Securing Smart Sensing Production System using Deep Neural Network Model

V. Chandini[1], U. Mounika[1], V. Akshaya[1], K. Aarathi[2]

[1]UG Student, [2]Assistant Professor, [1,2]Department of Computer Science and Engineering

[1,2]Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Secunderabad, Telangana, India

ABSTRACT

Internet of Things (IoT) enabled cyber physical systems such as Industrial equipment's and operational IT to send and receive data over internet. This equipment's will have sensors to sense equipment condition and report to centralized server using internet connection. Sometime some malicious users may attack or hack such sensors and then alter their data and this false data will be report to centralized server and false action will be taken. Due to false data many countries equipment and production system got failed and many algorithms was developed to detect attack, but all these algorithms suffer from data imbalance (one class my contains huge records (for example NORMAL records and other class like attack may contains few records which lead to imbalance problem and detection algorithms may failed to predict accurately). To deal with data imbalance, existing algorithms were using OVER and UNDER sampling which will generate new records for FEWER class only. To overcome from this issue, we are introducing novel technique without using any under or oversampling algorithms. The proposed technique consists of 2 parts which includes auto encoder and deep neural networks (DNNs).

Keywords: Internet of things, smart production system, sampling, auto encoder, deep neural network.

1. INTRODUCTION

Sensors are most commonly used in numerous applications ranging from body-parameters' measurement to automated driving. Moreover, sensors play a key role in performing detection- and vision-related tasks in all the modern applications of science, engineering and technology where the computer vision is dominating. An interesting emerging domain that employs the smart sensors is the Internet of Things (IoT) dealing with wireless networks and sensors distributed to sense data in real time and producing specific outcomes of interest through suitable processing. In IoT-based devices, sensors, and artificial intelligence (AI) are the most important elements which make these devices sensible and intelligent. In fact, due to the role of AI, the sensors act as smart sensors and find an efficient usage for a variety of applications, such as general environmental monitoring [1]; monitoring a certain number of environmental factors; weather forecasting; satellite imaging and its use; remote sensing-based applications; hazard events' monitoring such as landslide detection; self-driving cars; healthcare and so on. In reference to this latter sector, recently the usage of smart devices has been hugely increased in hospitals and diagnostic centers for evaluating and monitoring various health conditions of affected patients, remotely as well as physically [2].

Practically, there is no field of science or research which performs smartly without using the modern sensors. The wide usage and need of sensors; and IoT employed in remote sensing, environment and human health monitoring make the applications as intelligent. In the last decade, the agriculture applications have also included [3] the utilization of many types of sensors for monitoring and controlling various types of environmental parameters such as temperature, humidity, soil quality, pollution, air quality, water contamination, radiation, etc. This paper also aims to highlight the use of

the sensors and IoT for remote sensing and agriculture applications in terms of extensive discussion and review.

In recent years, SHM of civil structures has been a critical topic for research. SHM helps to detect the damage of a structure, and it also provides early caution of a structure that is not in a safe condition for usage. Civil infrastructure like [4] bridges get damaged with time, and the reason for the damage is heavy vehicles, loading environmental changes, and dynamic forces such as seismic. These types of changes mainly occur at existing structures constructed long ago, and various methods will detect that damage. The strategy of SHM involves observing the structure for a certain period to notice the condition of the structure and the periodic measurements of data will be collected, and the features of data will be extracted from these computation results, and the process of analysis can be done with the help of a featured data to find out the present-day health of the structure. The information collected from the process can be updated periodically to monitor the structure and based on the data collected through monitoring a structure, and the structure can be strengthened and repaired, and rehabilitation and maintenance can be completed [5].

