A Privacy Preserving Anomaly Detection Framework for Cooperative Smart Farming Ecosystem

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

***Abstract*—The agriculture sector has seen growing applications of AI and data intensive systems. Typically, individual farm owners join together to form agricultural cooperatives to share resources, data, and domain knowledge. These data intensive cooperatives help generate AI supported insights for member farmers. However, this leads to a rising concern among individual smart farm owners about the privacy of their data, especially while sharing the data with the co-op. In this paper, we present a framework where the individual smart farm owner's privacy is preserved, as it is shared to train robust anomaly detection models at the cooperative level. Here, we preserve the privacy of each farm owner by adding noise to their data through data perturbation techniques such as white Gaussian noise. Our experimental results show that the anomaly detection models can identify various anomalous events even when the training data is transformed with white noise. Further, we evaluate our framework and compare the detection performance on non-transformed and transformed data that belongs to multiple smart farms present in a cooperative.**

I. Introduction

Artificial Intelligence (AI) and data intensive applications have become pervasive in the agriculture sector. Smart farming integrates internet connected devices and various technologies to meet the market demands, improve crop productivity, predict yield [1], and use resources efficiently [2]. Deployed applications use the smart farm data to diagnose critical conditions of the farm such as crop diseases, soil conditions, etc., and also assist the individual farmers in tackling critical problems. In recent times, individual farmers are joining agricultural cooperatives, which are formal enterprises, financed, owned, and controlled by members for mutual benefit. These co-ops can aid the member farms by alerting them to events like, crop diseases, pest management, weather, changing labor costs, price fluctuations, etc. [3]. Other advantages of joining a co-op include resource sharing, machine use and maintenance, hiring farm labor, specialized machine operators, coordinating market visits, estimating price/purchase data, etc. [4]–[7].

Although, agricultural cooperatives are boon to member farm owners there is a major drawback in regards to data sharing. As member farm owners have to share data with cooperative ecosystem to get insights from the AI applications, many of them are concerned about their data security and privacy [8]–[10]. On a smart farm, an enormous amount of complex, dynamic and spatial data gets generated from many heterogeneous sensors, devices and equipment. Data sharing can be a risk to the individual farm owners as many of them consider specific farm information as a competitive edge over competitors. This data can also include farm specific practices and personal information about the farm owner. Any situation resulting in data leakage from the cooperative will impact the members farms causing potential economic loss. As such data security and privacy is a very important requirement and one of the primary objectives to ensure reliable operations in a smart cooperative ecosystem.

To tackle privacy concern of the smart farm owners, we designed a privacy preserving framework. Our framework incorporates white Gaussian noise technique [11] for perturbing the date generated by sensors. This technique generates additive noise to the existing sensor data without effecting the data distribution. In general, randomization techniques add noise to sensor data based on the boundary limits causing an impact on the actual data distribution. Due to this drawback, white Gaussian noise technique is also widely used among the signal processing community for adding noise to the electronic systems. Individual smart farm owners can also leverage our framework to preserve privacy of data generated from the sensors deployed on their farms. This automatically encourages farm owners to share data with the cooperative ecosystem and completely utilize the benefits provided by the cooperatives.

The effectiveness of our framework is measured by identifying simulated anomalous events present in the perturbed test data. This is done by utilizing the perturbed data present in the cooperative ecosystem to train anomaly detection model. Further we compare the performance of the anomaly detection model on transformed and non-transformed data. The results indicate that our framework identifies most of the anomalous events in the data. In terms of performance, the model trained on transformed data has slightly lower margin of difference when compared with non-transformed data.

We considered anomaly detection models in our framework since these models offers multiple insights to the farmers with the help of AI applications present in the cooperative ecosystem. For example, weed detection model identifies the presence weeds in the crop. Similarly, crop seeding, quality of crop can also be detected to alert the respective member farm owners. Overall, anomaly detection models are the most popular AI applications which have the ability to identify abnormal conditions on any given data when trained appropriately with the larger data as input. Moreover, anomaly detection models play crucial role proper functioning of the member farms present in the cooperative ecosystem.

The key contributions of this paper are:

- We have set up individual member smart farms units where multiple CPS are deployed for 10 days to capture normal and simulated abnormal events.
- We utilize cooperative agriculture ontology from our previous work [3] to perform data transformation by adding white Gaussian noise to data generated by all individual smart farms.
- We populate a knowledge graph by adding the transformed data of smart farms to the cooperative agriculture ontology in the cooperative ecosystem. The knowledge graph helps retrieve the integrated smart farm transformed data to train the anomaly detection model and further use the detection model to identify a wide range of abnormal events.
- We explain our experimental setup in detail and evaluate the performance of the anomaly detection models on both perturbed and non-perturbed retrieved from member farms that are part of cooperative smart farm ecosystem.

