ORIGINAL ARTICLE

Journal Section

Knowledge Discovery for Enabling Smart Internet-of-Things: A Survey

Sudip Misra[1*] | **Anandarup Mukherjee**[1*] | **Arijit Roy**[2†]

[1]Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, W. Bengal, 721302, India

[2]Advanced Technology Development Center, Indian Institute of Technology Kharagpur, W. Bengal, 721302, India

Correspondence
Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur,India
Email: smisra@sit.iitkgp.ernet.in

Present address
[†]Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, W. Bengal, 721302, India

Funding information

The utility of knowledge discovery on the Internet of Things (IoT) and its allied domains is undeniably one of the most indispensable ones, which results in optimized placement architectures, efficient routing protocols, device energy savings, and enhanced security measures for the implementation. The absence of knowledge discovery in IoT results in just an implementation of large-scale sensor networks, which generates a huge amount of data, and which needs, an often under-optimized, processing for actionable outputs. In this survey, we explore the various domains of IoT for which knowledge discovery is inseparable from the application, and how it benefits the overall implementation of the IoT architecture.

KEYWORDS

IoT, *Knowledge Discovery*, Sensor Networks, Mobile agents , Machine Learning, Deep Learning

1 | INTRODUCTION

The need for unification of technical and social domains, which were already connected to the Internet and inclusion of newer domains led to the emergence of the Internet of Things (IoT) paradigm (Minerva et al., 2015). Various international regulatory and standardization organizations define IoT according to their domain areas. For example, the Internet Engineering Task Force (IETF) defines IoT as *"The basic idea is that IoT will connect objects around us (electronic, electrical, non-electrical) to provide seamless communication and contextual services provided by them. Development of RFID tags, sensors, actuators, mobile phones makes it possible to materialize IoT which interact and co-operate each other to make the service better and accessible anytime, from anywhere."* (IETF, 2010). Similarly, the International Telecommunication Union – Telecommunication standardization unit's (ITU-T) Study group 13, which studies next-generation networks, and

IEEE define IoT as "A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies." (ITU-T Internet Reports, 2005), and as "A network of items each embedded with sensors – which are connected to the Internet." (IEEE), respectively. Essentially, IoT platforms are a collection of networked sensing, computation, and storage devices, which have the ability to sense, compute, decide and remotely actuate devices in order to integrate physical entities to digital systems for faster, accurate, reliable and efficient (with respect to cost, energy, speed, and size) task accomplishments.

IoT fosters reduced human intervention to achieve system and process-level automation, and aims to increase economic benefits for the subscribed industries and entities in the long run. The communication between these networked IoT devices can be a local or a global one, which ultimately unifies the physical devices, such as sensing motes (Filipponi et al., 2010), smartphones, vehicles (Zhang et al., 2015), home appliances, industrial devices (Majeed and Rupasinghe, 2017), software (Morin et al., 2017) and other embedded electronic items with remotely located storage and computation devices. IoT constitutes a mixture of legacy technologies, present technologies in massive use, and upcoming technologies. Legacy technologies, such as wireless sensor networks (WSNs) (Filipponi et al., 2010), machine-to-machine communications (M2M) (Sarigiannidis et al., 2017), (Wan et al., 2012), peer-to-peer networks (P2P) (Andreini et al., 2011), cyber-physical systems (CPS) (Shih et al., 2016) are concerned with providing connectivity between devices and users and form the core of IoT ecosystem.

Current technologies, such as cloud computing (Jiang et al., 2014), big data (Klein, 2017), software-defined networking (SDN) (Flauzac et al., 2015) are being rapidly adopted for use with the legacy technologies to achieve increasingly robust, smart and energy-efficient IoT solutions. The futuristic and upcoming technologies, such as integration of 5G communication systems (Vukobratovic et al., 2016), interoperability (Desai et al., 2015), and context awareness (Goel et al., 2017), aim to address the challenges of denser, yet faster and smarter IoT solutions. The extent of infusion of IoT and its importance to day-to-day lives of people can be highlighted by exploring the nuances of its implementation technology. The present-day trend in IoT is witnessing the emergence of new terminologies, which are essentially based upon the application areas of IoT. Technologoies such as the Internet of Vehicles (IoV) (Dandala et al., 2017), Internet of Environment (IoE), Internet of People (IoP) (Miranda et al., 2015), Internet of Health (IoH), and Industrial IoT (Industry 4.0) (Majeed and Rupasinghe, 2017) consider the use of IoT-based solutions and their inherent advantages for their specific domains only. IoV majorly deals with automated fault diagnosis, fault tolerance, vehicle to vehicle (V2V) communications, and vehicle to infrastructure (V2I) communications. IoE concerns itself with the applications of IoT in environmental monitoring (Arora et al., 2018)(Gómez et al., 2017) and management tasks. IoP focuses mainly on social networking and behavioral aspects of connected people and eventually, the society at large. IoH deals with the specific challenges of the healthcare domain and makes use of the advantages provided by IoT – faster data rate (Satija et al., 2017), miniaturized sensing devices in the form of wearables (Soraya et al., 2017), privacy, and security – in the pursuit of solving these challenges. Finally, Industry 4.0 deals with the design, deployment and management of IoT in industries and harsh industrial environments.

The existence of a wide range of IoT connectible paradigms, each of which, in turn, are composed of a considerable number of architectures, sensing, transmission, and processing technologies, in addition to data types, necessitates the use of knowledge discovery tools for efficient usage and connectivity. This knowledge discovery has to be associated/implemented at all possible levels of the IoT architecture for each of these associated paradigms for enabling a truly smart IoT. In this paper, we provide a broader and collective view of knowledge discovery in IoT, incorporating all these terminologies and technologies under the purview of IoT. Towards this objective, we divide the various focus areas of IoT into the following four categories – 1) *IoT application domains*, 2) *Networking components in IoT*, 3) *Data generation in IoT*, and 4) *Data types in IoT*. In the context of these four categorized focus areas, this paper explores the means and feasibility of knowledge discovery to achieve partial or full smart IoT systems. The application of knowledge discovery for each of these four domains may not always take place after processing of acquired data. Unlike traditional knowledge discovery tools, where knowledge gets discovered post-processing of data, smart IoT-based knowledge discovery may sometimes require its application before data processing such as sensor node discovery, the discovery of network components, processing data offload, and others. This makes knowledge discovery in smart IoT systems significantly different from regular data-based knowledge discovery applications.

1.1 | IoT Application Domains

The various application domains of IoT covering all IoT implementations revolve around the six broad classes of Smart city, Healthcare, Agriculture, Transportation, Industry, and Environment, as shown in Fig. 1. The smart city class consists of all works including, but not limited to smart homes (excluding human activity and health detection and tracking) (Shih et al., 2016), buildings (Shih et al., 2016), surveillance (Juhana and Anggraini, 2016), security (Moskvitch, 2017), retail (Jiménez-Zarco et al., 2017), waste management (Anagnostopoulos et al., 2017) and education (Gupta et al., 2015).

The smart grid is yet another domain, which is considered as the driving force of smart cities (Hernández et al., 2013). Smart grid encompasses a network of transmission lines, substations, transformers, and other components that deliver electricity from the power plant to homes and businesses. These grids enable two-way digital communication between utilities and its customers, which eventually depends on sensing along transmission lines for the operation of smart grids. Smart grid enables computer-based autonomous automation and control of electric grids based on the temporally changing demands of consumers.

Healthcare (Bhatia and Sood, 2017), although a functional part of the smart city domain, has been classified as a separate category altogether, owing to the large volume and variety of works in this domain alone. Agriculture is yet another domain in which IoT and IoT-based systems are making huge impact. IoT-based solutions for agricultural operations starting from evaluation of land for cropping, to monitoring crop growth (Petkovics et al., 2017), and finally the collection and distribution of the harvest to various markets and warehouses have been successfully architected (Brewster et al., 2017). The users of these IoT-based solutions report significant reduction in crop wastage, increased crop yields and increased economic benefits, as compared to the traditional agricultural practices.

Transportation is another crucial domain of IoT which has significant overlap with the smart city, agricultural and industrial domains. However, as this domain can be defined independently of the overlapping domains, we limit our focus to all works pertaining to vehicular IoT systems (Rani et al., 2016), vehicle to vehicle communications (Dandala et al., 2017), vehicle to infrastructure communications (Zhang et al., 2015), logistical planning and management (Brewster et al., 2017) in this domain. The industrial IoT domain constitutes works related to industrial monitoring (Shinde and Bhagat, 2017), safety (Kanzaki et al., 2017), supply chain management (Majeed and Rupasinghe, 2017), and other industrial applications (Kiran et al., 2017) making use of IoT-based solutions for their operation and management. Finally, the environmental IoT domain addresses challenges pertaining to monitoring of air quality (Black and White, 2017), water quality (Encinas et al., 2017), wildlife (Elias et al., 2017), and other such applications.

1.2 | Networking Components in IoT

Architecting IoT deployment summarily consists of network components in three layers – 1) Remote storage and processing, 2) Proximal routing and processing infrastructure, and 3) End devices. Previously, legacy IoT implementations made use of remote servers (Zhang et al., 2017) connected over backbone IP-based network for data storage and processing. However, the majority of the present day IoT's remote storage and processing is performed in the

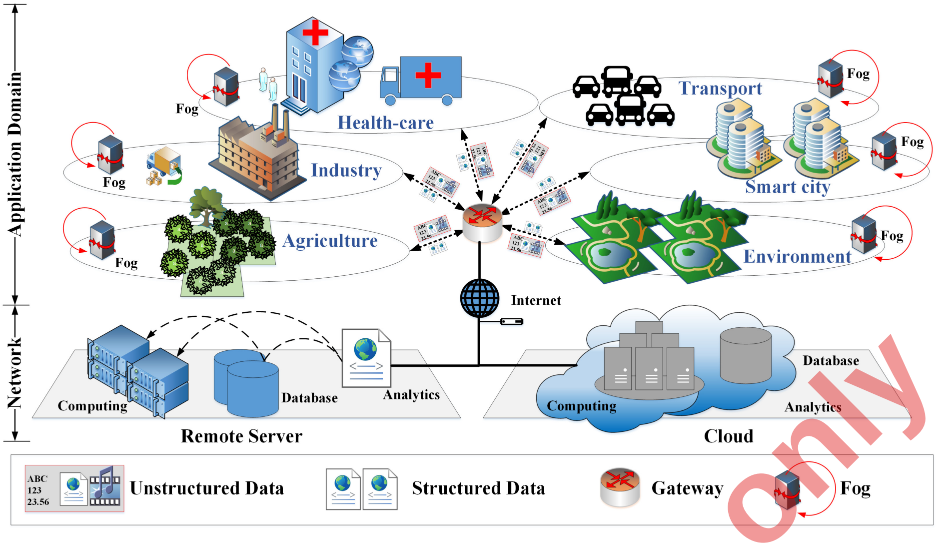


FIGURE 1 Heirarchy of the IoT networking ecosystem along with the various actors making-up the ecosystem.

cloud (Jiang et al., 2014). The cloud's capabilities – ease of access, faster deployment time, low maintenance, and on-demand upgrade of computational resources – from the standpoint of a user, makes it the most popular choice of platform for remote data integration, storage, and processing (Botta et al., 2016). This forms the topmost layer in the IoT ecosystem, as shown in Fig. 2. The second layer in the IoT ecosystem is termed as the proximal routing and processing infrastructure. This layer consists of networking component such as switches, routers, and gateways (Razafimandimby et al., 2017). Originally, this layer had the sole function of providing global IP-based connectivity to distributed local networks. In simpler terms, this layer is tasked with providing connectivity between the topmost and bottom-most layer of the IoT ecosystem (Fig. 2), as well as enabling communication between various distributed local networks without forwarding data and the topmost layer. However, more recent allied technologies such as the fog (Raafat et al., 2017), are becoming hugely popular owing to their added capability of processing and analysis of locally gathered data besides proving the regular unification and forwarding services and having the intelligence of selectively forwarding packets to the remote infrastructure in the topmost layer of the IoT ecosystem. This ensures speed-up of decision making and reduction in unnecessary congestion in the IP-based backbone network, which still relies on legacy IP-based protocols and infrastructure for all operations. The fog technology is beneficial to IoT sensors networks which make use of multimedia sensing, such as camera-based surveillance networks.

