Cyber-Physical System Security Surveillance using Knowledge Graph based Digital Twins - A Smart Farming Usecase

Sai Sree Laya Chukkapalli*, Nisha Pillai †, Sudip Mittal†, Anupam Joshi*
*Dept. of Computer Science & Electrical Engineering, University of Maryland Baltimore County,
Email: {saisree1, joshi}@umbc.edu
† Department of Computer Science & Engineering, Mississippi State University,
Email: npillai1@umbc.edu, mittal@cse.msstate.edu

***Abstract*—Rapid advancements in Cyber-Physical System (CPS) capabilities have motivated farmers to deploy this ecosystem on their farms. However, there is a growing concern among users regarding the security risks associated with CPS. Especially with rising number of cyber-attacks on CPS, such as modifying sensor readings, interrupting operations, etc. Therefore, this paper describes a security surveillance framework to detect deviations in the ecosystem by incorporating a digital twin supported anomaly detection model. The reason for incorporating digital twins is that they add value by enabling real-time monitoring of connected smart farms. We pre-process the collected data from sensors deployed on the smart farm setup. The pre-processed data is fused with our smart farm ontology to populate a knowledge graph. The generated graph is further queried to extract the necessary sensor data. We utilize the extracted normal data to train the anomaly detection model. Further, we tested our model if it identifies abnormal values from sensors by simulating anomalous use case scenarios specific to our ecosystem.**

***Index Terms*—Security Surveillance, Cyber-Physical System, Digital Twins, Knowledge Graphs, Artificial Intelligence**

I. Introduction

Farmers are shifting towards smart farming techniques by integrating modern Cyber-Physical Systems (CPS) with existing farming practices. In addition, efficient use of resources and improved crop yield can be achieved with the current technological advancements in the CPS-based Artificial Intelligence (AI) applications [1], [2]. The role of AI applications is to provide insights for increasing operational functionality and product quality which is beneficial to the farmers. For example, an agriculture drone named AGRAS MG-1[1] sprays fertilizers, pesticides evenly over the agricultural land at a much faster rate so as to achieve lower pesticide use. However, existing frameworks focus more on the application possibilities of the data collected from CPS, and less emphasis is laid on the security of CPS. Gupta et al. [3] have elaborated multiple cyber-attacks that the adversaries can perform on CPS in a smart farm ecosystem. Therefore, protecting the CPS from security attacks and risks in real-time is still a challenging task [4], [5].

A critical step to address this challenge is utilizing digital twins. In recent years, the integration of real-time data from physical sensors and knowledge specific to the domain for building a "*Digital Twin*" has gained much attention in the global market. Fundamentally, digital twins are defined as a virtual copy of the physical sensors that connects both the physical and virtual entity [6]. Moreover, digital twins aid in surveillance and controlling the physical entity there by improving operational efficiency, productivity, etc. To continuously monitor real-time data generated by sensors present on the smart farm we need to design a *surveillance system*. Several fields, such as transportation, medical, military, etc., have developed surveillance applications. Surveillance for CPS is considered as an important application [7] in order to protect CPS from cyber-attacks and downtime. To achieve this in a smart farm ecosystem, we integrate our digital twin and CPS to detect abnormalities in sensors. Here, we perceive sensors as an element of the CPS ecosystem.

In this paper, we present a security surveillance framework for CPS in a smart farm environment that monitors sensor data generated and issues security alerts to farm owners. Alerts generated can help farm owners take necessary action beforehand to reduce failures or sensor downtime. We also describe our collected data, smart farm ontology, and digital twin modules present in our surveillance framework. First, we begin by utilizing the collected data from the sensors deployed in the smart farm setup [8]. In the next stage, we integrate the pre-processed data with a smart farm ontology [9] for generating knowledge graphs in RDF (Resource Description Framework) format [10]. Our reason for incorporating an ontology and knowledge graphs is that these enables knowledge reuse across domains [11], [12]. Knowledge graphs are also known for solving time sensitive data modelling problems. Using digital twins, we extract information from the knowledge graph and forward it to the anomaly detection model for training on normal observations of sensor data. For creating an anomaly detection model, we use Principal Component Analysis (PCA) [13] as it is the most popular technique for multivariate data. Later, the trained anomaly detection model is used on test data which contains simulated abnormal conditions of sensors to identify anomalous points that deviate from the normal data. Finally, we show that our digital twin representation based anomaly detection model can be used to identify various anomalous conditions on the smart farm.