The rest of the paper is organized as follows: Section II presents related work. Section III explains our architecture and Section IV describes the experimental setup. Finally, Section V summarizes our work.

II. Related Work

In this section we describe related work in the areas of cooperative-based farming, smart farming ecosystems, anomaly detection models for Cyber Physical Systems (CPS), and privacy preserving technologies.

A. Cooperatives and Co-op Farming

Cooperatives (co-ops) are worker-owned enterprises, with the goal of providing mutual investment, profit, and shared technology benefits to co-owners [6]. Mutual benefits are typically defined through *membership agreements*, which include operational rules and conditions for use and distribution of shared resources [12], [13]. Cooperatives exist in several supply based industries such as produce, water, credit unions, utilities, transportation, childcare, and farming. Our research focuses on cooperative services in the agriculture industry [14].

The United States Department of Agriculture (USDA) communicates the importance of utilizing co-ops as tools for building sustainable communities. A large part of USDA's cooperative services involve providing supplies such as electricity, capital, e-connectivity, and technology to *rural* America [6], [7]. In addition to providing several beneficial services to larger, rural communities, agricultural co-ops also provide numerous benefits to farmers themselves, as well as their internal communities.

According to California Center for Cooperative Development [15], co-ops allow farmers to participate in assisted supply chain activities such as marketing and processing of produce, purchasing and borrowing agriculture equipment, as well as offloading supplies and farmer skill sets. Studies from USDA show that for every sector in the United States economy they are approximately 29,000 cooperatives [7]. In 2014, It was estimated that farmers are part of 3,000 agricultural co-ops, employing 191,000 people. Outside of the American agriculture cooperative system, there have also been observed benefits of using co-ops in supply-based farms in several other geographic regions. For example, Cameron et al. [16] discussed the role of agriculture cooperatives in Canada and Cuba which led to the strengthening of local food system.

Though cooperatives are becoming standardized among several agricultural regions, there is growing research performed to utilize Artificial Intelligence and Internet of Things (IoT) advancements to improve cooperative operations. In our previous work [3], we formalized a connected cooperative ecosystem which defines several embedded sensors and their communication with a diversity of entities in a cloud-based co-op hub. The work also describes various AI-based applications that can be deployed to aid member farmers in co-op based activities. In the next section, we provide several examples of existing smart agriculture ecosystems.

B. Smart Farm Ecosystem

The integration of cloud computing, big data, machine learning, and IoT-based tools with traditional farming environments is known as *smart farming* [17]. Smart farms incorporate a variety of sensors that provide on-premise insights such as crop condition monitoring, and operation life-cycle updates [2]. The diversity of sensors allow for optimized productivity across a multitude of farming workloads. For example, Jagannathan et al. [18] propose an automated task that leverages sensors located on large farm sprayers. The proposed system monitors existing water content present in produce soil, and sprays a calculated amount of water after initial observation. This allows farmers to quickly detect and react to changes across large farmland on a periodic basis. Another example of using equipment sensors to streamline farming tasks is a framework developed by Rupanagudi et al. [19], that identifies borer insects in tomatoes through *continuous monitoring* of the crops. This prevents food wastage by alerting farmers of signs of pesticides. Similarily, Jhuria et al. [20] propose an image processing tool to detect diseases in plants in all life-cycle phases, of intiial planting, to harvesting grown crops.

It is evident there are clear benefits of proposed smart-farming solutions to real-world agricultural ecosystems. However, these applications introduce several cybersecurity challenges as new tools, processes, and workflows are integrated. In the next section, we describe work done in the overarching security problem of anomalous CPS event detection and prevention.

C. Anomaly detection Models for Cyber-Physical Systems

Smart agriculture ecosystems rely on the use of stream *events* to provide real-time and accurate information to farmers [21]. A security flaw associated with this model is the potential of adversaries incorporating *anomalous* events to insecure event logs. Our goal is to create an anomaly detection model for agriculture-based smart ecosystems.

Several CPS ecosystems outside of the agriculture domain have considered this flaw, and as a result, developed anomaly detection models for identifying anomalous events. In the industrial and automobile industry, there are a variety of examples of using statistical based approaches for indicating anomalous events. For example, Zeng et al. [22] introduced a machine learning based intrusion detection method that detects malicious nodes in Vehicular Ad-Hoc Networks. Simiarily, Narayanan et al. [23] also study methods of detecting malicious behaviors in Vehicular Ad-Hoc Networks. The authors study vehicles in operation by collecting real-time data flowing between components of a Vehicular Ad-hoc network using a Hidden Markov Model (HMM) to detect anomalous events and alert users of malicious behavior. Lastly, Hao et al. [24] 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. In the smart home industry also incorporates similar machine-learning based methods to identify anomalous behavioral data. Ramapatruni et al. [25] propose HMM based anomaly detection models that identify anomalous events detected in smart home behavioral data. Dutta et al. [26] use similar methods to detect anomalous smart home events, but also incorporate an alert system to notify the home owner of suspicious behavior. One project related to agriculture-based smart ecosystems detects anomalous events in clean water supply systems. Robles et al. [27], create a testbed of water supply system events and identify anomalous occurrences through multiple machine learning approaches such as Support Vector Machine (SVM), K-Nearest-Neighbors (KNN), and Random Forest (RF).