Finally, the third layer, which is depicted as the bottom layer in Fig. 2, consists of various distributed edge devices and edge-device networks (Karthikeya et al., 2018). The magnanimous permeation of IoT on a global scale is in the order of 8 Billion connected devices in 2017 [1], which is more than what 32-bit IPv4-based addressing can handle (IPv4 can have a maximum of $2^{32} \simeq 4.2$ billion addresses). However, as a significant portion of the global networking infrastructure still relies on IPv4-based technologies, which cannot be changed or terminated overnight, but are being utilized using various IP reuse strategies within the local networks, along with the gradual introduction of IPv6, which has an address range of 2^{64}. Each of the constituent networks in the bottom layer in the IoT ecosystem has its own individual IP ranges, which is locally unique within the network. Simply put, two networked edge devices may have the same IP

[1]Gartner Research, February 2017

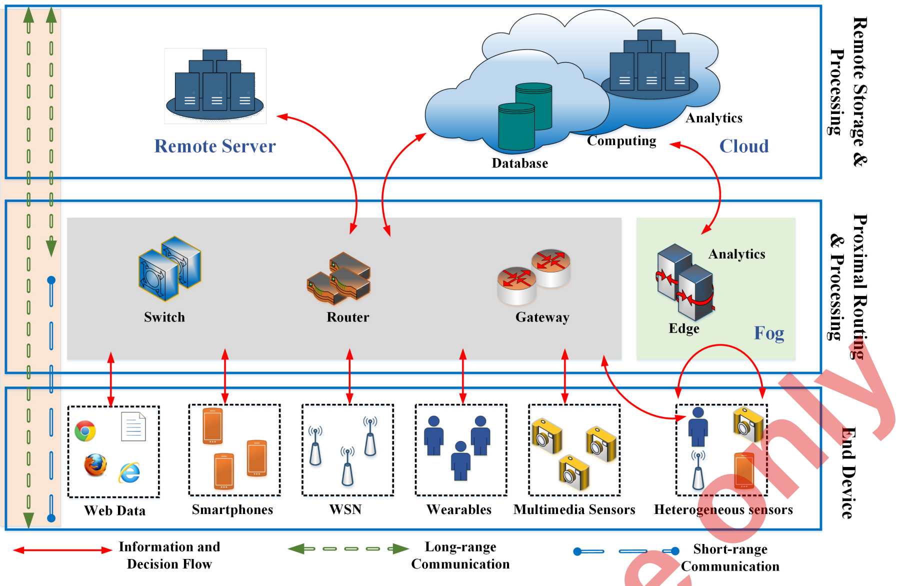


FIGURE 2 Heirarchy of the IoT networking ecosystem along with the various actors making-up the ecosystem.

address in separate networks, but not within the same local network. These local networks may be homogeneous or heterogeneous in nature.

The unifying network in an IoT ecosystem can be designated into two broad classes — *Long-range* and *Short-range* communication. The long-range communication technologies require access fees for their usage and consist of wired backbone networks, wireless cellular service networks, satellite-based networks and others. In contrast, short-range networks are generally without any access fee or licenses. Short-range networks can be achieved over wired local area networks (LAN) or wireless local area networks (WLAN). WLANs are more commonly formed in regular IoT applications using Bluetooth (Kasten and Langheinrich, 2001), Zigbee (Alliance, 2009), LoWPAN (Payer et al., 2009), LoRA (Vangelista et al., 2015), Wi-Fi (Baykas et al., 2011) or other license-free radio technologies.

1.3 | Data Types in IoT

The huge volume of data generated from all the networked devices in a global, a well as local, IoT ecosystem is attributed to sensors, devices, or equipment which may output or quantify their measurements in the form of text, sound, image, video or a combination of these. We delineate the IoT data into two groups – 1) Structured data and 2) Unstructured data, the source of which may be traced back to legacy devices as well as fairly modern ones. Most of the present day homogeneous IoT devices generate data in a structured format (Jiang et al., 2014), which have relational interdependencies. The structured data format restricts data length, data fields, as well as the data type (string, integer, floating point numbers) being used, which results in uniformly sized data packets, having precisely the same position of bits in the transmitted packets for all applications. This makes it easier to extract, organize and interprete of the data. For example, data from an online application form will be structured. In contrast, the use of legacy device networks, a mix of legacy and new devices, or heterogeneous device networks tend to generate data, which is inherently unstructured in nature (Jiang et al., 2014), (Datta et al., 2017). More often, the heterogeneous nature of the data makes it impossible to

TABLE 1 Data-based knowledge discovery based on sensing methodology.

Sensing Type	Structured Data	Unstructured Data
Explicit sensing (physical sensors)	(Jiang et al., 2014)	(Jiang et al., 2014), (Kuang et al., 2016)
Implicit sensing (virtual sensors)	(Guijarro et al., 2017), (Kaiwartya et al., 2017)	(Datta et al., 2017)
Group-based sensing (social sensing)	(Sharma and Kaur, 2017)	(Girau et al., 2017)
Collaborative sensing	(Ruta et al., 2017)	(Akpınar et al., 2017)

maintain uniformity during transmission or storage. An IoT-based smart city traffic monitoring system will typically have video feeds of intersections, sound-based sensors, traffic sensors, location sensors, and proximity sensors. Automating the process of event detection at a central station monitoring the information from a city-wide implementation of such systems will face the challenge of segregating and organizing the required information from all these sensors to gather requisite intelligence for each intersection.

1.4 | Data Generation in IoT

The multitude of actors in an IoT ecosystem is the main cause of variety and volume of data in IoT systems. Prior to analyzing data types in IoT, the differentiation of the sources and generators of these data is essential. Considering the various actors in an IoT ecosystem — sensors, actuators, people, virtual objects, services, platforms, and networks – the generation of data can be broadly classified into three major groups based on *mobility*, *sensor type*, and *sensing type* being used.

Depending on the mobility of the sensing platforms, data can be generated from – 1) static platforms, or 2) mobile platforms. The static platforms (Raafat et al., 2017) mainly constitute legacy IoT technologies such as WSN deployments in agriculture, M2M communication in smart offices and buildings, and other such systems and implementations. The data generation from these platforms mainly address the challenges of node/platform lifetime enhancement, addressing communication requirements, data amalgamation, deployment strategies, and deployment architectures. In contrast, the mobile platforms (Farris et al., 2017), (Verhelst and Moons, 2017) which constitute smartphones, vehicles, wearables and other such spatio-temporally dynamic systems deal with challenges in the areas of minimizing power consumption, localization, tracking, connectivity establishment, and maintenance, and processing offloading.

Depending on the sensor types used in the sensing platforms, data can be generated from either homogenous or heterogeneous sensors. However, for the sake of finer delineation, we group the sensing types into four groups – 1) scalar sensors, 2) multimedia sensors, 3) heterogeneous sensors, and 4) virtual sensors. Sensing using scalar sensors (Arora et al., 2018), such as proximity sensors, distance measurement sensors, pressure sensors, magnetic flux sensors, light sensors, color sensors, and others quantify information from surroundings or events into either digital or analog data, albeit without the information of the position, direction or nature of the sensed entity or phenomenon. Multimedia sensors (Costa et al., 2017) such as cameras and LiDARs, in contrast, encode information including information of the direction and position of the sensed entities and phenomenon. However, many sensing tasks are beyond the scope of operation of multimedia sensors such as sensing atmospheric pressure, temperature, distance, weight, and others. Heterogeneous sensors are a combination of scalar and multimedia sensors (Parida et al., 2012), which have the advantages of both of these sensor types. Finally, virtual sensors (Suciu et al., 2013) are a paradigm which makes use of information inferred from other physical sensors to generate and report data. For example, as the atmospheric temperature gradient is quite low within a given geographical location such as a town or city, virtual sensors can be suitably used to report temperature information from various parts of the city, without actually deploying too many physical sensors. Table 3 lists some of the commonly used sensors in the various IoT application domains.

Finally, depending on the sensing methodology used, data can be generated from sensors using – 1) explicit sensing, 2) implicit sensing, 3) social sensing, and 4) collaborative sensing. Explicit or direct sensing uses physical sensors in place to quantify actions (Jiang et al., 2014), natural phenomenon or event. Implicit sensing (Guijarro et al., 2017), in contrast, utilizes and virtualizes physically placed sensors to infer physical measurements for locations lacking actual sensors in certain scenarios, such as atmospheric condition measurements within a very small geographical area. Social sensing makes use of huge volumes of data aggregated from social media platforms (Sharma and Kaur, 2017) to generate trends and analyze behavior in order to understand events in quasi-real-time. Finally, collaborative sensing makes use of multiple distributed and heterogeneous sensors and sensor networks in order to generate actionable data (Akpınar et al., 2017).

1.5 | Use-case: Knowledge Discovery in Connected Self-driving Vehicles

The recent surge in research and developments in the domain of self-driving vehicles is coming to fruition due to the developments in areas of sensors, cloud computing, and IoT. These self-driving vehicles are capable of driverless navigation in real-world environments. The navigation of these vehicles is supported by a complex array of sensors, processors, and knowledge discovery processes. A typical self-driving car consists of sensors such as differential global positioning system (GPS), cameras, radio detection and ranging (RaDAR) and light detection and ranging (LiDARS). The combination of multiple such units of sensors in a single vehicle enables the vehicle to map its surroundings in real-time. The raw, heterogeneous data generated from these sensor combinations are fed to local computing infrastructures onboard the car, as well as, offloaded to remote processing infrastructures such as cloud or fog. The offloaded data is run through sophisticated algorithms, which discover the knowledge of the surroundings and obstructions recorded by the sensor arrays on the vehicle at every instant of time.

The process of knowledge discovery is applied in three sections of the connection between the vehicle and the remote processing infrastructure in an end-to-end manner:

1. **Sensor to local processing infrastructure:** The data collected from individual sensors need to be fused to generate essential actionable data. The local processing infrastructure onboard the vehicle ensures that the primary operations for operation and critical decision making in the vehicle are performed locally. To achieve this goal, the discovery of knowledge from raw sensorial data for data fusion and local discovery becomes imminent and critical for the operation of the vehicle.
2. **Selection of remote processing infrastructure:** The gathered and fused data onboard the vehicle, which can be offloaded for processing to a more powerful and resourceful remote infrastructure has to be initially channelized through the fastest possible route to the remote support. This requires detection of link qualities, the discovery of available and suitable network components and resources which can handle the data to be offloaded for near real-time knowledge discovery.
3. **Processing in the cloud/fog:** The data from the vehicle once offloaded to the remote infrastructure, which may be a cloud or a fog server, is used for discovery of information. This information is further processed through complex and robust algorithms to generate actionable inputs for the car, which is transmitted back to the vehicle.

1.6 | Motivation

IoT is a vast network of communicating devices connecting and exchanging data with each other in order to achieve various functionalities and operate in varying environments (Sulyman et al., 2017). Collectively, such vast numbers of devices, simultaneously generating heterogenous operational data, creates huge volumes of data, which gives rise to certain unique challenges associated with IoT – transmission, storage, searching, computation, and offloading. Despite the devices and entities in an IoT ecosystem being uniquely identifiable by means of network prefixes and hardware identifiers, the identification of data and extraction of information from such volumes of data is extremely challenging. The extracted information or knowledge from raw data in IoT networks is the driving force for achieving automation in various implementation scenarios. In this paper, we focus on the various aspects of knowledge discovery in smart IoT systems, which may have deployments spanning across cities as well as industries.

2 | KNOWLEDGE DISCOVERY FOR IOT

The challenges of discovering knowledge from data associated with IoT implementations rely heavily upon research in statistics, databases, pattern recognition, machine learning, data visualization, optimization, and high-performance computing. The discovered knowledge helps in enhancing the operational efficiency of the devices being controlled by the IoT network, generate advanced intelligence and analytics for industries and businesses, and enable faster discovery of services and systems online.