2. LITERATURE SURVEY

Ullo et. al [6] focused on an extensive study of the advances in smart sensors and IoT, employed in remote sensing and agriculture applications such as the assessment of weather conditions and soil quality; the crop monitoring; the use of robots for harvesting and weeding; the employment of drones. The emphasis has been given to specific types of sensors and sensor technologies by presenting an extensive study, review, comparison, and recommendation for advancements in IoT that would help researchers, agriculturists, remote sensing scientists and policy makers in their research and implementations. Sivasuriyan et. al [7] provides a detailed understanding of bridge monitoring, and it focuses on sensors utilized and all kinds of damage detection (strain, displacement, acceleration, and temperature) according to bridge nature (scour, suspender failure, disconnection of bolt and cables, etc.) and environmental degradation under static and dynamic loading. This paper presents information about various methods, approaches, case studies, advanced technologies, real-time experiments, stimulated models, data acquisition, and predictive analysis. Future scope and research also discussed the implementation of SHM in bridges. The main aim of this research is to assist researchers in better understanding the monitoring mechanism in bridges.

Dazhe Zhao et. al [8] proposed an easy-fabricated and compact untethered triboelectric patch with Polytetrafluoroethylene (PTFE) as triboelectric layer and human body as conductor. We find that the conductive characteristic of human body has negligible influence on the outputs, and the untethered triboelectric patch has good output ability and robustness. The proposed untethered triboelectric patches can work as sensor patches and energy harvester patches. Three typical applications are demonstrated, which are machine learning assisted objects distinguishing with accuracy up to 93.09–94.91 %, wireless communication for sending typical words to a cell phone, and human motions energy harvesting for directly powering electronics or charging an energy storage device. Bacco et. al [9] described, both analytically and empirically, a real testbed implementing IEEE 802.15.4-based communications between an UAV and fixed ground sensors. In our scenario, we found that aerial mobility limits the actual IEEE 802.15.4 transmission range among the UAV and the ground nodes to approximately 1/3 of the nominal one. We also provide considerations to design the deployment of sensors in precision agriculture scenarios.

Verma et. al [10] discussed the existing state-of-the-art practices of improved intelligent features, controlling parameters and Internet of things (IoT) infrastructure required for smart building. The

main focus is on sensing, controlling the IoT infrastructure which enables the cloud clients to use a virtual sensing infrastructure using communication protocols. The following are some of the intelligent features that usually make building smart such as privacy and security, network architecture, health services, sensors for sensing, safety, and overall management in smart buildings. As we know, the Internet of Things (IoT) describes the ability to connect and control the appliances through the network in smart buildings. The development of sensing technology, control techniques, and IoT infrastructure give rise to a smart building more efficient. Therefore, the new and problematic innovation of smart buildings in the context of IoT is to a great extent and scattered. The conducted review organized in a scientific manner for future research direction which presents the existing challenges, and drawbacks. Hu et. al [11] presented a real-time, fine-grained, and power-efficient air quality monitor system based on aerial and ground sensing. The architecture of this system consists of the sensing layer to collect data, the transmission layer to enable bidirectional communications, the processing layer to analyze and process the data, and the presentation layer to provide a graphic interface for users. Three major techniques are investigated in our implementation for data processing, deployment strategy, and power control. For data processing, spatial fitting and short-term prediction are performed to eliminate the influences of incomplete measurement and the latency of data uploading. The deployment strategies of ground sensing and aerial sensing are investigated to improve the quality of the collected data. Power control is further considered to balance between power consumption and data accuracy. Our implementation has been deployed in Peking University and Xidian University since February 2018, and has collected almost 100,000 effective values thus far. Famila et. al [12] proposed an Improved Artificial Bee colony optimization based ClusTering(IABCOCT) algorithm by utilizing the merits of Grenade Explosion Method (GEM) and Cauchy Operator. This incorporation of GEM and Cauchy operator prevents the Artificial Bee Colony (ABC) algorithm from stuck into local optima and improves the convergence rate. The benefits of GEM and Cauchy operator are embedded into the Onlooker Bee and scout bee phase for phenomenal improvement in the degree of exploitation and exploration during the process of CH selection. The simulation results reported that the IABCOCT algorithm outperforms the state of art methods like Hierarchical Clustering-based CH Election (HCCHE), Enhanced Particle Swarm Optimization Technique (EPSOCT) and Competitive Clustering Technique (CCT) in-terms of different measures such as throughput, packet loss, delay, energy consumption and network lifetime.