These systems describe methods for identifying anomalous events and alerting users of potential malicious behavior. However, there is a lack of privacy preserving technologies in their implementations. This paper focuses on integrating privacy preserving technique to anomaly detection systems. Examples of existing privacy preserving methods are described in the next section.

D. Privacy Preserving Techniques

The goal of privacy preserving techniques is the protection or control of sharing sensitive data. There is a growing body of literature work [28]–[30] based on methodologies for preserving privacy. These methods have applications in numerous security and privacy areas such as, anonymization [31], access control [10], [26], encrytption [32], secure multi-party computation (SMC) [33]. Privacy preserving techniques can also have applications in a combination of the above methods. For example, a combination of access control and encryption based approaches to protect privacy of the data shared by utilizing attribute-based encryption is explained by Xu et al. [34]. Another example, Malina et al. [35] presented a privacy preserving technique that protects the privacy of user while giving access to utilize services offered by the cloud. The technique presented is based on advanced cryptographic components and access control.

In general, cryptographic components are known to be more secure due to their complex encryption algorithms. Gong et al. [36] designed a framework by combining differential privacy and homomorphic encryption to prevent data leakage between the central server and participants. Similarly, fully homomorphic encrytption was developed by Marcano et al. [37] to protect functionality offered by deep learning methods such as classification of images.

Though, the above approaches are known to be highly secure they are computationally expensive. Moreover, edge devices connected to the CPS have low memory. Therefore, we have chosen white Gaussian noise to transform the data generated by sensors making it suitable to run on low power devices. Another advantage of white Gaussian noise technique is that noise added to sensors doesn't impact the data distribution.

III. System Architecture

Cyber-Physical Systems (CPS) and Internet of Thing (IoT) generate large amounts of data spurring the rise of Artificial Intelligence (AI) based smart applications. Driven by rapid advancements in technologies that support smart devices, agriculture and farming sector is shifting towards IoT connected ecosystem to balance the increase in demand for food supply. However, the *privacy* of data generated is compromised and raises concerns among smart farm owners, specially when part of a co-op.

Therefore, in this paper, we describe our framework designed to protect the individual smart farms' privacy and evaluate the effectiveness of anomaly detection models at the cooperative level, by collating data from multiple smart farms. In this work, we have focused on anomaly detection models, as they can be utilized to create a plethora of AI/ML assisted applications that benefits cooperative member farmers. The co-op can monitor the quality of the crops produced by member farmers by analyzing the historic sensor values for a particular crop, and use AI to compare it with other member farms.

Figure 1, gives the overall architecture of a privacy-preserving anomaly detection framework for a cooperative ecosystem. Our framework consists of four modules that are interlinked with each other:

- *Data Collection and Transformation*: In this phase, we collect and pre-process the data obtained from smart farms that are part of the cooperative ecosystem. Further, we transform the data using white Gaussian noise technique to preserve the privacy of the individual smart farms.
- *Ontology for Cooperative Smart Farm Ecosystem*: This phase updates the existing cooperative agriculture ontology from our previous work [3]. The reason to extend our ontology is to make it suitable for achieving the goal of preserving the privacy of individual smart farm data before sharing it to the cooperative. We also provide