The knowledge discovery process is associated with certain unique features, especially when dealing with data of high velocity, variety, and volume. Features such as large data volume, the efficiency of knowledge discovery, the accuracy of discovered knowledge and discovery automation are all linked together, and have to be addressed together for reliable knowledge discovery (Soundararajan et al., 2005) in IoT networks. We group the IoT knowledge discovery types into two broad categories – *Posteriori* and *Priori*, such that

1. **Posteriori**: The knowledge extending from experience, empirical evidence, or historical data is considered as a priori knowledge (Liqiang et al., 2011). For example, the success of a crop harvest in a particular geographical and climatic region depends on historical data of the region's seasonal and climatic variations, water availability and soil type. This knowledge discovery class can be further fragmented into four types – *Situated*, *Dispersed*, *Empirical*, and *Procedural*, such that
   a. **Situated**: It is a highly specific knowledge type, which may be temporal or spatial in nature. For example, the trend followed by the city-wide office-hour vehicular congestion at traffic intersections (Reid et al., 2017) during specific times in a day may be considered as posteriori specific knowledge type. Other IoT domains, which fall under this category include agricultural crop monitoring (Liqiang et al., 2011), (Yan-e, 2011), and environmental monitoring (Mois et al., 2017).
   b. **Dispersed**: A posteriori dispersed knowledge type has fragmented information having no single source of truth, but makes use of historical trends and information from various locations to achieve decision making. For example, the behavior of traffic on city roads during a city-wide traffic management (Mehmood et al., 2017) (Dandala et al., 2017), or managing supply chain logistics in agricultural sectors (Brewster et al., 2017) can be considered as posteriori dispersed knowledge types.
   c. **Empirical**: The posteriori empirical knowledge type is formulated from quantitative and qualitative observations of experimental measurements occurring over a period of time. For example, the knowledge of a person's normal blood pressure range, based on the person's age and weight is used for generating healthcare alerts and

advisories, ensuring medication compliance (Bharadwaj et al., 2017), and others (Mezghani et al., 2017). Other domains constituting the use of this knowledge type include wearable health monitoring (Ravì et al., 2017), machine/ plant process monitoring (Shinde and Bhagat, 2017), agricultural water requirement analysis and monitoring (Indira et al., 2018), and forest ecosystem monitoring (Bayne et al., 2017).

   d. **Procedural**: The posteriori procedural knowledge type is formulated from the sequence of events to be followed under certain boundary conditions in order to accomplish a task such as the sequence of events to be followed in order to power-on a computer can be considered as posteriori procedural knowledge. Other examples constituting the use of procedural data include real-time QoS-aware healthcare data transmission (Satija et al., 2017), industrial process failure alerts (Xu et al., 2017), agricultural produce management across vast geographical regions (Brewster et al., 2017), and automated vehicular fault diagnosis (Rani et al., 2016).

2. **Priori**: It is the knowledge that is deduced irrespective of experience or historical evidence, and purely on the basis of first principles. For example, the ability of a computing machine to identify between an integer and a string data type or traffic flow control on city roads during peak traffic hours (Goel et al., 2017). This class of knowledge discovery can be further divided into two types — *Dispersed* and *Situated*.
   a. **Dispersed**: A priori dispersed knowledge type in IoT has fragmented instantaneous information having no single source of truth but needs to assemble truths from various sources to arrive at a decision. For example, the traffic sensors in a smart city gather information from various distributed locations to achieve traffic flow rerouting (Goel et al., 2017). Other similar examples include industrial process monitoring (Shinde and Bhagat, 2017), power distribution management (Morello et al., 2017), and disaster management (Kumar et al., 2017).
   b. **Situated**: A priori situated knowledge type makes use of instantaneous information available from a location to achieve decision making. For example, parking management (Ling et al., 2017) in a smart city relies on real-time information from various parking sensors spread across parking lots in a city in order to allow parking spaces to vehicles. Other examples of this knowledge discovery type include post-disaster response (Kumar et al., 2017), industrial incident monitoring (Xu et al., 2017), and public transportation services engagement and utilization (Fraga-Lamas et al., 2017), (Prandi et al., 2017).

Focusing on IoT specific application domain information, the major bulk of the knowledge discovery tasks can be distinguished into the following five tasks/challenges — *Node discovery*, *Service discovery*, *Data discovery*, *Resource discovery*, and *Information discovery*. These five challenges are elaborated in Sections 2.1 - 2.5. Additionally, Table 2 summarizes the various knowledge discovery tasks/challenges according to the IoT application domains.

2.1 | Node Discovery

The bulk of IoT nodes (end-devices) in various scenarios are designed and deployed considering the prominent factor of reduction of energy consumption. The most common scheme of conservation of energy in IoT deployments at the level of end devices is through sleep scheduling, wherein the nodes power-off or power-down for designated intervals of time in order to prolong their deployment lifetime. Any node in an ad-hoc configuration needs to discover its neighbors in order to forward its sensed data to a local sink, which eventually forwards it to a local server or cloud (refer Fig. 2) by means of intermediate networking components such as routers, switches, or gateways. Within the layer of end-devices, node discovery is additionally used for estimating the shortest path, the quickest path, and fault-resilient path between a source and sink node in ad-hoc configuration. The same tasks are also applicable for data transmission between the three layers – from the end devices to the remote server or cloud. Various popular solutions to node discovery in IoT include time synchronized protocols (e.g., recursive binary tree partitioning, wakeup scheduling), deterministic approaches (e.g., Searchlight), probabilistic approaches (e.g. Bloom filters (Dautov et al., 2017)), colocation-based approaches (e.g., randomized discovery, context-aware power management discovery), fully distributed opportunistic approaches (e.g., efficient application-layer discovery protocol), and learning-based approaches (e.g., Q-learning) (Valarmathi et al., 2016).

2.2 | Resource Discovery

The ubiquitous nature of IoT presents another challenge in the form of nature and location of available resources within the connected IoT ecosystem. This resource information helps in deciding the requirements and operational boundaries of the applications supported by an IoT deployment. Resources may include information about available sensors and their types, available computing and storage resources, middlewares, and networking components available within the specific IoT ecosystem. Various resource discovery mechanisms in IoT include — distributed discovery services, centralized discovery, Constrained Application Protocol (CoAP) based discovery, semantic discovery, and others such as Object Name Service (ONS) and Domain Name Service (DNS) (Datta et al., 2015). As resource discovery can be employed to optimize the energy consumption of the IoT-based solution, it can be similarly used to look for optimum sources of energy to power the IoT solution (Kraemer et al., 2017). Solutions such as these highlight the robustness and vast scope for knowledge discovery in IoT.

2.3 | Service Discovery

The various services associated with IoT deployments can be categorized into four groups — (a) Identity-related Services, (b) Information Aggregation Services, (c) Collaborative-Aware Services, and (d) Ubiquitous Services (Al-Fuqaha et al., 2015). The identity-related services are the most primary ones in an IoT ecosystem, which provide all IoT components being networked and connected to the Internet with assigned identities/tags for participation in an application. The information aggregation services deal with collection and aggregation of data from sensor units and other end-devices. Collaboration-aware services primarily aim to organize and execute application-oriented decision making based on the collected data from sensors or end-device groups. These work in conjunction with information aggregation services. Finally, ubiquitous services are focused on spreading the reach and availability of collaborative-aware services to all aspects of IoT.

2.4 | Data Discovery

The collection of heterogeneous data generation and storage entities in an IoT ecosystem enables ubiquitousness in applications. However, it gives rise to the problem of data redundancy. For example, an application-specific data in social sensing may be already present in the system, without the need for it to be freshly sensed or mined. However, if this data is not discovered, the IoT application will redundantly try to collect the data everytime it is queried by an user, which gives rise to wasteful expenditure of compution and energy. Besides addressing the problem of redundancy, the data discovery process entails details such as type, size, distribution and representation of the data (Tsai et al., 2014), the discovery of which is highly critical in the process of resource and service discovery in IoT.

TABLE 2 Knowledge discovery techniques in IoT domains for various sensor types.

IoT Domains	Scalar	Multimedia	Heterogeneous	Virtual
Smart City	SD (Filipponi et al., 2010)	SD-DD-ID-RD (Costa et al., 2017)	SD-RD (Cirani et al., 2014), SD (Wu et al., 2014), ND-SD-DD-ID (Puiu et al., 2016), ND-SD-DD-RD-ID (Sanchez et al., 2014)	ND-SD-DD-RD (Suciu et al., 2013)
Smart-grid	ND-ID (Matta et al., 2018), RD (Ambikapathy et al., 2018)	-	RD-ID (Ou et al., 2012) (Zhou and Rodrigues, 2013)	RD-ID (Hernández et al., 2013)
Health-care	SD (Ngu et al., 2017), SD-DD-ID (Bhatia and Sood, 2017)	SD-DD-RD-ID (Li et al., 2017), (Hassan et al., 2017)	DD-ID (Parida et al., 2012)	SD-DD-RD-ID (Singh et al., 2017)
Agriculture	ND-DD-ID (Baggio, 2005)	ND-SD-DD (Petkovics et al., 2017)	DD-RD-ID (Popović et al., 2017)	ND-SD (Ojha et al., 2017)
Transportation	ND-DD-ID (Sun et al., 2017), (Arora et al., 2018)	ND-DD-RD (Al-Turjman et al., 2017)	ND-DD-ID (Al-Dweik et al., 2017)	-
Industry	ND-RD (Kiran et al., 2017)	ND-SD-DD-RD-ID (Kanzaki et al., 2017)	ND-SD-DD-RD-ID (Höller et al., 2017)	-
Environment	ND-DD-ID (Arora et al., 2018)	DD-ID (Elias et al., 2017)	ND-SD-DD-ID (Gómez et al., 2017)	-

ND, node discovery; **SD**, service discovery; **DD**, data discovery; **RD**, resource discovery; **ID**, information discovery.

2.5 | Information Discovery

The vast numbers of devices and device types generating data for IoT presents a major challenge in terms of increase in complexity of the discovery of knowledge as the data travels from the end-device layer to the remote processing layer (refer Fig. 2). Many present-day IoT applications make use of middleware (Perera et al., 2014) to collect data and derive relevant information or context (Abowd et al., 1999) from it, before connecting the data to a remote processing infrastructure. This way, the data is shaped into manageable proportions which are already partially processed, instead of remaining as raw sensorial values. Information discovery solutions are being used successfully for detecting security threats (Tsitsiroudi et al., 2016) and providing robust security alert generations (Sarigiannidis et al., 2015) during transfer of data over IP enabled WSNs.

TABLE 3 Various sensor types and sensors associated with applications in IoT domains.

IoT Domains	Scalar sensors	Multimedia sensors
Smart City	Ultrasonic, RFID, GPS, Temperature, Humidity, Pressure, Proximity, Barometer, Strain, Light, Color, Speed, Magnetic flux, Electric flux	Visual camera, Depth camera, Thermal camera, LiDAR, RaDAR
Smart-grid	Magnetic flux, Electric flux, Temperature, Proximity	Visual camera, Thermal camera
Health-care	Electrocardiogram (ECG), Electroencephalogram (EEG), Pulse oximeter, Barometer, Sphygmomanometer, Temperature, Proximity, Infrared temperature detector	Visual cameras, Depth cameras
Agriculture	Soil moisture, Soil temperature, Barometer, Temperature, Rainfall, Anemometer, Color, Strain, GPS	Visual cameras, Thermal cameras, Multispectral cameras, Hyperspectral cameras
Transportation	Speed, Electric flux, Magnetic flux, Temperature, Pressure, Airflow, Strain, Color, Rainfall, RFID, Ultrasonic	Visual cameras, LiDARs, RaDARs, SoNARs, Depth cameras
Industry	Radiation, Color, Pressure, Temperature, Humidity, Proximity, Light, Strain, Electric flux, Magnetic flux, Ultrasonic	Visual cameras, Thermal cameras, Multispectral cameras, LiDARs, SoNARs
Environment	Soil moisture, Soil temperature, Barometer, Temperature, Rainfall, Solar radiation, Anemometer, Color, Strain, GPS, Proximity	-Visual camera, Thermal camera, LiDARs, Multispectral cameras

3 | THE KNOWLEDGE DISCOVERY PROCESS

Smart IoT systems are characterized by its reduced human intervention and dynamic capabilities, which, although universal, can adapt to applications and various architecture in order to function. The crux of the smart dynamism of IoT is attributed to the knowledge discovery process followed within an IoT ecosystem. The knowledge discovery process for IoT networks can be fragmented into the following four sequential steps – 1) Application domain information, 2) Pre-processing, 3) Tools for knowledge discovery, and 4) Data interpretation and Application. Each of these steps is further discussed in detail in Sections 3.1 - 3.5.

