be published. [Online]. Available: https://doi.org/10.1016/j.aci.2018.09.002

[81] S. Ahmad and S. Purdy, "Real-time anomaly detection for stream-ing analytics," 2016. [Online]. Available: arXiv:1607.02480 [cs.AI].

[82] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, Nov. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231217309864

[83] J. Wu, W. Zeng, and F. Yan, "Hierarchical temporal memory method for time-series-based anomaly detection," *Neurocomputing*, vol. 273, pp. 535–546, Jan. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231217313887

[84] M. Thill, W. Konen, and T. Bäck, "Online anomaly detection on the webscope S5 dataset: A comparative study," in *Proc. Evol. Adapt. Intell. Syst. (EAIS)*, May 2017, pp. 1–8.

[85] H. Ringberg, A. Soule, J. Rexford, and C. Diot, "Sensitivity of PCA for traffic anomaly detection," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 35, no. 1, pp. 109–120, 2007.

[86] T. Yu, X. Wang, and A. Shami, "Recursive principal component analysis-based data outlier detection and sensor data aggregation in IoT systems," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2207–2216, Dec. 2017.

[87] D. Galeano and R. S. Tsay, "Outlier detection in multivariate time series by projection pursuit AU—Galeano, Pedro," *J. Amer. Stat. Assoc.*, vol. 101, no. 474, pp. 654–669, Jun. 2006. [Online]. Available: https://amstat.tandfonline.com/doi/abs/10.1198/016214505000001131

[88] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proc. MLSDA 2nd Workshop Mach. Learn. Sensory Data Anal. (MLSDA)*, 2014, pp. 4–11. [Online]. Available: http://doi.acm.org/10.1145/2689746.2689747

[89] C. Fan, F. Xiao, Y. Zhao, and J. Wang, "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data," *Appl. Energy*, vol. 211, pp. 1123–1135, Feb. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0306261917317166

[90] Y. Guo, W. Liao, Q. Wang, L. Yu, T. Ji, and P. Li, "Multidimensional time series anomaly detection: A GRU-based Gaussian mixture variational autoencoder approach," in *Proc. Asian Conf. Mach. Learn.*, 2018, pp. 97–112.

[91] T. Kieu, B. Yang, and C. S. Jensen, "Outlier detection for multidimensional time series using deep neural networks," in *Proc. 19th IEEE Int. Conf. Mobile Data Manag. (MDM)*, Jun. 2018, pp. 125–134.

[92] D. Li, D. Chen, J. Goh, and S.-K. Ng, "Anomaly detection with generative adversarial networks for multivariate time series," 2018. [Online]. Available: arXiv:1809.04758 [cs, stat].

[93] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. IEEE 8th Int. Conf. Data Min.*, 2008, pp. 413–422.

[94] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Disc. Data*, vol. 6, no. 1, pp. 1–39, Mar. 2012. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2133360.2133363

[95] S. D. Bay and M. Schwabacher, "Mining distance-based outliers in near linear time with randomization and a simple pruning rule," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2003, pp. 29–38.

[96] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.

[97] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Rec.*, vol. 29, 2000, pp. 93–104.

[98] T. Shi and S. Horvath, "Unsupervised learning with random forest predictors," *J. Comput. Graph. Stat.*, vol. 15, no. 1, pp. 118–138, 2006.

[99] S. Das, B. L. Matthews, A. N. Srivastava, and N. C. Oza, "Multiple kernel learning for heterogeneous anomaly detection: Algorithm and aviation safety case study," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2010, pp. 47–56.

[100] S. Budalakoti, S. Budalakoti, A. N. Srivastava, M. E. Otey, and M. E. Otey, "Anomaly detection and diagnosis algorithms for discrete symbol sequences with applications to airline safety," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 39, no. 1, p. 101, Jan. 2009.

[101] W. Zhang *et al.*, "LSTM-based analysis of industrial IoT equipment," *IEEE Access*, vol. 6, pp. 23551–23560, 2018.