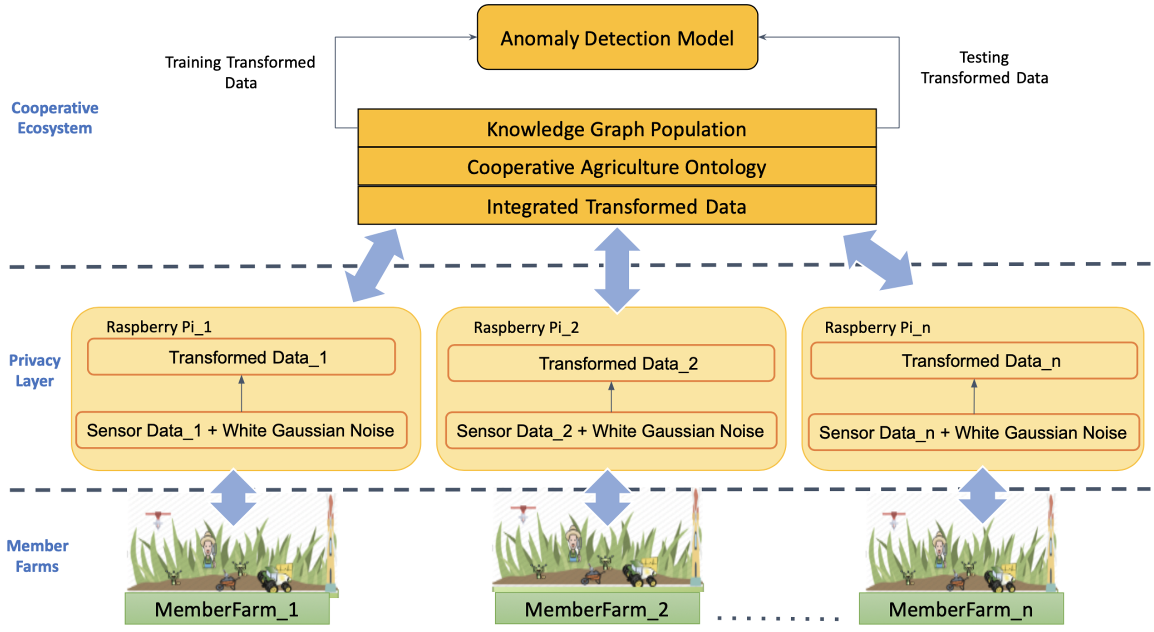


Fig. 1. Architecture of our privacy preserving anomaly detection framework.

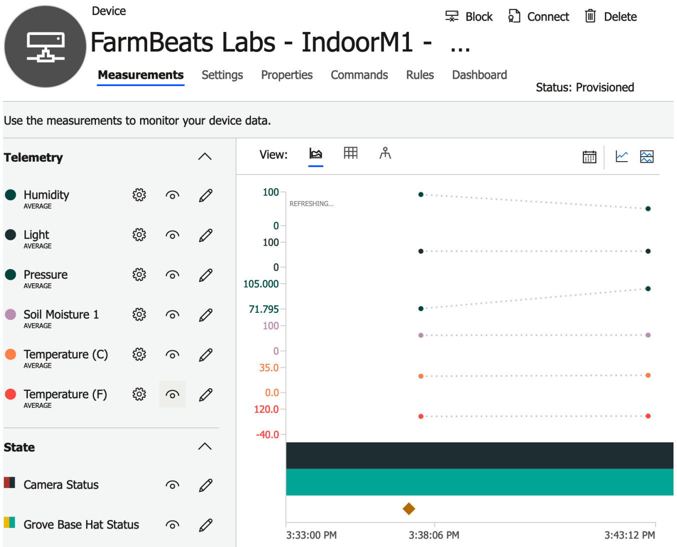
information about entities and relationships in the smart farm ecosystem.

- *Knowledge Graph Population*: In this phase, we utilize the cooperative agriculture ontology schema mentioned in Section III-B and integrate with individual smart farm transformed data III-A to populate a knowledge graph.
- *Anomaly Detection Model*: This is the final phase, where we query the knowledge graph using SPARQL Protocol and RDF Query Language (SPARQL) [38] to retrieve the required data. We utilize the transformed data from all individual farms to create the anomaly detection model at the cooperative level. These models can then be used by the cooperative to add value to the member farms.

Next, we describe these four modules in detail.

A. Data Collection and Transformation

We utilize data collected from individual smart farms by replicating the infrastructure setup described by Sontowski et al. [9]. Each smart farm infrastructure setup had devices such as grove base hat and camera. The sensor list deployed includes barometer, capacitive moisture sensor, air quality, temperature and grove light sensor. The functionality of the grove base hat is to determine whether it is connected to a Raspberry Pi or not based on the labels such as *Enabled* and *Not Enabled*. Here, a Raspberry Pi acts as the edge device that pushes all the sensor data collected to the cloud. The light intensity measurements across the surroundings are recorded by the light sensor. While moisture content present in the soil is recorded by capacitive moisture sensor and quality of air is recorded by the air quality sensor. Likewise, barometer and temperature sensors record the pressure and temperature for every timestamp. Furthermore, data collected from all the sensors is pushed to the Raspberry Pi and stored in the form of a csv file.


Device

FarmBeats Labs - IndoorM1 - ...

Status: Provisioned

Use the measurements to monitor your device data.

Fig. 2. Visualization of transformed data stored in the Microsoft Azure Cloud at the cooperative.

We also run pre-processing techniques on the integrated csv file where we drop rows that have invalid data or missing values. Transform columns with categorical labels to numeric values by applying label encoding.