The rest of the paper is organized as follows: Section II contains related work. Section III explains our architecture, while Section IV describes the experimental setup for our use cases to demonstrate our anomaly detection model's effectiveness. Finally, we conclude and discuss the possible

[1]https://www.dji.com/mg-1

future work in Section V.

II. Related Work

In this section, we describe some related work on the role of CPS in agriculture and digital twins. We also discuss applications of knowledge graphs and anomaly detection models.

A. Cyber-physical systems in agriculture

Driven by rapid technological advancements in cloud computing, big data, machine learning, etc., the agriculture sector started shifting towards smart farming. Multiple subsystems deployed on the farm when integrated provide numerous insights about condition monitoring, planning of farm operations, optimizing productivity, etc. For example, Jagannathan et al. [14] described an automated agriculture task system. Here the system sprayed the required water by monitoring the existing soil water content. Rupanagudi et al. [15] developed a framework to identify insects in tomatoes at an earlier phase through continuous monitoring of smart farms. Another image processing tool to detect diseases in plants right from the stage of planting to harvesting was proposed by Jhuria et al. [16]. However, these applications impose new challenges to smart farms in terms of security threats [3]. Sontowski et al. [8] also showcased a real time attack on smart farms. Therefore, it is essential to have security surveillance for smart farms.

B. Digital twin for cyber-physical systems

Digital Twins are known for simulating real-time behavior of physical systems. The concept of using twins was first introduced by NASA's Apollo program, where a virtual replica of a space vehicle was created to carry out flight operations. Incorporation of digital twins by various CPS sectors can help in real-time monitoring, identifying equipment failures, etc. For example, in manufacturing [17], these models can be utilized for quality management where continuous monitoring of product data from various devices has an added advantage over traditional inspection based techniques. In healthcare, simulations play an essential role in medical surgery and equipment design. Elayan et al. [18] designed a digital twin based anomaly detection model that detects heart conditions based on the data obtained from wearable devices. Xu et al. [19] proposed a anomaly detection framework based on digital twins where anomalous events were detected on critical infrastructure testbeds. In the paper [20], digital twins for smart farm has been described in order to identify soil moisture content. However, their work was mainly investigative in nature.

C. Knowledge graphs and their applications

Knowledge graphs store large amounts of data in the form of triplets where *Subject* and *Object* represent head and tail nodes, and the relation between them is represented as *Predicate*. Moreover, knowledge graphs are known for allowing easier querying of the required information as data is interconnected. The querying is done by utilizing SPARQL Protocol and RDF Query Language (SPARQL) [21] where data is represented as instances of the knowledge graph. Some of the widely used knowledge graphs that are supported by semantic schemas include YAGO [22], DBpedia [23], NELL [24], Google Knowledge Graph [25], etc. In the recent years, knowledge graphs are widely being incorporated in multiple domains such as healthcare [26], manufacturing [27], cybersecurity [28], [29], etc. Several knowledge graphs supported real-world applications [30] for language representation learning, question answering and recommendation systems have been built to provide efficient services. For instance, in the manufacturing domain, an application was developed by Banerjee et al. [31] where they extract and infer knowledge from large-scale production line data to enhance process management in manufacturing. Their paper described a semantic query mechanism utilized to estimate the minimum response rate and turnaround time for a non-defective product.