[102] R. Maia. (2018). *Multivariate temporal data analysis for vessels behavior anomaly detection*. [Online]. Available: conferences.inf.ed.ac.uk/EuroDW2018/papers/eurodw18-Maia.pdf

[103] M. Yadav, P. Malhotra, L. Vig, K. Sriram, and G. Shroff, "ODE-augmented training improves anomaly detection in sensor data from machines," 2016. [Online]. Available: arXiv:1605.01534 [cs.AI].

[104] A. Nanduri and L. Sherry, "Anomaly detection in aircraft data using recurrent neural networks (RNN)," in *Proc. IEEE Integr. Commun. Navig. Surveillance (ICNS)*, 2016, p. 5.

[105] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min. (KDD)*, 2018, pp. 387–395. [Online]. Available: http://arxiv.org/abs/1802.04431

[106] Z. Ebrahimzadeh and S. Kleinberg, "Multi-scale change point detection in multivariate time series," presented at the NIPS Time Series Workshop, Long Beach, CA, USA, Dec. 2017.

[107] D. J. Hill, B. S. Minsker, and E. Amir, "Real-time Bayesian anomaly detection for environmental sensor data," in *Proc. Congr. Int. Assoc. Hydraulic Res.*, vol. 32, 2007, p. 503.

[108] H. Cheng, P.-N. Tan, C. Potter, and S. Klooster, "Detection and characterization of anomalies in multivariate time series," in *Proc. SIAM Int. Conf. Data Min.*, Apr. 2009, pp. 413–424. [Online]. Available: https://epubs.siam.org/doi/10.1137/1.9781611972795.36

[109] S. Zhang, C. Fung, S. Huang, Z. Luan, and D. Qian, "PSOM: Periodic self-organizing maps for unsupervised anomaly detection in periodic time series," in *Proc. IEEE/ACM 25th Int. Symp. Qual. Service (IWQoS)*, Jun. 2017, pp. 1–6.

[110] F. Y. Edgeworth, "XLI. On discordant observations," *The London Edinburgh Dublin Philos. Mag. J. Sci.*, vol. 23, no. 143, pp. 364–375, 1887.

Andrew A. Cook was born in York, U.K. in 1987. He received the B.Sc. degree (Hons.) in astrophysics with mathematics and the M.Sc. degree in advanced computer science from the University of Keele, Newcastle-under-Lyme, U.K. in 2010 and 2018, respectively, where he is currently pursuing the Ph.D. degree in computer science as part of the Smart Energy Network Demonstrator Project.

His current research interests are in the detection of anomalous events in time-series data

Göksel Mısırlı received the Ph.D. degree in computing science from Newcastle University, Newcastle upon Tyne, U.K., in 2013.

He continued to work as a Researcher with the Interdisciplinary Computing and Complex BioSystems group, Newcastle University, until 2017. He then joined Keele University, Newcastle-under-Lyme, U.K., as a Lecturer. His research interests include the development of novel algorithms, computational design paradigms and frameworks, utilizing large, heterogeneous and complex datasets. Knowledge representation, ontologies, data integration and mining, visualization, computational workflows, data and model-driven design approaches are some of the areas that he explored. Particularly, he demonstrated the utility of these approaches to computationally design synthetic genetic circuits via the use of mathematical simulations and time series data.

Zhong Fan (Senior Member, IEEE) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1992 and 1994, respectively, and the Ph.D. degree in telecommunication networks from Durham University, Durham, U.K., in 1997.

He was a Chief Research Fellow with Toshiba Research Europe, Cambridge, U.K., and led multiple projects on IoT, smart grid, data analytics, and 5G communications. He is a Professor and the Academic Director of the Smart Energy Network Demonstrator with Keele University, Newcastle-under-Lyme, U.K. He is an Advisory Board Member of the IEEE Industry Community on IoT. He has worked extensively with academia and industry in many collaborative projects funded by Innovate U.K., EPSRC, and EU. He has published over 160 journals and conference articles and filed 35 patents. His current research interests are IoT and applications of AI to smart energy systems.