In the next step, we perform data transformation to preserve privacy of the data before uploading it to the Azure cloud. The data transformation is done both on normal and simulated abnormal data. We have reconstructed most of the simulated anomalous scenarios for smart farms mentioned in our previous paper [39]. For data transformation, we have chosen to use white Gaussian noise technique [11]. This method is widely used in generating noise for electronic systems by signal processing community. As the underlying distribution

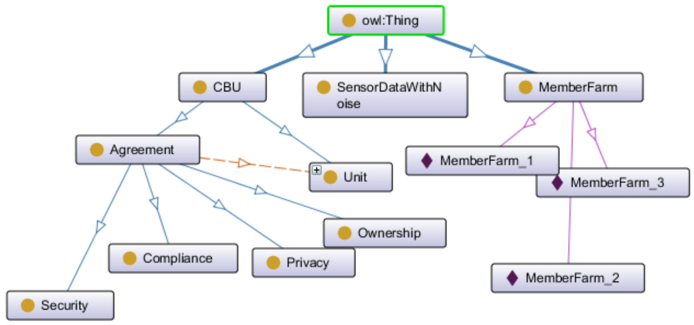


Fig. 3. Some of the classes and instances in the Cooperative Agriculture Ontology.

of data still remains intact even after adding generated white noise to the sensors.

We generate white Gaussian noise [11] in time domain with constant power spectral density (PSD) at all frequencies. Here white indicates that distribution of power is independent at all frequencies while Gaussian refers to probability distribution function for generated noise in time domain. This process of random noise generation has zero mean where every noise sample is uncorrelated with other samples present in the sequence.

The mathematical notion of data transformation process by adding white Gaussian noise is given by:

$$U(t) = w(t) + x(t)$$

In the above equation, U(t) is the transformed sensor data, readings of a actual sensor is denoted by w(t) and random sample generation of white noise for the sensor is denoted by x(t). We follow the same process of adding white noise to each sensor data column present in the csv file located in the Raspberry Pi. Later, only transformed data is pushed to the cloud shown in Figure 2. for anomaly detection described in Section III-D.

B. *Ontology for Cooperative Smart Farm Ecosystem*

We incorporate cooperative agriculture ontology created in our previous work [3] and further develop on it based on our current requirements. We develop a semantically rich ontology to assist our framework by re-using classes and relationships from the existing cooperative agriculture ontology and member farm ontology. Figure 3, provides an overview of our extended ontology and the associated relationships. In order to capture the transformed data, we had to add additional entities and relationships. Some of the important entities and their relationships utilized required for our framework are described below:

1) *Classes:*

- *CBU*: An entity that represents the centralized unit of the cooperative ecosystem. This entity has shared information about individual farm owners that are part of co-op in order to provide better insights such as prediction, detection, etc.
- *MemberFarm*: An entity that represents the individual member farm part of the cooperative ecosystem that follows the rules and functions according to the co-op. This *MemberFarm* class belongs to the member farm ontology. It monitors the actions performed on the individual smart farm and has access to all resources that belong to the farm.
- *Agreement*: An entity that describes the polices structured by the cooperative ecosystem. These policies cover aspects such as security, privacy, ownership and compliance. Any individual smart farm owner part of the cooperative ecosystem should abide by the policies in order to utilize the benefits of the co-op.
- *Observation*: This entity belongs to the *MemberFarm* ontology that provides us with recordings of the sensors deployed in the individual farm. The recordings include readings associated with *SensorData* class and timestamp associated with *Time* class.
- *Time*: An entity that provides us with temporal information for all the sensors deployed in the individual smart farm. This entity also belongs to the *MemberFarm* ontology.
- *SensorDataWithNoise*: An entity that belongs to the *MemberFarm* ontology. This entity has readings for every physical sensor added with white Gaussian noise.

2) *Relationships:*

- *hasMember*: This property provides us with information of individuals smart farms that are part of the cooperative ecosystem. Here, subject entity belongs to the *CBU* class and the object entity belongs to the *MemberFarm* class.
- *presents*: This property presents the relation between the *CBU* class and *Agreement* class where information about terms and conditions to be followed are provided for a secured cooperative ecosystem.
- *hasTransformedData*: This property has subject entity as *SensorData* and object entity as *SensorDataWithNoise* wherein values generated have noise added by utilizing white Gaussian noise technique.
- *hasTime*: This property has subject entity as *Observation* class and object entity as *Time* class. It provides us with temporal information of physical sensor.

C. *Knowledge Graph Population*

We populate a knowledge graph by utilizing Semantic Web technologies such as RDF [40] and SPARQL [38]. Knowledge graphs play an essential role in capturing the relation between entities in order to better understand the underlying information based on context. We provide domain knowledge by integrating data generated from sensors with our updated cooperative agriculture ontology. As described in the above Section III-A, readings are obtained from the CPS sensors deployed in each individual farms. The schema of the knowledge graph is based on entities and relationships described in Section III-B. A graphical representation of the knowledge graph populated based on our framework is shown in Figure 4.