D. Security for cyber-physical systems

With the increasing connectivity of CPS, researchers face new challenges regarding the recent exposure of CPS to several cyber-attacks and threats. To combat this problem, multiple CPS domains have incorporated anomaly detection models to identify anomalous events and further keep their CPS secure and reliable. Zeng et al. [32] introduced a machine learning based intrusion detection method that detects malicious nodes in vehicular ad-hoc networks. Dutta et al. [33] explained a Hidden Markov Model (HMM) based anomaly detection model that identifies and alerts the owner of a smart home whenever an anomalous event is detected in the network or behavioral data. In the paper [34], anomalous events in clean water supply systems obtained from the testbed were identified and evaluated by utilizing multiple machine learning approaches such as support vector machines, k-nearest neighbours, and random forest. Hao et al. [35] developed a statistical machine learning approach to detect abnormal patterns that have low false omission rates in industrial control systems by utilizing Seasonal Auto-Regressive Integration Moving Average (SARIMA) based dynamic threshold model.

Extensive work has been done in various domains that utilize digital twins for anomaly detection. However, incorporating digital twins and knowledge graphs for anomaly detection in a smart farm environment has not been done to the best of our knowledge. As discussed earlier, we believe that our digital twin supported anomaly detection approach offers a new surveillance perspective to identify real-time anomalous events from multiple sensors deployed in a smart farm environment.

III. System Architecture

In this section, we explain our surveillance framework, which contains a digital twin supported anomaly detection model to automatically detect an abnormal value in a sensor while thoroughly monitoring the captured data from sensors deployed in a smart farm environment. The system architecture of our proposed security surveillance framework for CPS to

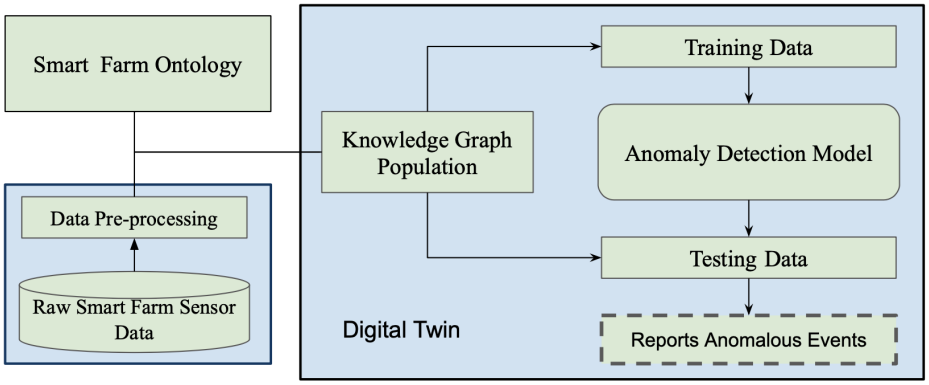

Fig. 1. Architecture of our security surveillance framework for CPS.

tackle the security risks in the smart farm ecosystem is shown in Figure 1.

Our framework consists of three modules that are interlinked with each other :

- *Data collection & pre-processing*: This is the first step in which physical sensors' real-time data is read from the data source. We pre-process the data, and further extract the required features for our work.
- *Smart farm ontology*: In this phase, we utilize an existing ontology [9] to provide information about entities and relationships in the smart farm ecosystem.
- *Digital twin*: This is the final phase where the digital twin contains two sub-modules: knowledge graph and anomaly detection. We populate a knowledge graph by linking the real time data and entities defined in our smart farm ontology. Further, we extract the required data from the knowledge graph through SPARQL Protocol and RDF Query Language (SPARQL) [21] queries. The data then serves as an input to generate an anomaly detection model. Finally, we use the generated model to detect abnormal measurements in sensors.

We elaborate on various components in the following sections.

A. Data collection & pre-processing

For this work, we have used the data collected by Sontowski et al. [8]. The authors have deployed sensors in a smart farm setup connected via Raspberry Pi that pushes data to the cloud whenever there is a change in the observation. The data given was captured for 10 days and contains information such as the status of grove base hat, camera, and measurements from sensors like barometer, grove light sensor, air quality sensor, capacitive moisture sensor, and air quality sensor. Changes in physical properties such as temperature, pressure, humidity, light intensity and air quality of a sensor are updated immediately.