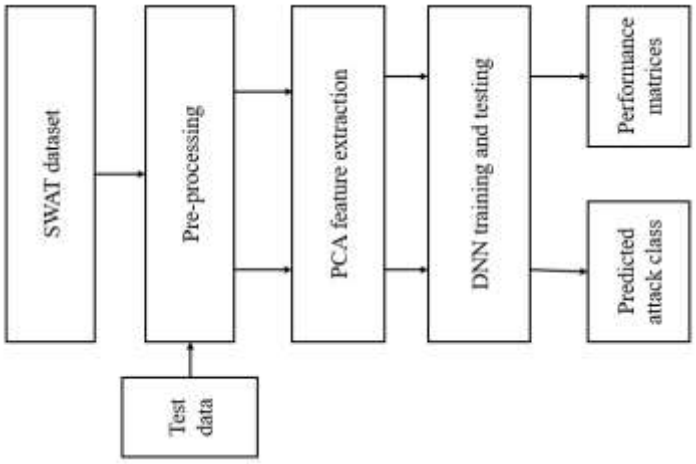


Fig. 1: Block diagram of proposed system.

3. PROPOSED SYSTEM

Data Preprocessing in Machine learning

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

PCA feature reduction

The Principal Component Analysis is a popular unsupervised learning technique for reducing the dimensionality of data. It increases interpretability yet, at the same time, it minimizes information loss. It helps to find the most significant features in a dataset and makes the data easy for plotting in 2D and 3D. PCA helps in finding a sequence of linear combinations of variables.

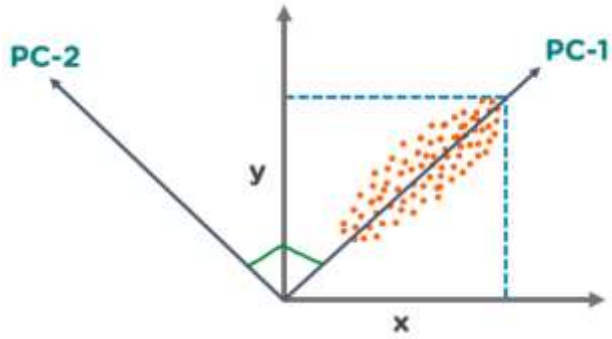


Fig. 2: PCA analysis.

In the above figure, we have several points plotted on a 2-D plane. There are two principal components. PC1 is the primary principal component that explains the maximum variance in the data. PC2 is another principal component that is orthogonal to PC1.

Fig. 3: Applications of PCA in Machine Learning.

- PCA is used to visualize multidimensional data.
- It is used to reduce the number of dimensions in healthcare data.
- PCA can help resize an image.
- It can be used in finance to analyze stock data and forecast returns.
- PCA helps to find patterns in the high-dimensional datasets.

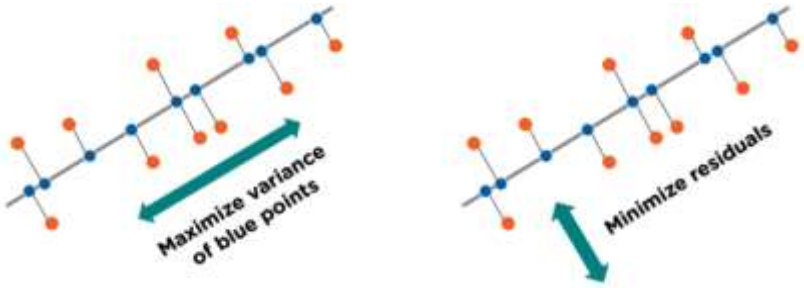


Fig. 4: PCA working.