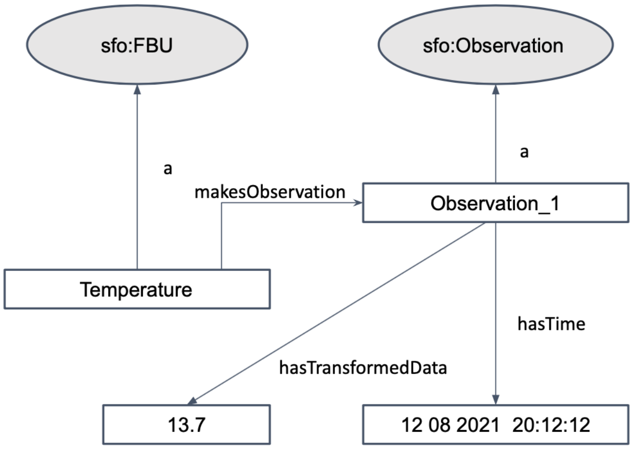


Fig. 4. A graphical view of knowledge graph populated in RDF format for reading recorded by temperature sensor.

We query the knowledge graph using SPARQL query language, a language similar to SQL in order to get the desired output. For example, if we want to query list of individual member farms present in cooperative ecosystem:

```
SELECT ?x WHERE {
    ?x :type :CBU.
    ?y :type :MemberFarm.
    ?x :hasMember ?y.
```

The above query will return the following values:

```
MemberFarm_1, MemberFarm_2,
MemberFarm_3
```

D. Anomaly Detection Model

In this work, we have focused on anomaly detection models, as they can be utilized to create a plethora of AI/ML assisted applications that benefits cooperative member farmers. The smart farm owners, part of the co-op can get alerts on the occurrence of abnormal events in the smart farm ecosystem. Hence, we create a deep learning based anomaly detection model to identify anomalous events. To build co-op level models, individual smart farm owners need to share their privacy preserved data with cooperative ecosystem to train a robust anomaly detection model.

For example, to detect anomalies within a smart farm crop productions, information of environmental temperature, humidity, soil water level and pH level can be used, and obtained from multiple sensors across a farm. A sample x_t collected from a set of sensors of a particular measurement type (e.g., soil sensors) at time t is denoted by: $x_t = (v_1, v_2, \ldots, v_m)$, where $v_{1..m}$ are the values of sensors readings. Considering different types of sensors, we represent a sample as a 2d vector, where a sample x_t collected at time t records n values of a particular measurement type and m sensor measurements.

Autoencoders [41] are unsupervised learning algorithm that aims to produce output similar to input. Multiple variations of autoencoders have been developed to address various research problems. In this paper, we have chosen to use Long Short Term Memory (LSTM) autoencoder [41] as detection model to identify anomalous events. As this detection model has the ability to identify anomalous patterns over long sequences in the time series domain. In general, LSTM autoencoder consists of an encoder and decoder in the LSTM network. The encoder compresses the input data to latent space z and the decoder decompresses the output similar to the input data with low reconstruction error. The calculation of reconstruction error is done by taking the difference between original input and the following reconstruction and the formula is shown below:

$$L = \frac{1}{2} \sum_{x} ||x - \hat{x}||^2$$

Here, x represents the input data, $\hat{x}$ represents the output and L represents the reconstruction error.

For evaluation purpose (See Section IV), we train the above LSTM autoencoder model separately on two different data settings, transformed data extracted from the knowledge graph, and non-transformed data. Here, training set contains only normal observations of sensors. Likewise, we test both the model's capability to identify simulated anomalous events present in the test data (transformed and non-transformed) based on the reconstruction error. If the reconstruction error is large, then event is considered as anomalous event and farm owners are alerted regarding the event. But if the reconstructed error is low then the even is considered as a normal event. The parameters considered for the neural network model are explained in Section IV.

This kind of a generic anomaly detection scheme can be used to create multiple AI applications that can be developed for the co-op ecosystem that benefits member farmers, by alerting them to events like, crop diseases, pest management, weather, changing labor costs, price fluctuations, etc. The co-op can utilize the shared data to create an early warning system by using various AI tools to predict a crop disease or a pest problem. For example, if a member farm has a higher use of fertilizer than peers then the co-op can alert the member farm. Most of these applications are dependent on a robust anomaly detection framework.

IV. Evaluation

We evaluate our architecture described in the above Section III. The main goal of our framework is to identify the effectiveness of anomaly detection model while preserving the privacy shared data of individual smart farms that are a part of cooperative society. In this section, we describe the experimental setup on how we train the LSTM autoencoder to detect anomalous events in transformed data and non-transformed data retrieved using knowledge graphs. Further, we present the performance of our framework in regards to identification of anomalous points in both data settings such as perturbed and non-perturbed form where the collated data belongs to member farms present in the cooperative ecosystem.