The data collected from all the sensors, including their timestamp, are combined and stored in the Microsoft Azure cloud. A sample of measurements for each sensor and their status are given in Table I. The data present had 9 features

Timestamp	Sensor	Measurement	Status
29-07-2020 18:39:35	AirHumidity	40.7	
29-07-2020 18:39:35	AirPressure	100.59	
29-07-2020 18:39:35	AirTemperatureC	19.92	
29-07-2020 18:39:35	AirTemperatureF	9.27	
29-07-2020 18:39:35	Camera		Not Installed
29-07-2020 18:39:35	GroveBase HAT		Installed
29-07-2020 18:39:35	Light	26.84	
29-07-2020 18:39:35	Soil Moisture	25.47	
29-07-2020 18:40:49	AirHumidity	42.42	
29-07-2020 18:40:49	AirPressure	98.73	
29-07-2020 18:40:49	AirTemperatureC	19.84	
29-07-2020 18:40:49	AirTemperatureF	9.27	
29-07-2020 18:40:49	Camera		Not Installed
29-07-2020 18:40:49	GroveBase HAT		Installed
29-07-2020 18:40:49	Light	26.84	
29-07-2020 18:40:49	Soil Moisture	25.47	

TABLE I
OBSERVED DATA FOR EACH SENSOR.

such as Time, AirHumidity, AirPressure, AirTemperatureC, AirTemperatureF, CameraStatus, GroveBaseHatStatus, Light, and SoilMoisture1. Two distinct values, categorical and numerical, are seen across the columns of the dataset. CameraStatus and GroveBaseHatStatus columns are represented as categorical nominal values, and the rest of the columns are represented as numerical values.

Data pre-processing techniques were applied to the given data set to eliminate null values and convert categorical nominal data into numerical values using label encoding. As label encoding takes less processing time and does not affect the dimensions of the dataset.

B. Smart farm ontology

We re-use smart farm ontology created in our previous work [9] as it contains all the general concepts related to the smart farming ecosystem. The smart farm ontology developed is based on various physical sensors, and their relationships where the sensors are represented as classes and their relationships are represented as properties. We explain below some of the important classes and properties that we have used in our implementation.

The class *Time* is an essential class for our work as it indicates the timestamp for every recorded observation of a physical sensor deployed in the smart farm setup. For example, the barometer sensor's change in temperature at a particular time is captured by this class. A *MemberFarm* class monitors all the interactions happening in the farm, such as status and readings of all the sensors deployed in the smart farm setup. *Farm Based Units (FBU)* is an entity that refers to all the physical sensors deployed in the farm. In our case, soil moisture, barometer, light, air quality sensor are represented as individuals of this *Farm Based Units (FBU)* class. The *Farm Based Units (FBU)* class is a sub-class of the *Sensor* class. *Observation* class is associated with two classes such as *SensorData* and *Time* The *SensorData* class represents the measurement value of the physical sensors deployed in the smart farm and does not focus on the temporal aspect.

The property *hasValue* holds an instance of *Farm Based Units (FBU)* class that contains a data value. This data value represents the status or measurement of the physical sensor. The subject entity belongs to the *Observation* class and the object entity belongs to *SensorData* class. *hasTime* is a relationship between *Observation* class and *Time* class which provides the timestamp at which the data was recorded.

C. Digital twin

Our digital twin supports security surveillance of CPS by continuously monitoring the data generated from physical sensors. This is done by extracting real-time normal data through a knowledge graph for training the detection model. Later, we simulate the anomalous condition of sensors and add it to the test data for validating our detection model. A detailed explanation for each sub-module of digital twin is presented below.

1) Knowledge graph: In this sub-module, we describe our approach in populating a knowledge graph. We use the knowledge graph to obtain structured knowledge from the heterogeneous data generated by the physical sensors. Our approach integrates real-time data with our extended smart farm ontology to populate the knowledge graph. In which real-time data is obtained after pre-processing the raw data collected from physical sensors containing a stream of numerical values, including a timestamp. Furthermore, generated knowledge graph present in the RDF (Resource Description Framework) format [10] captures key information from structured data required for integration and reasoning. For example, a light sensor, an instance of *FBU* class records an observation. The change in sensor data and the recorded time are shown in the *Observation* class. Figure 2, shows a graphical representation of light intensity measured by the light sensor which is 26.8 on 29th July 2020 at 18:39:35 PM. Here, the measured value 26.8 is an instance of *SensorData* class, and timestamp is an individual of *Time* class.