3.1 | Application Domain Information

The first step towards knowledge discovery in IoT necessitates the comprehensive information and understanding of the domain for which knowledge discovery is to be performed. This involves the collection and evaluation of the end-users' requirements, application-specific requirements, resources in use, available resources, deployment architecture, criticality of the domain, and targets to be achieved. Traditionally, application domain information is modeled under one of the following three categories – 1) Academic research (Fayyad et al., 1996), 2) Industry (Cabena et al., 1998), and 3)

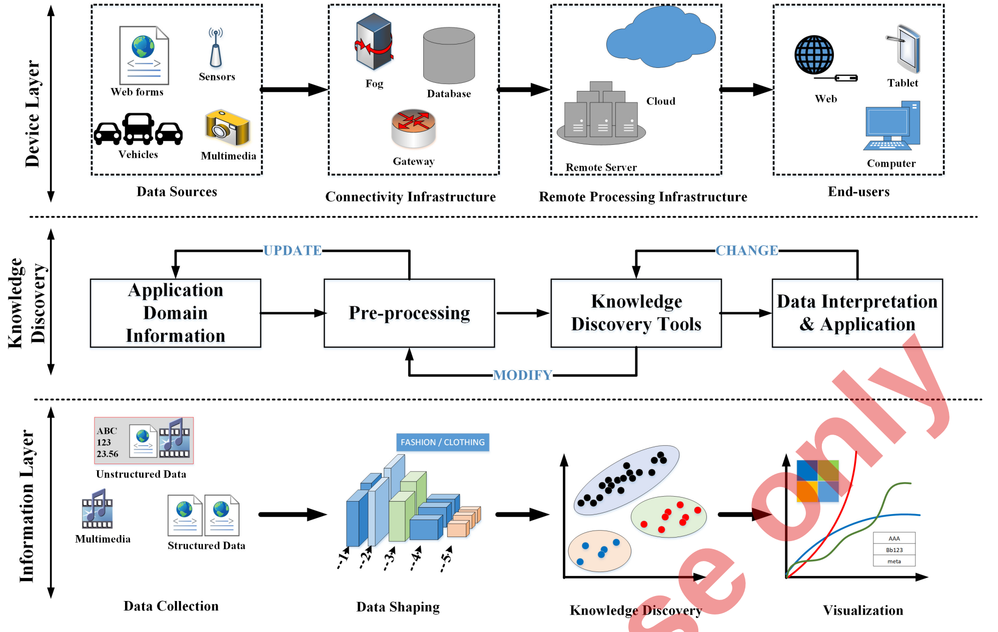


FIGURE 3 The knowledge discovery process in an IoT ecosystem.

Hybrid (Cios et al., 2000).

The academic research-based application domain information model is detailed with a rich technical and domain-specific knowledge, but lacks the aspects of business and economic perspectives. The industrial models are mainly focused on business understanding and processes and can be easily understood by non-domain experts. Finally, the hybrid models contain both the features of academic research and industrial models. The vast majority of application domain information models in IoT follow hybrid models. As the first step in domain-specific knowledge discovery, the collection and analysis of the domain requirements enable for a smoother transition from this step to the next one — *Pre-processing*.

3.2 | Pre-processing

The typically dense and heterogeneous nature of IoT deployments, as well as the unbiased use of new as well as legacy devices and data formats, necessitates the frequent, but judicious, use of data pre-processing. Additionally, as IoT implementations strive to have a unified global view of the deployment domain, the problem of data amalgamation is often encountered, which requires the unification of heterogeneous data, which may be structured or unstructured. One of the most powerful, yet nascent research areas in IoT is data interoperability, which can be Semantic Gateway-based (Desai et al., 2015), Hub-based (Blackstock and Lea, 2014), Across architectures (cloud-gateway-client) (Soursos et al., 2016), or Service-oriented architecture (SOA) based (Derhamy et al., 2017). These unification strategies further complicate knowledge discovery in IoT by introducing data with multi-dimensional and multi-modal features, which have interdependencies and need to be processed together. For example, deployment of IoT solutions in agricultural management generates data-set, which comprises of scalar data from soil moisture, soil temperature, and water level sensors, as well as multimedia data from visual, thermal and multispectral cameras, which measure the various vegetation indices and monitor plant growth. In addition to the complications in the nature of data, often technical disturbances within the network such as loss of signal, interference, sensor errors, and others corrupt the generated data stream. The application of data pre-processing prior to application of various tools for knowledge discovery on the data obtained ensures the removal of corrupted data, unbiasing the data, and the selection of the proper discovery tool as per the domain application.

3.3 | Tools for Knowledge Discovery

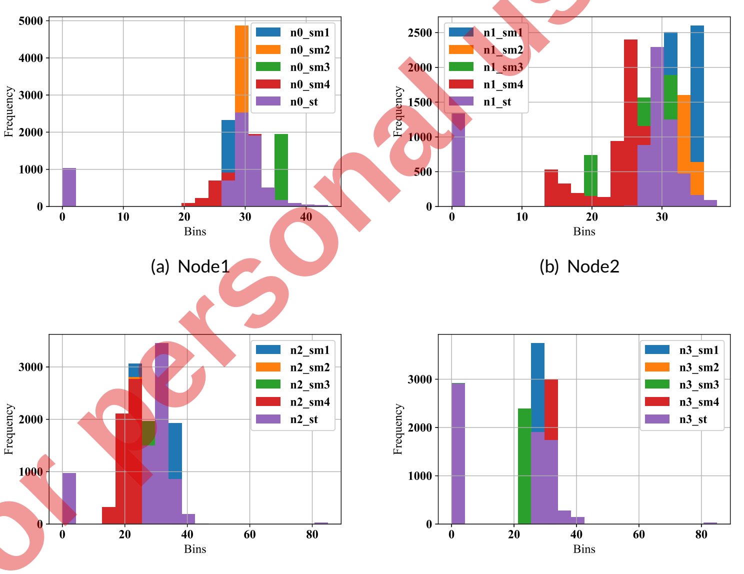
We broadly categorize the various tools of IoT knowledge discovery as follows — 1) Predictive to and 2) Forecasting. Prediction tools make use of prior information of an event to probabilistically estimate or characterize the outcome of the event at a specific point in time and is independent of time. On the other hand, forecasting can be considered as time-dependent prediction, which is able to estimate trends for a time-series data based on the previous behavior of the same time-series data.

1. **Predictive tools**: use explanatory variables for characterizing an event's output, and can be considered as one-off estimates of an event at a specific time in the future. For example, the outcome of the roll of an unbiased die can be characterized using predictive tools. Predictive tools for IoT can be further categorized as either machine learning-based or graph-based.
   a. **Machine learning-based**: These tools are based on algorithms which learn from the data to either infer patterns in the data or classify the data under various categories, without being explicitly programmed to do so. The machine learning algorithms have been categorized according to their usage in IoT deployment scenarios under the three classes – supervised, unsupervised, and reinforcement learning.
      i. **Supervised**: The supervised learning algorithms are predictive, and work on labeled data, where each prior data is tagged with some identifier *tag* or *label* or *class* during model generation/learning. New unlabelled data can be assigned to the learned models in this learning paradigm in order to categorize (classify or predict) them according to the previously defined labels. Some examples of supervised machine learning-based predictive tools being used in IoT are KNN (Cui et al., 2017), Decision Trees (Soraya et al., 2017), Neural Networks (Javed et al., 2017), Support Vector Machines (SVM) (Fekade et al., 2017), and Bayesian networks (Razafimandimby et al., 2017).
      ii. **Unsupervised**: The unsupervised learning algorithms are primarily descriptive in nature and work on unlabeled data. The majority of their operations are concerned with either clustering of data or discovering patterns in them. Some examples of unsupervised machine learning-based predictive tools being used in IoT are K-means clustering (Terán et al., 2017), Principal Component Analysis (PCA) (Yu et al., 2017), Self Organized Maps (SOM) (Tsirmpas et al., 2015), and Vector quantization (Mukherjee et al., 2017b).
      iii. **Reinforcement Learning**: This class of machine learning algorithms are initially trained in a coarse manner prior to deployment of the algorithm. Post-deployment, the algorithm updates its learning by continually learning from the incoming data. One of the most popular forms of reinforcement learning algorithms, which also finds use in IoT-based deployments is Q-learning (Zhu et al., 2017).
   b. **Graph-based**: These are highly relational and structured knowledge representations, which can be used for observing patterns or behavior of events. Some graph-based predictive tools in IoT are probabilistic graph models (Mohsin et al., 2017) and Markov models (Jia et al., 2017).
2. **Forecasting tools**: use observed trends in time-series data for characterizing an event's output, and tend to have generalized points in time. For example, the seasonal variations in weather over a geographical location can be

characterized using forecasting tools, with one specific feature as time. Regression is an example of a forecasting tool(Dharur et al., 2017).

3.4 | Case Study: Knowledge Discovery in Agricultural IoT

We consider an agricultural IoT implementation for comparing the performance of some of the common knowledge discovery tools. The dataset for our case-study is taken from our real-life implementation of agricultural IoT nodes, which follows a master-slave architecture at the field level. Four slave nodes ($n0$, $n1$, $n2$, and $n3$), each armed with four soil moisture sensors ($sm1$, $sm2$, $sm3$, and $sm4$) and a soil temperature sensor (st) are placed separately in various agricultural plots. The data from these nodes are periodically transmitted to a master node every 30 minutes through a Zigbee radio. The master node is connected to the Internet and periodically uploads the data collected from these nodes to a remote server. Fig. 4 shows the distribution of the collected data from these four nodes for over a period of 12 weeks. One of the most common data outliers obtained is from the underground temperature sensor, which shows sporadically sensed values almost equal to 0° C and greater than 40° C during some of the sensing instances. The distribution of received data from these four slave nodes is shown in Figs. 4(a), 4(b), 4(c), and 4(d). The different spread of the data is due to separate crop watering regimens followed at the four placement locations.



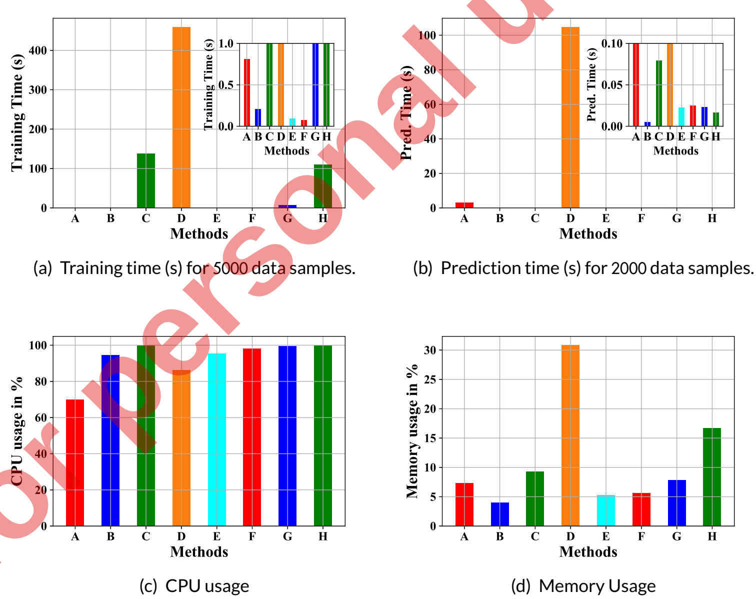
(a) Node1 (b) Node2

(c) Node3 (d) Node4

FIGURE 4 The distribution of data received at a remote server from various networked agricultural sensor nodes. Each of the four IoT nodes consist of 4 soil moisture sensors placed at additively increasing depths of 15cms. and a soil temperature sensor at a depth of 10 cms. below the ground.