A. Experimental Setup

For this work, we utilize the collected data explained in Section III-A from three individual smart farms labelled as MemberFarm1, MemberFarm2 and MemeberFarm3. The farm owners of all three farms are part of the cooperative ecosystem where they share information and borrow equipment. The data collected was for a period of ten days and pushed to the edge device such as Raspberry Pi owned by each individual farm owner. We perform transformation of data for sensors deployed in the farm by incorporating white Gaussian noise technique described in Section III-A. The transformed data is later integrated with cooperative agriculture ontology presented in Section III-B. This integration helps us to populate a knowledge graph hosted in the cloud. Further, we query the knowledge graph using SPARQL [38] for extracting required data. The process of populating knowledge graph and example query is detailed in Section III-C.

We extract normal conditions of sensor data from the individual smart farms that has additive white Gaussian noise and copy of data that has no white Gaussian noise is also retrieved for evaluation of the experiments. We train the LSTM autoencoder model described in Section III-D on both the datasets. The model is trained for 100 epochs with a batch size of 10 and stored in the cloud owned by the cooperative ecosystem to identify anomalous events in the transformed test data and non-transformed test data in the next stage. The test data contains simulated anomalous events replicated from our previous paper [39]. The activation function used in the model is Rectifiled Linear Unit (ReLU), RMSE (Root Mean Squared Error) was considered for loss function and optimizer is set to Adaptive Moment Estimation (ADAM).

B. Results

In this section, we present the results for evaluation of our framework on two different test data settings, namely, collated smart farm data from multiple farms without perturbations and with perturbations using white Gaussian noise technique respectively. The evaluation is based on the precision, recall and F1 scores. The secondary metrics include true positive, false positive, true negative and false negative present in the confusion matrix. Precision is calculated by TP / (TP+FP) which indicates only the relevant anomalous data points. The recall determines percentage of detected anomalous points from total of actual anomalous points and is calculated by TP / (TP+FN). F1 scores is calculated by 2*((precision*recall)/(precision+recall)). The training dataset for perturbed and non-perturbed data in the cooperative ecosystem had 19,152 data points which belonged to the normal class. While test data had 8,208 data points where points that belong to anomalous class are 2,700. We flag the event as anomaly based on the reconstruction error described in Section III-D.

As shown in Table I, we see that performance of the anomaly detection model when trained on perturbed data has slightly been affected in identifying the anomalous events when compared to the performance of the model on actual dataset. These results *indicate* that cloud based services owned by the cooperative ecosystem can perform their designated computations even when the data is perturbed. To summarize the results, recall and F1 scores have seen a significant drop when test data had white Gaussian noise. However, detection model trained with perturbed data was able to identify most of the anomalous events.

Data	Precision	Recall	F-1 Score
Non-transformed Data	0.97	0.92	0.94
Transformed Data	0.96	0.89	0.87

TABLE I
PRECISION, RECALL, AND F1 SCORES FOR NON-TRANSFORMED AND TRANSFORMED DATA ACROSS THE TEST SET.

V. CONCLUSION AND FUTURE WORK

Emerging AI applications for the Cyber-Physical Systems (CPS) deployed on smart farms have helped revolutionize the agriculture sector. Moreover, the rise of cooperative agriculture ecosystems that connects the individual farm owners have proven to be beneficial. As data sharing and the use of AI applications become pervasive, the need to effectively monitor real time data generated from individual smart farm sensors, without comprising the privacy of the data generated becomes a vital problem to tackle.

In this paper, we propose a privacy preserving algorithm with a white Gaussian noise perturbation process to ensure the integrity of individual smart farm data. We collect data from deployed sensors embedded in smart farms, and push the data to edge devices such as Raspberry Pis. After data collection, we transform the data stored in the edge devices by utilizing perturbation techniques such as white Gaussian noise. We selected white Gaussian noise to transform the sensor-generated data due to its sustainability to run on low power devices, as well as its ability to add noise to the dataset without altering the data distribution of the sensors. Transformed sensor data allows anomaly detection models to more easily identify anomalous events compared to the original (non-transformed) sensor data. Once the data is transformed, we are able to integrate it with the extended cooperative agriculture ontology. This allows us to query the knowledge graph for inputs to train an LSTM-based auto-encoder detection model. By using the transformed data as training input, the anomaly detection model was able to identify a majority of the anomalous events in comparison to using the original sensor data as training input.

In our ongoing work, we are exploring additional data perturbation techniques such as k-anonymity, L diversity, etc. We are also further evaluating the effectiveness of our framework by utilizing advanced anomaly detection models.

REFERENCES

[1] 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.

[2] 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.

[3] 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.

[4] Agrivi LTD. Cooperative management software. https://www.agriculture-xprt.com/software/cooperative-management-software-525842, 2020.

[5] ADAX COOP. The erp dedicated to agricultural cooperatives. https://www.adax-erp.com/adax-coop-agricultural/, 2020.