We use the SPARQL Protocol and RDF Query Language (SPARQL) [21], similar to Structured Query Language (SQL) to run queries for extracting information from the knowledge graph. The extracted knowledge is used for generating feature

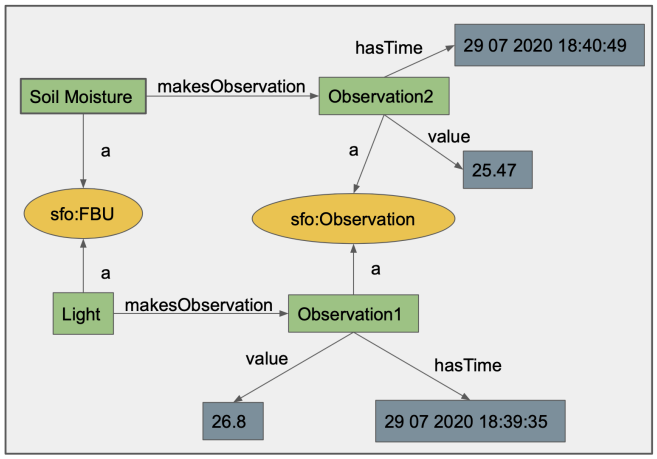

Fig. 2. Representation of knowledge graph in Resource Description Framework (RDF) for observations recorded by light sensor.

vectors which serves as an input to the anomaly detection model. For example, we can get information about measurements of soil moisture sensor on a particular day with the help of a SPARQL Protocol and RDF Query Language (SPARQL) query [21].

2) Anomaly detection: We utilize the data extracted from the knowledge graph that provides us with sensor readings for every timestamp. These readings represent the normal or abnormal behaviors of sensors deployed in a smart farm setup. Any deviation from the normal behavior of the data collected is reported as an anomaly. Our focus is to develop an anomaly detection model that continuously monitors the sensor data and alerts the farm-owner whenever an anomalous state is detected. However, anomaly detection models often find it challenging to detect anomalies when presented with high-dimensional data. Since the increase in number of features affect both the performance and accuracy of the models. To address this problem, we have chosen to use Principal Component Analysis (PCA) method to create a model.

PCA, as described by Li et al. [13] is a multivariate statistical analysis algorithm where the original data is mapped onto a lower-dimensional space by retaining most information from the original data. In our work, we perform PCA based anomaly detection by utilizing the above algorithm [13] and further calculate the covariance matrix only on the normal conditions of the dataset. To detect anomalies on the test data, we compute mahalanobis distance between the normal data and test data in principal component space. If the distance between them exceeds a certain threshold, we classify the data point as an anomaly. The threshold is determined by applying the distance function on the normal data, and its distribution. We used mahalanobis distance here instead of euclidean distance as the former includes mean values, variance and covariance of the variables.

The formula for computing mahalanobis distance (D) for normal data and it's distribution is:

$$\boldsymbol{D} = \sqrt{(\boldsymbol{x} - \boldsymbol{m})^\top \boldsymbol{T}^{-1} (\boldsymbol{x} - \boldsymbol{m})} \quad (1)$$

Here x is the observed data, mean values of the independent variables is represented by the vector m, and their covariance matrix inverse is denoted as T^{-1}.

IV. Use Case Scenario

We tested the effectiveness of our system described in Section III with multiple use cases in a real-time environment. In this section, we describe our experimental setup and steps taken to generate abnormal conditions that can happen with sensors deployed in the smart farm setup. Further, we also explain the implementation process and evaluate our model's ability to identify an abnormal condition.