Seven knowledge discovery tools – K Nearest Neighbors (KNN), Decision Trees, Neural Networks, Support Vector Machines (SVM), Linear Regression, Bayesian Networks, K-means Clustering, and Affinity Propagation – are chosen to uniquely identify the data originating nodes from the collection of the sensed data only. These seven methods are chosen as they can be directly used for learning patterns of the scalar data, especially in our scenario. These knowledge discovery tools are hosted on a low-power processor board (Raspberry Pi), which finds popular usage as an edge device in many IoT applications. Fig. 5 shows the four metrics chosen to evaluate the performance of these seven tools – training time, prediction time, CPU usage, and system memory usage. The dataset consisting of 7000 polled instances is split randomly into training and testing data, each with 5000 and 2000 data instances, respectively. In Fig. 5(a), it is seen that SVM incurs the maximum time for training followed by neural networks and affinity propagation. Similarly, a much higher time for prediction is again incurred by SVM as compared to others (as shown in Fig. 5(b)). Regarding CPU usage, KNN shows the least percentage of use (Fig. 5(c)), whereas most of the methods tend to have very high CPU use percentages. Interestingly, the CPU usage of SVM is lesser than most of the other methods, which fared better regarding training and prediction times. Finally, in Fig. 5(d), it is seen that decision trees have the smallest footprint, whereas SVMs tend to have an enormous memory footprint.

Summarizing the performance of these tools in terms of prediction accuracies, we observe that KNN, decision trees, and neural networks have a very high training and testing accuracy score (> 96%), whereas tools such as SVM and Bayesian networks show accuracy scores in the range of 60% approximately. As our dataset is highly location and trend specific, the unsupervised learning tools (K-means and affinity propagation) have a below-par performance of $<$ 10% concerning accuracy scores. A linear regression cannot be measured regarding accuracies, and we leave it out from the accuracy score comparison.



(a) Training time (s) for 5000 data samples.

(b) Prediction time (s) for 2000 data samples.

(c) CPU usage

(d) Memory Usage

FIGURE 5 A comparison of the various metrics on a resource constrained edge IoT node (Raspberry Pi) for the following methods – **A**: KNN, **B**: Decision Trees, **C**: Neural Networks, **D**: SVM, **E**: Linear Regression, **F**: Bayesian network, **G**: K-means clustering, **H**: Affinity Propagation.

Table 4 provides a side-by-side comparison of the seven knowledge discovery tools concerning their use in IoT-based systems. As IoT applications are defined for the constraints of time, processing, and memory, Table 4 compares the listed knowledge discovery tools against these parameters itself. It is to be noted that some methods in unsupervised learning such as SOM (Guo et al., 2017) and vector quantization (Pai et al., 2018) are too processing and time intensive to be of any useful comparison with the seven chosen tools. Similarly, semi-supervised tools such as Q-learning (Zhu et al., 2017) and Adversarial Networks (Alzantot et al., 2017) are also resource intensive to be of any practical use in resource-constrained edge devices such as the Raspberry Pi.

TABLE 4 A comparison of the performance of various knowledge discovery tools for their usage in IoT systems.

Method	Class	Training time	Prediction time	CPU usage	Memory usage	References
KNN	S	o o	o o	o o	o o o	(Galvão et al., 2017), (Cui et al., 2017)
Decision Trees	S	o	o	o o o o	o	(Wei et al., 2018), (Cui et al., 2017)
Neural Networks	S	o o o	o o	o o o o	o o	(Galvão et al., 2017)
SVM	S	o o o o	o o o o	o o o	o o o o	(Wei et al., 2018), (Galvão et al., 2017)
Linear Regression	S	o	o	o o o	o	(Cui et al., 2017)
Bayesian Networks	S	o	o	o o o o	o	(Wei et al., 2018), (Cui et al., 2017)
K-means clustering	U	o o	o	o o o o	o o	(Cui et al., 2017), (Shukla, 2017)
Affinity propagation	U	o o o	o	o o o o	o o o	(Cui et al., 2017)

S, Supervised; **U**, Unsupervised; **o**, Very low; **oo**, Low; **ooo**, Medium; **oooo**, High;

3.5 | Data Interpretation and Application

After the application of knowledge discovery algorithms/tools, the next (and final) stage in the IoT knowledge discovery process involves the interpretation of the information generated in order to achieve tasks outlined in Sections 2.1 - 2.5. Interpretation approaches such as graphs, charts, and tables may be used to decide rules or thresholds for updating the operational bounds of the agents in the IoT ecosystem. The updates can be directly mapped in the form of new relations in databases and modified rules in rule-based operations in order to smartly actuate devices, perform routing or disburse information to end-users. The interpreted IoT data is then used for various applications by enabling features such as context awareness, resource management, and analytics, in addition to empowering the existing features handling communication, security, data, routing, and sensing. The specific networking challenges addressed by knowledge discovery on various levels in the IoT ecosystem for static as well as mobile infrastructure are summarized in Table 5.

TABLE 5 Agent-based knowledge discovery techniques for various components of IoT.

IoT Component	Static	Mobile
Cloud	Se (Mukherjee et al., 2017a), (Zhou et al., 2017), Da (Ranjan et al., 2017), An (Patel et al., 2017), Da, Re (Bhandari et al., 2017), Re, Co, An (Petrolo et al., 2017), Ca (Mingozzi et al., 2016)	Se, Sn, Re (Yan et al., 2017), Re (Farris et al., 2017), (Zhang et al., 2015), Da, Co (Vukobratovic et al., 2016), Se, Da, Co (Ma et al., 2015)
Gateway	Re, Sn, An, Co (Billet and Issarny, 2017), (Petrolo et al., 2017)	Re, Sn, An, Co (Billet and Issarny, 2017)
Fog/ Edge	An (Patel et al., 2017), Da, An, Co, Re, Se (Sahni et al., 2017), (Verhelst and Moons, 2017), Se (Mukherjee et al., 2017a), Ro (Naranjo et al., 2017), An, Da (Raafat et al., 2017)	Re (Farris et al., 2017), Da, Re, Co, An (Sun and Ansari, 2016), Ca, Co (Roca et al., 2018), Re (Farris et al., 2017)
End Nodes	Da (Raafat et al., 2017), Ro (Naranjo et al., 2017), An, Re, Se (Verhelst and Moons, 2017), Ca, An (Chaochaisit et al., 2016)	An, Re, Se (Verhelst and Moons, 2017)

Ca: Context Awareness; **Co**: Communication; **Se**: Security; **Ay**: Analytics; **Da**: Data; **Ro**: Routing, **Sn**: Sensing, **Re**: Resource.

4 | SYNTHESIS

Synthesizing from the vast amounts of works being pursued in IoT, the challenges immediately pertaining to networking issues in order to enable smart IoT systems include *Node discovery*, *Resource discovery*, *Service discovery*, *Data discovery*, and *Information discovery*. An interesting aspect of the IoT ecosystem, which makes the knowledge discovery process quite challenging is mobility (static and mobile infrastructure), which is attributed to the ubiquitous nature of IoT. This ubiquitousness necessitated the emergence of numerous universal, as well as application-specific data formats which can be grouped into either of the two — structured and unstructured data. The complex nature of data generated from IoT data sources, which can be *scalar*, *multimedia*, *heterogeneous*, or *virtual*, results in data packets having varying packet lengths, varying packet structures, and has varying requirements in terms of power, security, privacy, and data-rate. For example, medical IoT deployments emphasize more on privacy and data-rate, as compared to focus on power consumption in agricultural IoT deployments, or focus on security and data-rate in surveillance-based IoT deployments.

The variety of options that are available in the hardware, software, and middlewares that are simultaneously integrated into a single end-to-end IoT solution poses immense challenges to the security of the whole solution. Additionally, the constrained nature of these IoT nodes, and access provided to various node discovery mechanisms are easy means to compromise the security of the IoT solution. Security threats may include Denial of Service (DoS), Physical damage, Eavesdropping, Node capture, and partial or full device control (Roman et al., 2013). Various solutions such as inclusion of active round-trip time measurements along with network scans for heterogeneous environments (Metongnon and Sadre, 2018), Security Fusion as a Service (SFaaS) based detection mechanisms working with software switches to detect attacks (Kuo et al., 2018), cognitive data analysis and route-discovery based threat detection (Tsitsiroudi et al., 2016), and other knowledge discovery based approaches are proposed to address the security issues during device discovery in IoT. Unlike traditional sensor network systems, such as WSN or wired networks, knowledge discovery in IoT is a complex task, involving the use of knowledge, which is spread across deployments or industries and is seldom restricted to a single layer in the IoT ecosystem. In the pursuit of discovering knowledge from IoT deployments for enabling smarter IoT systems, the operational span of these five networking challenges spread across various levels of IoT, such as *the remote storage and processing infrastructure*, *proximal routing and sensing layer*, and *end device layers*, as illustrated in Fig. 2. Table 6 summarizes the pros and cons of knowledge discovery for various components of IoT against specific IoT features – Processing power (Processing), Connectivity requirements (Connectivity), Energy consumption (Energy), Security, Cost of implementation (Cost), and Reliability.

The large density of IoT devices and its global perforation necessitates the judicious use of limited resources such as energy and network bandwidth. Ideally, IoT devices should have low energy (Energy) (Biason et al., 2018), and network bandwidth (Connectivity) demands. However, at the same time, it is immensely desirable for these IoT devices to have high computation power (Processing), high-security measures (Security) (Sarigiannidis et al., 2015), and high post-deployment device and data reliability (Reliability). Additionally, the increased dependence of technical solutions on IoT also necessitates that the cost of each device/ solution is kept to a minimum to achieve maximum usability for the masses. In order to provide a complete overview of the areas, which come under the immediate purview of knowledge discovery in IoT, Table 6 lists the various components to consider as, *Sensors*, *Sensing*, *Data*, *Services*, *Layer*, and *Applications*. *Sensors* evaluate the various sensor types, whereas *Sensing* evaluates the different sensing paradigms in the context of IoT. The sensed data (*Data*) explores the needs of the distinct data types generated – structured and unstructured. The *Services* component estimates the requirements for knowledge discovery in IoT from a discovery and management point of view, whereas the *Layers* component does it from a physical device point of view. Finally, *Applications* evaluate the requirements for IoT-based knowledge discovery under three sub-categories – (a) Very critical, (b) Critical, and (c) Non-critical. The first sub-category is for real-time systems such as guidance systems of vehicles, aircraft, and others. Critical systems are those requiring near real-time updates on data and decisions such as healthcare. Finally, non-critical systems have very low dependencies on time such as parking management, agriculture, and others.

5 | FUTURE TRENDS FOR KNOWLEDGE DISCOVERY IN IOT

The various work domains of IoT clearly indicate a trend of convergence towards smarter IoT systems. Future IoT systems will be inherently armed with enabling features allowing for seam interoperability across various deployment architectures. Knowledge discovery trends in these IoT systems will build upon the solutions addressing the following categories

1. **Knowledge discovery for denser IoT deployments**: Future IoT deployments will witness intelligent and dynamic addressing strategies, which will be able to assign application-specific addresses to devices as soon as they associate with a network, and reallocate it during their disassociation from the network. These strategies will have to extensively rely on knowledge discovery in order to identify the application context and scope of operation of the devices joining a network.
2. **Knowledge discovery for interoperability**: Future trends aiming towards ubiquitousness, but with the drastic reduction in energy consumption by optimizing data from sensors and data-flow within the networks will rely heavily on data interoperability. This interoperability will not only enable unification of data formats but also enable the inclusion of a vast variety of heterogeneous network and sensing technologies to the core infrastructure.
3. **Knowledge discovery for low-power analytics**: The ever-rising number of simultaneously connected IoT devices in

TABLE 6 Comparison of knowledge discovery against common IoT features for various IoT components.