[6] Directory of Rural Farmer, Rancher, and Fishery Cooperatives. https://ageconsearch.umn.edu/record/280616?ln=en, 2017. [Online].

[7] USDA. AGRICULTURAL COOPERATIVE STATISTICS 2017. https://www.rd.usda.gov/files/publications/SR81_CooperativeStatistics2018.pdf, 2017. [Online].

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

[9] 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.

[10] 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.

[11] Friedrich K Jondral. White gaussian noise–models for engineers. *Frequenz*, 72(5-6):293–299, 2018.

[12] Co-op Bylaws and Other Governance Documents. https://www.co-oplaw.org/legal-tools/cooperative-bylaws/.

[13] Cooperative Marketing Agreements. https://cccd.coop/sites/default/files/resources/Marketing-Agreement-USDA.pdf.

[14] Tennessee Farm Bureau. Cooperatives provide billion-dollar boost to state's economy. https://www.tnfarmbureau.org/cooperatives-provide-billion-dollar-boost-states-economy, 2017. [Online].

[15] Agricultural Co-ops. https://www.cccd.coop/co-op-info/co-op-types/agricultural-co-ops.

[16] Greg Cameron, Francisco Rogelio Pérez Rosado, and Dayni Deysi Díaz Mederos. Agricultural co-operatives in canada and cuba: trends, prospects and ways forward. *Environment, Development and Sustainability*, 22(2):643–660, 2020.

[17] Smart Farming—Automated and Connected Agriculture. https://www.engineering.com/DesignerEdge/DesignerEdgeArticles/ArticleID/16653/Smart-FarmingAutomated-and-Connected-Agriculture.aspx. [Online].

[18] 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.

[19] 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.

[20] 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.

[21] Daniel Lopez, Maria Uribe, Claudia Santiago, Andrés Torres, Nicolas Guataquira, Stefany Castro, Pantaleone Nespoli, and Felix Gomez Marmol. Shielding iot against cyber-attacks: An event-based approach using siem. *Wireless Communications and Mobile Computing*, 2018, 10 2018.

[22] 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.

[23] Sandeep Nair Narayanan, Sudip Mittal, and Anupam Joshi. Obd_securealert: An anomaly detection system for vehicles. In *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 1–6. IEEE, 2016.

[24] 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.

[25] Sowmya Ramapatruni, Sandeep Nair Narayanan, Sudip Mittal, Anupam Joshi, and Karuna Joshi. Anomaly detection models for smart home security. In *2019 IEEE 5th 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 19–24. IEEE, 2019.

[26] 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.

[27] 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.

[28] Hillol Kargupta, Souptik Datta, Qi Wang, and Krishnamoorthy Sivakumar. On the privacy preserving properties of random data perturbation techniques. In *Third IEEE international conference on data mining*, pages 99–106. IEEE, 2003.

[29] C Nalini and AR Arunachalam. A study on privacy preserving techniques in big data analytics. *International Journal of Pure and Applied Mathematics*, 116(10):281–286, 2017.

[30] Amine Boulemtafes, Abdelouahid Derhab, and Yacine Challal. A review of privacy-preserving techniques for deep learning. *Neurocomputing*, 384:21–45, 2020.

[31] Manolis Terrovitis, Nikos Mamoulis, and Panos Kalnis. Privacy-preserving anonymization of set-valued data. *Proceedings of the VLDB Endowment*, 1(1):115–125, 2008.

[32] Jinguang Han, Willy Susilo, Yi Mu, and Jun Yan. Privacy-preserving decentralized key-policy attribute-based encryption. *IEEE Transactions on Parallel and Distributed Systems*, 23(11):2150–2162, 2012.

[33] Alfredo Cuzzocrea and Elisa Bertino. A secure multiparty computation privacy preserving olap framework over distributed xml data. In *Proceedings of the 2010 ACM symposium on applied computing*, pages 1666–1673, 2010.

[34] 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.

[35] Lukas Malina and Jan Hajny. Efficient security solution for privacy-preserving cloud services. In *2013 36th International Conference on Telecommunications and Signal Processing (TSP)*, pages 23–27. IEEE, 2013.

[36] Maoguo Gong, Jialun Feng, and Yu Xie. Privacy-enhanced multi-party deep learning. *Neural Networks*, 121:484–496, 2020.

[37] Néstor J Hernández Marcano, Mads Moller, Soren Hansen, and Rune Hylsberg Jacobsen. On fully homomorphic encryption for privacy-preserving deep learning. In *2019 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2019.

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

[39] Sai Sree Laya Chukkapalli, Nisha Pillai, Sudip Mittal, and Anupam Joshi. Cyber-physical system security surveillance using knowledge graph based digital twins-a smart farming usecase. In *2021 IEEE Intelligence and Security Informatics (ISI)*. IEEE, 2021.

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

[41] Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. *arXiv preprint arXiv:1607.00148*, 2016.