A. Experimental setup

As described in Section III-A, sensors deployed in the smart farm setup generate data for every timestamp. We have integrated the data collected from sensors for 10 days with our smart farm ontology described in Section III-B to generate a knowledge graph. Then our digital twin model that supports anomaly detection continuously extracts the vital information from the knowledge graph as explained in Section III-C1, which contains only numerical values of sensors and time at which a change in observation is recorded. Further, we utilize PCA based anomaly detection model described in Section III-C2 to monitor the data generated from sensors and alert the farm owner whenever there is an anomaly observed in the operational condition of the sensors. We have chosen to use digital twin in our work to observe data generated from sensors in real-time and simulate anomalous conditions of sensors. In this way, digital twin also supports anomaly detection function by helping evaluate the model's performance during abnormal operating conditions of the sensors. An example of a simulated anomalous scenario is when soil moisture sensor had a sudden spike in its value and continued for few days caused by inducing anomalous data to the actual data in the digital twin. In this case, a sudden spike in the soil moisture sensor is considered abnormal and needs immediate attention as it could damage the crop. Another scenario is when the temperature sensor readings are varied drastically by simulating a temperature drop around the smart farm. Likewise, we have simulated anomalous sensor readings for about 900 observations out of 2281 in the test data.

B. Abnormal scenario detection

We ran the digital twin supported PCA based anomaly detection model described in Section III-C to evaluate our model. We started our experiment by extracting data from knowledge graph and normalize the obtained input features with the help of the min-max function [36]. Further, we split our data into training and test data. The training data consists of data points collected from sensors during normal conditions. This data is used to train the PCA based anomaly detection model and estimate the threshold score of acceptance. The test data contains the data points from simulated anomalous scenarios

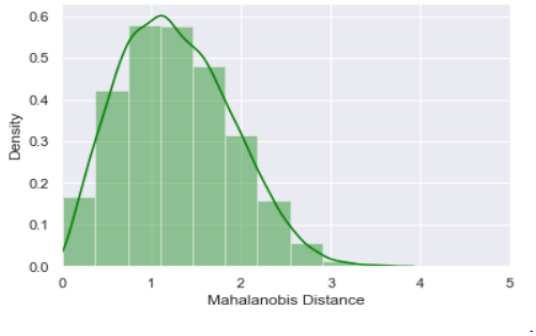

Fig. 3. Distribution of mahalanobis distance for normal observations.

described in Section IV-A. In our next step, we trained the PCA based anomaly detection model on training data where the number of principal components is 2. Simultaneously, the threshold score is set to 3.2 based on the distribution of mahalanobis distance for normal conditions of sensor data as shown in Figure 3. Any new data point whose mahalanobis distance exceeds this threshold is flagged as an anomaly. We validate our work by running our test data on the PCA based anomaly detection model in our final step. We detected 837 anomalous observations in the sensor, and simultaneously alerts were issued out of 900 simulated anomalous readings of the sensors. We also noticed that all the detected anomalous data points exceeded way above our threshold score.

V. Conclusion and Future Work

The adoption of CPS in the agriculture sector has provided significant benefits to the farmers but led to the rise of security threats and vulnerabilities. In this paper, we describe our security surveillance framework where a digital twin supported anomaly detection model addresses security problems CPS ecosystem faces in the agriculture sector. We utilize real-time data from multiple sensors deployed in a farm setting and our existing smart farm ontology to populate a knowledge graph. Our digital twin setup supports the generation, population, querying of the knowledge graph and further aids in building an anomaly detection model. Therefore, the generated knowledge graph is queried to obtain normal data as an input to train the PCA based anomaly detection model. The trained model is used to detect any deviations in data generated by the physical sensors. To verify the efficacy of our model, we also simulated abnormal scenarios in our digital twin that the sensors might measure. Our initial results show that the model was able to identify deviations in the sensors. In the future, we plan to extend our work by applying other anomaly detection models on more extensive smart farm data. We would also like to analyze more real-time anomalous use case scenarios in the smart farming ecosystem while fully utilizing the benefits of the digital twin.

Acknowledgment

The authors would like to thank Dr. Maanak Gupta for providing us with the dataset used in this work.