Components	Specifics	Processing	Connectivity	Energy	Security	Cost	Reliability
Sensor	Scalar	Low	Low	Low	-	Low	Low/High
	Multimedia	High	High	High	-	High	High
	Heterogeneous	High	High	High	-	High	High
	Virtual	Low	Low/High	Low	-	Low	Low/High
Sensing	Explicit	Low/High	Low/High	Low	High	Low/High	High
	Implicit	Low	Low/High	Low	Low	Low	Low/High
	Group	Low	High	High	Low	High	Low/High
	Collaborative	Low	Low	Low	Low	High	High
Data	Structured	Low	-	Low	Low	-	High
	Unstructured	High	-	Low	High	-	Low
Services	Node discovery	Low	High	Low	Low	-	High
	Service discovery	High	Low	Low	Low	-	High
	Data discovery	High	High	Low	High	-	High
	Resource discovery	High	High	Low	High	-	High
	Information discovery	High	Low	High	High	-	Low
Layer	Cloud	High	High	High	High	High	High
	Gateway	High	High	Low	Low	Low/High	High
	Fog/Edge	High	Low	Low	Low	Low	Low/High
	End nodes	Low	Low/High	Low	Low	Low	Low/High
Applications	Very critical	High	High	Low/High	High	High	High
	Critical	High	High	Low/High	High	Low/High	Low/High
	Non-critical	Low	Low	Low	Low	Low	Low/High

High: signifies high to very high costs; **Low**: signifies low to very low costs; **Low/High**: signifies conditional dependence; **-**: not applicable.

all walks of life will necessitate the need for solutions aiming to reduce network transmission costs and transmission energy by limiting the distance a data has to travel in order to be organized into actionable knowledge. The rise of fog and edge computing paradigms have been successful in significantly reducing the processing load at remote servers and cloud data centers by introducing local, and on-site analytics and decision making. However, future improvements in this category will have to maintain a balance between the power consumption and size of edge devices and the load on the remote infrastructure.

REFERENCES

Abowd, G., Dey, A., Brown, P., Davies, N., Smith, M. and Steggles, P. (1999) Towards a better understanding of context and context-awareness. In *Handheld and ubiquitous computing*, 304–307. Springer.

Akpınar, K., Ballard, T., Hua, K. A., Li, K., Tarnpradab, S. and Ye, J. (2017) COMMIT: A Multimedia Collaboration System for Future Workplaces with the Internet of Things. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, 349–360. ACM.

Al-Dweik, A., Muresan, R., Mayhew, M. and Lieberman, M. (2017) IoT-based multifunctional Scalable real-time Enhanced Road Side Unit for Intelligent Transportation Systems. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 1–6.

Al-Fuqaha, A., Guizani, M., Mohammadi, M., Aledhari, M. and Ayyash, M. (2015) Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials*, **17**, 2347–2376.

Al-Turjman, F., Radwan, A., Mumtaz, S. and Rodriguez, J. (2017) Mobile traffic modelling for wireless multimedia sensor networks in IoT. *Computer Communications*, **112**, 109–115.

Alliance, Z. (2009) IEEE 802.15. 4, ZigBee standard.

Alzantot, M., Chakraborty, S. and Srivastava, M. (2017) Sensegen: A deep learning architecture for synthetic sensor data generation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 188–193. IEEE.

Ambikapathy, A., Singh, G. and Tiwari, P. (2018) Smart Switching Algorithm Between IC and PO Algorithms for Grid-Connected PV System. In *Advances in Smart Grid and Renewable Energy*, 83–92. Springer.

Anagnostopoulos, T., Zaslavsky, A., Kolomvatsos, K., Medvedev, A., Amirian, P., Morley, J. and Hadjiefthymiades, S. (2017) Challenges and Opportunities of Waste Management in IoT-enabled Smart Cities: A Survey. *IEEE Transactions on Sustainable Computing*.

Andreini, F., Crisciani, F., Cicconetti, C. and Mambrini, R. (2011) A scalable architecture for geo-localized service access in smart cities. In *Future Network & Mobile Summit (FutureNetw), 2011*, 1–8. IEEE.

Arora, C., Arora, N., Choudhary, A. and Sinha, A. (2018) Intelligent Vehicular Monitoring System Integrated with Automated Remote Proctoring. In *Intelligent Communication and Computational Technologies*, 325–332. Springer.

Baggio, A. (2005) Wireless sensor networks in precision agriculture. In *ACM Workshop on Real-World Wireless Sensor Networks (REALWSN 2005), Stockholm, Sweden*, 1567–1576.

Baykas, T., Sum, C.-S., Lan, Z., Wang, J., Rahman, M. A., Harada, H. and Kato, S. (2011) IEEE 802.15. 3c: the first IEEE wireless standard for data rates over 1 Gb/s. *IEEE Communications Magazine*, **49**.

Bayne, K., Damesin, S. and Evans, M. (2017) The internet of things–wireless sensor networks and their application to forestry. *NZ Journal of Forestry*, **61**, 37.

Bhandari, S., Sharma, S. K. and Wang, X. (2017) Cloud-assisted device clustering for lifetime prolongation in wireless IoT networks. In *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 1–4. IEEE.

Bharadwaj, S. A., Yarravarapu, D., Reddy, S. C. K., Prudhvi, T., Sandeep, K. S. P. and Reddy, O. S. D. (2017) Enhancing healthcare using m-care box (monitoring non-compliance of medication). In *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 167–171.

Bhatia, M. and Sood, S. K. (2017) A comprehensive health assessment framework to facilitate IoT-assisted smart workouts: A predictive healthcare perspective. *Computers in Industry*, **92**, 50–66.

Biason, A., Pielli, C., Zanella, A. and Zorzi, M. (2018) Access Control for IoT Nodes with Energy and Fidelity Constraints. *IEEE Transactions on Wireless Communications*.

Billet, B. and Issarny, V. (2017) Spinel: An Opportunistic Proxy for Connecting Sensors to the Internet of Things. *ACM Transactions on Internet Technology (TOIT)*, **17**, 21.

Black, I. and White, G. (2017) Citizen Science, Air Quality, and the Internet of Things. *The Internet of Things: Breakthroughs in Research and Practice: Breakthroughs in Research and Practice*, 137.

Blackstock, M. and Lea, R. (2014) IoT interoperability: A hub-based approach. In *2014 International Conference on the Internet of Things (IOT)*, 79–84.

Botta, A., De Donato, W., Persico, V. and Pescapé, A. (2016) Integration of cloud computing and Internet of Things: a survey. *Future Generation Computer Systems*, **56**, 684–700.

Brewster, C., Roussaki, I., Kalatzis, N., Doolin, K. and Ellis, K. (2017) IoT in Agriculture: Designing a Europe-Wide Large-Scale Pilot. *IEEE Communications Magazine*, **55**, 26–33.

Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. and Zanasi, A. (1998) *Discovering data mining: from concept to implementation*. Prentice-Hall, Inc.

Chaochaisit, W., Bessho, M., Koshizuka, N. and Sakamura, K. (2016) Human Localization Sensor Ontology: Enabling OWL 2 DL-Based Search for User's Location-Aware Sensors in the IoT. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, 107–111.

Cios, K. J., Teresinska, A., Konieczna, S., Potocka, J. and Sharma, S. (2000) Diagnosing myocardial perfusion from PECT bull's-eye maps-A knowledge discovery approach. *IEEE Engineering in Medicine and Biology Magazine*, **19**, 17–25.

Cirani, S., Davoli, L., Ferrari, G., Léone, R., Medagliani, P., Picone, M. and Veltri, L. (2014) A Scalable and Self-Configuring Architecture for Service Discovery in the Internet of Things. *IEEE Internet of Things Journal*, **1**, 508–521.

Costa, D. G., Collotta, M., Pau, G. and Duran-Faundez, C. (2017) A fuzzy-based approach for sensing, coding and transmission configuration of visual sensors in smart city applications. *Sensors*, **17**, 93.

Cui, W., Kim, Y. and Rosing, T. S. (2017) Cross-platform machine learning characterization for task allocation in IoT ecosystems. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, 1–7. IEEE.

Dandala, T. T., Krishnamurthy, V. and Alwan, R. (2017) Internet of Vehicles (IoV) for traffic management. In *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, 1–4. IEEE.

Datta, S. K., Bonnet, C. and Haerri, J. (2017) Extending DataTweet IoT Architecture for Virtual IoT Devices. In 10th *IEEE International Conference on Internet of Things (iThings-2017)*, 1–6.

Datta, S. K., Da Costa, R. P. F. and Bonnet, C. (2015) Resource discovery in Internet of Things: Current trends and future standardization aspects. In *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, 542–547. IEEE.

Dautov, R., Distefano, S., Senko, O. and Surnin, O. (2017) Property-Based Network Discovery of IoT Nodes Using Bloom Filters. In *International Conference on Human Centered Computing*, 394–399. Springer.

Derhamy, H., Eliasson, J. and Delsing, J. (2017) IoT Interoperability On-Demand and Low Latency Transparent Multiprotocol Translator. *IEEE Internet of Things Journal*, **4**, 1754–1763.

Desai, P., Sheth, A. and Anantharam, P. (2015) Semantic Gateway as a Service Architecture for IoT Interoperability. In *2015 IEEE International Conference on Mobile Services*, 313–319.

Dharur, S., Hota, C. and Swaminathan, K. (2017) Energy efficient IoT framework for Smart Buildings. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 793–800. IEEE.

Elias, A. R., Golubovic, N., Krintz, C. and Wolski, R. (2017) Where's the Bear?-Automating Wildlife Image Processing Using IoT and Edge Cloud Systems. In *2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI)*, 247–258. IEEE.

Encinas, C., Ruiz, E., Cortez, J. and Espinoza, A. (2017) Design and implementation of a distributed IoT system for the monitoring of water quality in aquaculture. In *Wireless Telecommunications Symposium (WTS), 2017*, 1–7. IEEE.

Farris, I., Militano, L., Nitti, M., Atzori, L. and Iera, A. (2017) MIFaaS: a Mobile-IoT-Federation-as-a-Service Model for Dynamic Cooperation of IoT Cloud Providers. *Future Generation Computer Systems*, **70**, 126–137.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (1996) *Advances in knowledge discovery and data mining*, vol. 21. AAAI press Menlo Park.

Fekade, B., Maksymyuk, T., Kryk, M. and Jo, M. (2017) Probabilistic Recovery of Incomplete Sensed Data in IoT. *IEEE Internet of Things Journal*.

Filipponi, L., Vitaletti, A., Landi, G., Memeo, V., Laura, G. and Pucci, P. (2010) Smart City: An Event Driven Architecture for Monitoring Public Spaces with Heterogeneous Sensors. In *2010 Fourth International Conference on Sensor Technologies and Applications*, 281–286.

Flauzac, O., González, C., Hachani, A. and Nolot, F. (2015) SDN based architecture for IoT and improvement of the security. In *2015 IEEE* 29^{th} *International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, 688–693. IEEE.

Fraga-Lamas, P., Fernández-Caramés, T. M. and Castedo, L. (2017) Towards the internet of smart trains: a review on industrial IoT-connected railways. *Sensors*, **17**, 1457.

Galvão, Y. M., Albuquerque, V. A., Fernandes, B. J. and Valença, M. J. (2017) Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting. In *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–6. IEEE.

Girau, R., Martis, S. and Atzori, L. (2017) Lysis: A platform for IoT distributed applications over socially connected objects. *IEEE Internet of Things Journal*, **4**, 40–51.

Goel, D., Chaudhury, S. and Ghosh, H. (2017) An IoT approach for context-aware smart traffic management using ontology. In *Proceedings of the International Conference on Web Intelligence*, 42–49. ACM.