REFERENCES

[1] Rahul Dagar, Subhranil Som, and Sunil Kumar Khatri. Smart farming–iot in agriculture. In *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 1052–1056. IEEE, 2018.

[2] Nitu Kedarmal Choudhary, Sai Sree Laya Chukkapalli, Sudip Mittal, Maanak Gupta, Mahmoud Abdelsalam, and Anupam Joshi. Yieldpredict: A crop yield prediction framework for smart farms. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 2340–2349. IEEE, 2020.

[3] Maanak Gupta, Mahmoud Abdelsalam, Sajad Khorsandroo, and Sudip Mittal. Security and privacy in smart farming: Challenges and opportunities. *IEEE Access*, 8:34564–34584, 2020.

[4] Hichem Sedjelmaci, Fateh Guenab, Sidi-Mohammed Senouci, Hassnaa Moustafa, Jiajia Liu, and Shuai Han. Cyber security based on artificial intelligence for cyber-physical systems. *IEEE Network*, 34(3):6–7, 2020.

[5] Aritran Piplai, Sai Sree Laya Chukkapalli, and Anupam Joshi. Nattack! adversarial attacks to bypass a gan based classifier trained to detect network intrusion. In *2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*, pages 49–54. IEEE, 2020.

[6] Fei Tao, Fangyuan Sui, Ang Liu, Qinglin Qi, Meng Zhang, Boyang Song, Zirong Guo, Stephen C-Y Lu, and AYC Nee. Digital twin-driven product design framework. *International Journal of Production Research*, 57(12):3935–3953, 2019.

[7] Hongpeng Wang, Jingtai Liu, and Jianda Han. Rs-cps: A distributed architecture of robotic surveillance cyber-physical system in the nature environment. In *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 1287–1292, 2015.

[8] Sina Sontowski, Maanak Gupta, Sai Sree Laya Chukkapalli, Mahmoud Abdelsalam, Sudip Mittal, Anupam Joshi, and Ravi Sandhu. Cyber attacks on smart farming infrastructure. In *2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC)*, pages 135–143. IEEE, 2020.

[9] Sai Sree Laya Chukkapalli, Sudip Mittal, Maanak Gupta, Mahmoud Abdelsalam, Anupam Joshi, Ravi Sandhu, and Karuna Joshi. Ontologies and artificial intelligence systems for the cooperative smart farming ecosystem. *IEEE Access*, 8:164045–164064, 2020.

[10] W3. Resource Description Framework. https://www.w3.org/RDF/.

[11] Sai Sree Laya Chukkapalli, Aritran Piplai, Sudip Mittal, Maanak Gupta, and Anupam Joshi. A smart-farming ontology for attribute based access control. In *6th IEEE International Conference on Big Data Security on Cloud (BigDataSecurity 2020)*, 2020.

[12] Sai Sree Laya Chukkapalli, Shaik Barakhat Aziz, Nouran Alotaibi, Sudip Mittal, Maanak Gupta, and Mahmoud Abdelsalam. Ontology driven ai and access control systems for smart fisheries. In *Proceedings of the 2021 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems*, pages 59–68, 2021.

[13] Fan Li, Belle R Upadhyaya, and Sergio RP Perillo. Fault diagnosis of helical coil steam generator systems of an integral pressurized water reactor using optimal sensor selection. *IEEE Transactions on Nuclear Science*, 59(2):403–410, 2012.

[14] S Jagannathan, R Priyatharshini, et al. Smart farming system using sensors for agricultural task automation. In *2015 IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR)*, pages 49–53. IEEE, 2015.

[15] Sudhir Rao Rupanagudi, BS Ranjani, Prathik Nagaraj, Varsha G Bhat, and G Thippeswamy. A novel cloud computing based smart farming system for early detection of borer insects in tomatoes. In *2015 international conference on communication, information & computing technology (ICCICT)*, pages 1–6. IEEE, 2015.

[16] Monika Jhuria, Ashwani Kumar, and Rushikesh Borse. Image processing for smart farming: Detection of disease and fruit grading. In *2013 IEEE second international conference on image information processing (ICIIP-2013)*, pages 521–526. IEEE, 2013.

[17] Five use cases of Digital Twins in manufacturing. https://blogs.opentext.com/five-use-cases-for-digital-twins-in-manufacturing. [Online].

[18] Haya Elayan, Moayad Aloqaily, and Mohsen Guizani. Digital twin for intelligent context-aware iot healthcare systems. *IEEE Internet of Things Journal*, 2021.

[19] Qinghua Xu, Shaukat Ali, and Tao Yue. Digital twin-based anomaly detection in cyber-physical systems. In *2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST)*, pages 205–216. IEEE, 2021.

[20] Rafael Gomes Alves, Gilberto Souza, Rodrigo Filev Maia, Anh Lan Ho Tran, Carlos Kamienski, Juha-Pekka Soininen, Plinio Thomaz Aquino, and Fabio Lima. A digital twin for smart farming. In *2019 IEEE Global Humanitarian Technology Conference (GHTC)*, pages 1–4. IEEE, 2019.

[21] W3. Sparql query language. https://www.w3.org/TR/rdf-sparql-query/.

[22] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706, 2007.

[23] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick Van Kleef, Sören Auer, et al. Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic web*, 6(2):167–195, 2015.

[24] Andrew Carlson, Justin Betteridge, Richard C Wang, Estevam R Hruschka Jr, and Tom M Mitchell. Coupled semi-supervised learning for information extraction. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 101–110, 2010.

[25] Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web*, 8(3):489–508, 2017.

[26] Maya Rotmensch, Yoni Halpern, Abdulhakim Tlimat, Steven Horng, and David Sontag. Learning a health knowledge graph from electronic medical records. *Scientific reports*, 7(1):1–11, 2017.

[27] Longlong He and Pingyu Jiang. Manufacturing knowledge graph: a connectivism to answer production problems query with knowledge reuse. *IEEE Access*, 7:101231–101244, 2019.

[28] Sandeep Nair Narayanan, Ashwinkumar Ganesan, Karuna Joshi, Tim Oates, Anupam Joshi, and Tim Finin. Early detection of cybersecurity threats using collaborative cognition. In *2018 IEEE 4th international conference on collaboration and internet computing (CIC)*, pages 354–363. IEEE, 2018.

[29] Aditya Pingle, Aritran Piplai, Sudip Mittal, Anupam Joshi, James Holt, and Richard Zak. Relext: Relation extraction using deep learning approaches for cybersecurity knowledge graph improvement. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 879–886, 2019.

[30] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[31] Agniva Banerjee, Raka Dalal, Sudip Mittal, and Karuna Pande Joshi. Generating digital twin models using knowledge graphs for industrial production lines. In *Proceedings of the 2017 ACM on Web Science Conference*, WebSci '17, page 425–430, New York, NY, USA, 2017. Association for Computing Machinery.

[32] Yi Zeng, Meikang Qiu, Zhong Ming, and Meiqin Liu. Senior2local: A machine learning based intrusion detection method for vanets. In *International conference on smart computing and communication*, pages 417–426. Springer, 2018.

[33] Sofia Dutta, Sai Sree Laya Chukkapalli, Madhura Sulgekar, Swathi Krithivasan, Prajit Kumar Das, Anupam Joshi, et al. Context sensitive access control in smart home environments. In *6th IEEE International Conference on Big Data Security on Cloud (BigDataSecurity 2020)*, 2020.

[34] Andres Robles-Durazno, Naghmeh Moradpoor, James McWhinnie, and Gordon Russell. A supervised energy monitoring-based machine learning approach for anomaly detection in a clean water supply system. In *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pages 1–8. IEEE, 2018.

[35] Weijie Hao, Tao Yang, and Qiang Yang. Hybrid statistical-machine learning for real-time anomaly detection in industrial cyber-physical systems. *IEEE Transactions on Automation Science and Engineering*, 2021.

[36] Shivam Agarwal. Data mining: Data mining concepts and techniques. In *2013 International Conference on Machine Intelligence and Research Advancement*, pages 203–207. IEEE, 2013.