Gómez, J. E., Marcillo, F. R., Triana, F. L., Gallo, V. T., Oviedo, B. W. and Hernández, V. L. (2017) IoT FOR ENVIRONMENTAL VARIABLES IN URBAN AREAS. *Procedia Computer Science*, **109**, 67–74.

Guijarro, L., Pla, V., Vidal, J. R. and Naldi, M. (2017) Game Theoretical Analysis of Service Provision for the Internet of Things Based on Sensor Virtualization. *IEEE Journal on Selected Areas in Communications*, **35**, 691–706.

Guo, S., Guo, Z., Qiu, Z., Liu, Y. and Wang, Y. (2017) IFRAT: An IoT Field Recognition Algorithm based on Time-series Data. In *2017* 3^{rd} *International Conference on Big Data Computing and Communications (BIGCOM)*, 251–255. IEEE.

Gupta, A., Gupta, P. and Chhabra, J. (2015) IoT based power efficient system design using automation for classrooms. In *2015 Third International Conference on Image Information Processing (ICIIP)*, 285–289.

Hassan, M. M., Lin, K., Yue, X. and Wan, J. (2017) A multimedia healthcare data sharing approach through cloud-based body area network. *Future Generation Computer Systems*, **66**, 48–58.

Hernández, L., Baladron, C., Aguiar, J. M., Carro, B., Sanchez-Esguevillas, A., Lloret, J., Chinarro, D., Gomez-Sanz, J. J. and Cook, D. (2013) A multi-agent system architecture for smart grid management and forecasting of energy demand in virtual power plants. *IEEE Communications Magazine*, **51**, 106–113.

Höller, J., Tsiatsis, V. and Mulligan, C. (2017) Toward a Machine Intelligence Layer for Diverse Industrial IoT Use Cases. *IEEE Intelligent Systems*, **32**, 64–71.

IEEE () Special Report: The Internet of Things. URL: `http://theinstitute.ieee.org/static/special-report-the-internet-of-things`.

IETF (2010) The Internet of Things – Concept and Problem Statement. URL: `http://tools.ietf.org/id/draft-lee-iot-problem-statement-00.txt`.

Indira, D., Harshita, M., Pranav, D. S. and Sai, J. P. M. (2018) TILLAGE DRIP: An Efficient Seed Selection and Conservative Irrigation with Crop Defective Alert by IOT. In *Smart Computing and Informatics*, 53–62. Springer.

ITU-T Internet Reports (2005) ITU Internet Reports 2005: The Internet of Things. URL: `http://handle.itu.int/11.1002/pub/800eae6f-en`.

Javed, A., Larijani, H., Ahmadinia, A. and Gibson, D. (2017) Smart random neural network controller for HVAC using cloud computing technology. *IEEE Transactions on Industrial Informatics*, **13**, 351–360.

Jia, B., Li, W. and Zhou, T. (2017) A Novel P2P Service Discovery Algorithm Based on Markov in Internet of Things. In *2017 IEEE International Conference on Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC)*, vol. 2, 26–31. IEEE.

Jiang, L., Da Xu, L., Cai, H., Jiang, Z., Bu, F. and Xu, B. (2014) An IoT-oriented data storage framework in cloud computing platform. *IEEE Transactions on Industrial Informatics*, **10**, 1443–1451.

Jiménez-Zarco, A. I., Rospigliosi, A., Martínez-Ruiz, M. P. and Izquierdo-Yusta, A. (2017) Marketing 4.0: Enhancing Consumer-Brand Engagement. *Socio-Economic Perspectives on Consumer Engagement and Buying Behavior*, 94.

Juhana, T. and Anggraini, V. G. (2016) Design and implementation of Smart Home Surveillance system. In *2016 10^{th} International Conference on Telecommunication Systems Services and Applications (TSSA)*, 1–5.

Kaiwartya, O., Abdullah, A. H., Cao, Y., Lloret, J., Kumar, S., Shah, R. R., Prasad, M. and Prakash, S. (2017) Virtualization in wireless sensor networks: fault tolerant embedding for Internet of Things. *IEEE Internet of Things Journal*.

Kanzaki, H., Schubert, K. and Bambos, N. (2017) Video streaming schemes for industrial IoT. In *2017 26^{th} International Conference on Computer Communication and Networks (ICCCN)*, 1–7. IEEE.

Karthikeya, S. A., Narayanan, R. et al. (2018) Power-aware gateway connectivity in battery-powered dynamic IoT networks. *Computer Networks*, **130**, 81–93.

Kasten, O. and Langheinrich, M. (2001) First experiences with bluetooth in the smart-its distributed sensor network. In *Workshop on Ubiquitous Computing and Communications, PACT*, vol. 1.

Kiran, M., Subrahmanyam, V. and Rajalakshmi, P. (2017) Novel power management scheme and effects of constrained on-node storage on performance of MAC layer for industrial IoT networks. *IEEE Transactions on Industrial Informatics*.

Klein, S. (2017) The World of Big Data and IoT. In *IoT Solutions in Microsoft's Azure IoT Suite*, 3–13. Springer.

Kraemer, F. A., Ammar, D., Braten, A. E., Tamkittikhun, N. and Palma, D. (2017) Solar energy prediction for constrained IoT nodes based on public weather forecasts. In *Proceedings of the Seventh International Conference on the Internet of Things*, 2. ACM.

Kuang, L., Yang, L. T. and Qiu, K. (2016) Tensor-based software-defined Internet of Things. *IEEE Wireless Communications*, **23**, 84–89.

Kumar, J. S., Zaveri, M. A. and Choksi, M. (2017) Task Based Resource Scheduling in IoT Environment for Disaster Management. *Procedia Computer Science*, **115**, 846–852.

Kuo, C.-T., Chi, P.-W., Chang, V. and Lei, C.-L. (2018) SFaaS: Keeping an eye on IoT fusion environment with security fusion as a service. *Future Generation Computer Systems*.

Li, Y., Jeong, Y. S., Shin, B. S. and Park, J. H. (2017) Crowdsensing Multimedia Data: Security and Privacy Issues. *IEEE MultiMedia*, **24**, 58–66.

Ling, X., Sheng, J., Baiocchi, O., Liu, X. and Tolentino, M. E. (2017) Identifying parking spaces & detecting occupancy using vision-based IoT devices. In *Global Internet of Things Summit (GIoTS), 2017*, 1–6. IEEE.

Liqiang, Z., Shouyi, Y., Leibo, L., Zhen, Z. and Shaojun, W. (2011) A crop monitoring system based on wireless sensor network. *Procedia Environmental Sciences*, **11**, 558–565.

Ma, R., Li, J., Guan, H., Xia, M. and Liu, X. (2015) EnDAS: Efficient Encrypted Data Search as a Mobile Cloud Service. *IEEE Transactions on Emerging Topics in Computing*, **3**, 372–383.

Majeed, A. A. and Rupasinghe, T. D. (2017) Internet of things (IoT) embedded future supply chains for industry 4.0: An assessment from an ERP-based fashion apparel and footwear industry. *International Journal of Supply Chain Management*, **6**, 25–40.

Matta, N., Rahim-Amoud, R., Merghem-Boulahia, L. and Jrad, A. (2018) Putting Sensor Data to the Service of the Smart Grid: From the Substation to the AMI. *Journal of Network and Systems Management*, **26**, 108–126.

Mehmood, Y., Ahmad, F., Yaqoob, I., Adnane, A., Imran, M. and Guizani, S. (2017) Internet-of-things-based smart cities: Recent advances and challenges. *IEEE Communications Magazine*, **55**, 16–24.

Metongnon, L. and Sadre, R. (2018) Fast and efficient probing of heterogeneous IoT networks. *International Journal of Network Management*, **28**.

Mezghani, E., Exposito, E. and Drira, K. (2017) A Model-Driven Methodology for the Design of Autonomic and Cognitive IoT-Based Systems: Application to Healthcare. *IEEE Transactions on Emerging Topics in Computational Intelligence*, **1**, 224–234.

Minerva, R., Biru, A. and Rotondi, D. (2015) Towards a definition of the Internet of Things (IoT). *IEEE Internet Initiative*, **1**.

Mingozzi, E., Tanganelli, G., Vallati, C., Martínez, B., Mendia, I. and González-Rodríguez, M. (2016) Semantic-based context modeling for quality of service support in IoT platforms. In *2016 IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 1–6.

Miranda, J., Mäkitalo, N., Garcia-Alonso, J., Berrocal, J., Mikkonen, T., Canal, C. and Murillo, J. M. (2015) From the Internet of Things to the Internet of People. *IEEE Internet Computing*, **19**, 40–47.

Mohsin, M., Sardar, M. U., Hasan, O. and Anwar, Z. (2017) IoT Risk Analyzer: A Probabilistic Model Checking Based Framework for Formal Risk Analytics of the Internet of Things. *IEEE Access*.

Mois, G., Folea, S. and Sanislav, T. (2017) Analysis of Three IoT-Based Wireless Sensors for Environmental Monitoring. *IEEE Transactions on Instrumentation and Measurement*.

Morello, R., De Capua, C., Fulco, G. and Mukhopadhyay, S. C. (2017) A Smart Power Meter to Monitor Energy Flow in Smart Grids: The Role of Advanced Sensing and IoT in the Electric Grid of the Future. *IEEE Sensors Journal*, **17**, 7828–7837.

Morin, B., Harrand, N. and Fleurey, F. (2017) Model-Based Software Engineering to Tame the IoT Jungle. *IEEE Software*, **34**, 30–36.

Moskvitch, K. (2017) Securing IoT: In your smart home and your connected enterprise. *Engineering Technology*, **12**, 40–42.

Mukherjee, B., Neupane, R. L. and Calyam, P. (2017a) End-to-End IoT Security Middleware for Cloud-Fog Communication. In *Cyber Security and Cloud Computing (CSCloud), 2017 IEEE 4th International Conference on*, 151–156. IEEE.

Mukherjee, P., Samant, T., Swain, T. and Datta, A. (2017b) SEP-V: A solution to energy efficient technique in intra-cluster cooperative communication for wireless sensor network. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 204–208. IEEE.

Naranjo, P. G. V., Shojafar, M., Mostafaei, H., Pooranian, Z. and Baccarelli, E. (2017) P-SEP: a prolong stable election routing algorithm for energy-limited heterogeneous fog-supported wireless sensor networks. *The Journal of Supercomputing*, **73**, 733–755.

Ngu, A. H., Gutierrez, M., Metsis, V., Nepal, S. and Sheng, Q. Z. (2017) IoT Middleware: A Survey on Issues and Enabling Technologies. *IEEE Internet of Things Journal*, **4**, 1–20.

Ojha, T., Misra, S. and Raghuwanshi, N. S. (2017) Sensing-cloud: Leveraging the benefits for agricultural applications. *Computers and Electronics in Agriculture*, **135**, 96–107.

Ou, Q., Zhen, Y., Li, X., Zhang, Y. and Zeng, L. (2012) Application of internet of things in smart grid power transmission. In *Mobile, Ubiquitous, and Intelligent Computing (MUSIC), 2012 Third FTRA International Conference on*, 96–100. IEEE.

Pai, N.-S., Chen, J.-Y., Chen, P.-Y. and Hong, J.-H. (2018) Speech Processing Based on Hidden Markov Model and Vector Quantization Techniques Applied to Internet of Vehicles. *Sensors and Materials*, **30**, 803–820.

Parida, M., Yang, H. C., Jheng, S. W. and Kuo, C. J. (2012) Application of RFID Technology for In-House Drug Management System. In *2012 15th International Conference on Network-Based Information Systems*, 577–581.

Patel, P., Ali, M. I. and Sheth, A. (2017) On Using the Intelligent Edge for IoT Analytics. *IEEE Intelligent Systems*, **32**, 64–69.

Payer, U., Kraxberger, S. and Holzer, P. (2009) IPv6 Label Switching on IEEE 802.15.4. In *Third International Conference on Sensor Technologies and Applications. SENSORCOMM'09.*, 650–656. IEEE.

Perera, C., Zaslavsky, A., Christen, P. and Georgakopoulos, D. (2014) Context aware computing for the internet of things: A survey. *IEEE Communications Surveys & Tutorials*, **16**, 414–454.

Petkovics, I., Simon, J., Petkovics, Á. and Čović, Z. (2017) Selection of unmanned aerial vehicle for precision agriculture with multi-criteria decision making algorithm. In *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*, 000151–000156. IEEE.

Petrolo, R., Morabito, R., Loscrì, V. and Mitton, N. (2017) The design of the gateway for the Cloud of Things. *Annals of Telecommunications*, **72**, 31–40.

Popović, T., Latinović, N., Pešić, A., Zečević, Ž., Krstajić, B. and Djukanović, S. (2017) Architecting an IoT-enabled platform for precision agriculture and ecological monitoring: A case study. *Computers and Electronics in Agriculture*, **140**, 255–265.

Prandi, C., Nisi, V. and Nunes, N. (2017) Bus Stops as Interactive Touchpoints: Improving Engagement and Use of Public Transport. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter*, 20. ACM.

Puiu, D., Barnaghi, P., Tönjes, R., Kümper, D., Ali, M. I., Mileo, A., Parreira, J. X., Fischer, M., Kolozali, S., Farajidavar, N., Gao, F., Iggena, T., Pham, T. L., Nechifor, C. S., Puschmann, D. and Fernandes, J. (2016) CityPulse: Large Scale Data Analytics Framework for Smart Cities. *IEEE Access*, **4**, 1086–1108.

Raafat, H. M., Hossain, M. S., Essa, E., Elmougy, S., Tolba, A. S., Muhammad, G. and Ghoneim, A. (2017) Fog Intelligence for Real-Time IoT Sensor Data Analytics. *IEEE Access*, **5**, 24062–24069.

Rani, M. D., Pradeepa, J. J. and Shaby, S. M. (2016) Measurement and fault detection in intelligent wireless system using wireless devices. In *2016 International Conference on Communication and Signal Processing (ICCSP)*, 2236–2240.

Ranjan, R., Thakker, D., Haller, A. and Buyya, R. (2017) A note on exploration of IoT generated big data using semantics.

Ravì, D., Wong, C., Lo, B. and Yang, G. Z. (2017) A Deep Learning Approach to on-Node Sensor Data Analytics for Mobile or Wearable Devices. *IEEE Journal of Biomedical and Health Informatics*, **21**, 56–64.

Razafimandimby, C., Loscrí, V., Vegni, A. M. and Neri, A. (2017) A Bayesian and smart gateway based communication for noisy IoT scenario. In *2017 International Conference on Computing, Networking and Communications (ICNC)*, 481–485. IEEE.

Reid, A. R., Pérez, C. R. C. and Rodríguez, D. M. (2017) Inference of vehicular traffic in smart cities using machine learning with the Internet of Things. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 1–14.

Roca, D., Milito, R., Nemirovsky, M. and Valero, M. (2018) Tackling IoT Ultra Large Scale Systems: Fog Computing in Support of Hierarchical Emergent Behaviors. In *Fog Computing in the Internet of Things*, 33–48. Springer.

Roman, R., Zhou, J. and Lopez, J. (2013) On the features and challenges of security and privacy in distributed internet of things. *Computer Networks*, **57**, 2266–2279.

Ruta, M., Scioscia, F., Pinto, A., Gramegna, F., Ieva, S., Loseto, G. and Di Sciascio, E. (2017) A CoAP-based framework for collaborative sensing in the Semantic Web of Things. *Procedia Computer Science*, **109**, 1047–1052.

Sahni, Y., Cao, J., Zhang, S. and Yang, L. (2017) Edge Mesh: A New Paradigm to Enable Distributed Intelligence in Internet of Things. *IEEE Access*, **5**, 16441–16458.

Sanchez, L., Muñoz, L., Galache, J. A., Sotres, P., Santana, J. R., Gutierrez, V., Ramdhany, R., Gluhak, A., Krco, S., Theodoridis, E. et al. (2014) SmartSantander: IoT experimentation over a smart city testbed. *Computer Networks*, **61**, 217–238.

Sarigiannidis, P., Karapistoli, E. and Economides, A. A. (2015) VisIoT: A threat visualisation tool for IoT systems security. In *IEEE International Conference on Communication Workshop (ICCW), 2015*, 2633–2638. IEEE.

Sarigiannidis, P., Zygiridis, T., Sarigiannidis, A., Lagkas, T. D., Obaidat, M. and Kantartzis, N. (2017) Connectivity and coverage in machine-type communications. In *2017 IEEE International Conference on Communications (ICC)*, 1–6. IEEE.

Satija, U., Ramkumar, B. and Manikandan, M. S. (2017) Real-Time Signal Quality-Aware ECG Telemetry System for IoT-Based Health Care Monitoring. *IEEE Internet of Things Journal*, **4**, 815–823.

Sharma, P. and Kaur, P. D. (2017) Effectiveness of web-based social sensing in health information dissemination — A review. *Telematics and Informatics*, **34**, 194–219.

Shih, C. S., Chou, J. J., Reijers, N. and Kuo, T. W. (2016) Designing CPS/IoT applications for smart buildings and cities. *IET Cyber-Physical Systems: Theory Applications*, **1**, 3–12.

Shinde, K. S. and Bhagat, P. H. (2017) Industrial process monitoring using IoT. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 38–42. IEEE.

Shukla, P. (2017) ML-IDS: A machine learning approach to detect wormhole attacks in Internet of Things. In *Intelligent Systems Conference (IntelliSys), 2017*, 234–240. IEEE.

Singh, D., Tripathi, G., Alberti, A. M. and Jara, A. (2017) Semantic edge computing and IoT architecture for military health services in battlefield. In *2017 14^{th} IEEE Annual Consumer Communications & Networking Conference (CCNC)*, 185–190. IEEE.

Soraya, S. I., Chiang, T.-H., Chan, G.-J., Su, Y.-J., Yi, C.-W., Tseng, Y.-C. and Ching, Y.-T. (2017) IoT/M2M wearable-based activity-calorie monitoring and analysis for elders. In *2017 39^{th} Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2390–2393. IEEE.

Soundararajan, E., Joseph, J., Jayakumar, C. and Somasekharan, M. (2005) Knowledge discovery tools and techniques. *Recent Advances in Information Technology*, 141.

Soursos, S., Žarko, I. P., Zwickl, P., Gojmerac, I., Bianchi, G. and Carrozzo, G. (2016) Towards the cross-domain interoperability of IoT platforms. In *2016 European Conference on Networks and Communications (EuCNC)*, 398–402.

Suciu, G., Vulpe, A., Halunga, S., Fratu, O., Todoran, G. and Suciu, V. (2013) Smart Cities Built on Resilient Cloud Computing and Secure Internet of Things. In *2013 19^{th} International Conference on Control Systems and Computer Science*, 513–518.

Sulyman, A. I., Oteafy, S. M. and Hassanein, H. S. (2017) Expanding the cellular-IoT umbrella: An architectural approach. *IEEE wireless communications*, **24**, 66–71.

Sun, W., Liu, J. and Zhang, H. (2017) When smart wearables meet intelligent vehicles: challenges and future directions. *IEEE wireless communications*, **24**, 58–65.

Sun, X. and Ansari, N. (2016) EdgeIoT: Mobile Edge Computing for the Internet of Things. *IEEE Communications Magazine*, **54**, 22–29.

Terán, M., Aranda, J., Carrillo, H., Mendez, D. and Parra, C. (2017) IoT-based system for indoor location using bluetooth low energy. In *2017 IEEE Colombian Conference on Communications and Computing (COLCOM)*, 1–6. IEEE.

Tsai, C.-W., Lai, C.-F., Chiang, M.-C., Yang, L. T. et al. (2014) Data mining for Internet of Things: A survey. *IEEE Communications Surveys and Tutorials*, **16**, 77–97.

Tsirmpas, C., Anastasiou, A., Bountris, P. and Koutsouris, D. (2015) A New Method for Profile Generation in an Internet of Things Environment: An Application in Ambient-Assisted Living. *IEEE Internet of Things Journal*, **2**, 471–478.

Tsitsiroudi, N., Sarigiannidis, P., Karapistoli, E. and Economides, A. A. (2016) EyeSim: A mobile application for visual-assisted wormhole attack detection in IoT-enabled WSNs. In *2016 9^{th} IFIP Wireless and Mobile Networking Conference (WMNC)*, 103–109. IEEE.

Valarmathi, M., Sumathi, L. and Deepika, G. (2016) A survey on node discovery in Mobile Internet of Things (IoT) scenarios. In *2016 3^{rd} International Conference on Advanced Computing and Communication Systems (ICACCS)*, vol. 1, 1–5. IEEE.

Vangelista, L., Zanella, A. and Zorzi, M. (2015) Long-range IoT technologies: The dawn of LoRa^{TM}. In *Future Access Enablers of Ubiquitous and Intelligent Infrastructures*, 51–58. Springer.

Verhelst, M. and Moons, B. (2017) Embedded Deep Neural Network Processing: Algorithmic and Processor Techniques Bring Deep Learning to IoT and Edge Devices. *IEEE Solid-State Circuits Magazine*, **9**, 55–65.

Vukobratovic, D., Jakovetic, D., Skachek, V., Bajovic, D., Sejdinovic, D., Kurt, G. K., Hollanti, C. and Fischer, I. (2016) CONDENSE: A Reconfigurable Knowledge Acquisition Architecture for Future 5G IoT. *IEEE Access*, **4**, 3360–3378.

Wan, J., Li, D., Zou, C. and Zhou, K. (2012) M2M communications for smart city: An event-based architecture. In *2012 IEEE 12^{th} International Conference on Computer and Information Technology (CIT)*, 895–900. IEEE.

Wei, S., Zhao, X. and Miao, C. (2018) A comprehensive exploration to the machine learning techniques for diabetes identification. In *2018 IEEE 4^{th} World Forum on Internet of Things (WF-IoT)*, 291–295. IEEE.

Wu, Q., Ding, G., Xu, Y., Feng, S., Du, Z., Wang, J. and Long, K. (2014) Cognitive Internet of Things: A New Paradigm Beyond Connection. *IEEE Internet of Things Journal*, **1**, 129–143.

Xu, Y., Sun, Y., Wan, J., Liu, X. and Song, Z. (2017) Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications. *IEEE Access*.

Yan, H., Hua, Q., Zhang, D., Wan, J., Rho, S. and Song, H. (2017) Cloud-Assisted Mobile Crowd Sensing for Traffic Congestion Control. *Mobile Networks and Applications*, 1–7.

Yan-e, D. (2011) Design of intelligent agriculture management information system based on IoT. In *2011 International Conference on Intelligent Computation Technology and Automation (ICICTA)*, vol. 1, 1045–1049. IEEE.

Yu, T., Wang, X. and Shami, A. (2017) Recursive Principal Component Analysis based Data Outlier Detection and Sensor Data Aggregation in IoT Systems. *IEEE Internet of Things Journal*.

Zhang, H., Zhang, Q. and Du, X. (2015) Toward Vehicle-Assisted Cloud Computing for Smartphones. *IEEE Transactions on Vehicular Technology*, **64**, 5610–5618.

Zhang, X., Zhang, J., Li, L., Zhang, Y. and Yang, G. (2017) Monitoring citrus soil moisture and nutrients using an iot based system. *Sensors*, **17**, 447.

Zhou, J., Cao, Z., Dong, X. and Vasilakos, A. V. (2017) Security and privacy for cloud-based IoT: challenges. *IEEE Communications Magazine*, **55**, 26–33.

Zhou, L. and Rodrigues, J. J. (2013) Service-oriented middleware for smart grid: Principle, infrastructure, and application. *IEEE Communications Magazine*, **51**, 84–89.

Zhu, J., Song, Y., Jiang, D. and Song, H. (2017) A New Deep-Q-Learning-Based Transmission Scheduling Mechanism for the Cognitive Internet of Things. *IEEE Internet of Things Journal*, 1–1.