Step 1: Normalize the data: Standardize the data before performing PCA. This will ensure that each feature has a mean = 0 and variance = 1.

$$Z = \frac{x - \mu}{\sigma}$$

Step 2: Build the covariance matrix: Construct a square matrix to express the correlation between two or more features in a multidimensional dataset.

Step 3: Find the Eigenvectors and Eigenvalues: Calculate the eigenvectors/unit vectors and eigenvalues. Eigenvalues are scalars by which we multiply the eigenvector of the covariance matrix.

Step 4: Sort the eigenvectors in highest to lowest order and select the number of principal components.

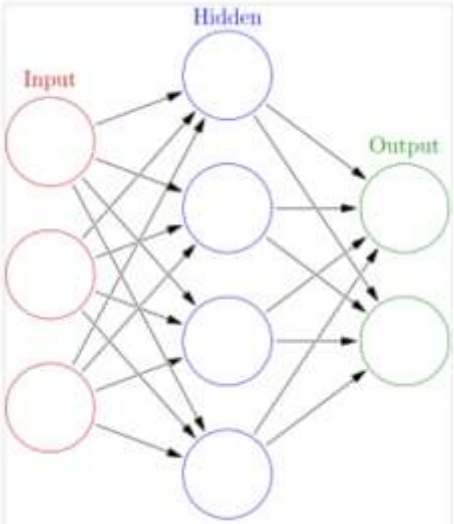
What is a deep neural network?

At its simplest, a neural network with some level of complexity, usually at least two layers, qualifies as a deep neural network (DNN), or deep net for short. Deep nets process data in complex ways by employing sophisticated math modeling.

To truly understand deep neural networks, however, it's best to see it as an evolution. A few items had to be built before deep nets existed.

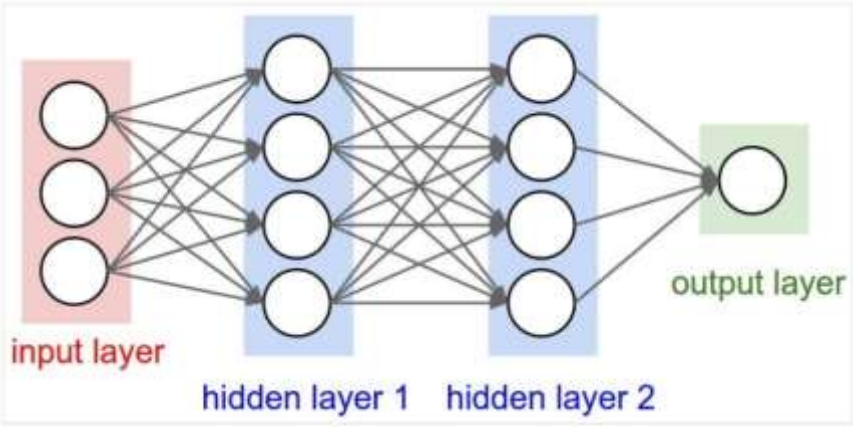
First, machine learning had to get developed. ML is a framework to automate (through algorithms) statistical models, like a linear regression model, to get better at making predictions. A model is a single model that makes predictions about something. Those predictions are made with some accuracy. A model that learns—machine learning—takes all its bad predictions and tweaks the weights inside the model to create a model that makes fewer mistakes.

The learning portion of creating models spawned the development of artificial neural networks. ANNs utilize the hidden layer as a place to store and evaluate how significant one of the inputs is to the output. The hidden layer stores information regarding the input's importance, and it also makes associations between the importance of combinations of inputs.



Deep neural nets, then, capitalize on the ANN component. They say, if that works so well at improving a model—because each node in the hidden layer makes both associations and grades importance of the input to determining the output—then why not stack more and more of these upon each other and benefit even more from the hidden layer?

So, the deep net has multiple hidden layers. 'Deep' refers to a model's layers being multiple layers deep.



4. RESULTS AND DISCUSSION

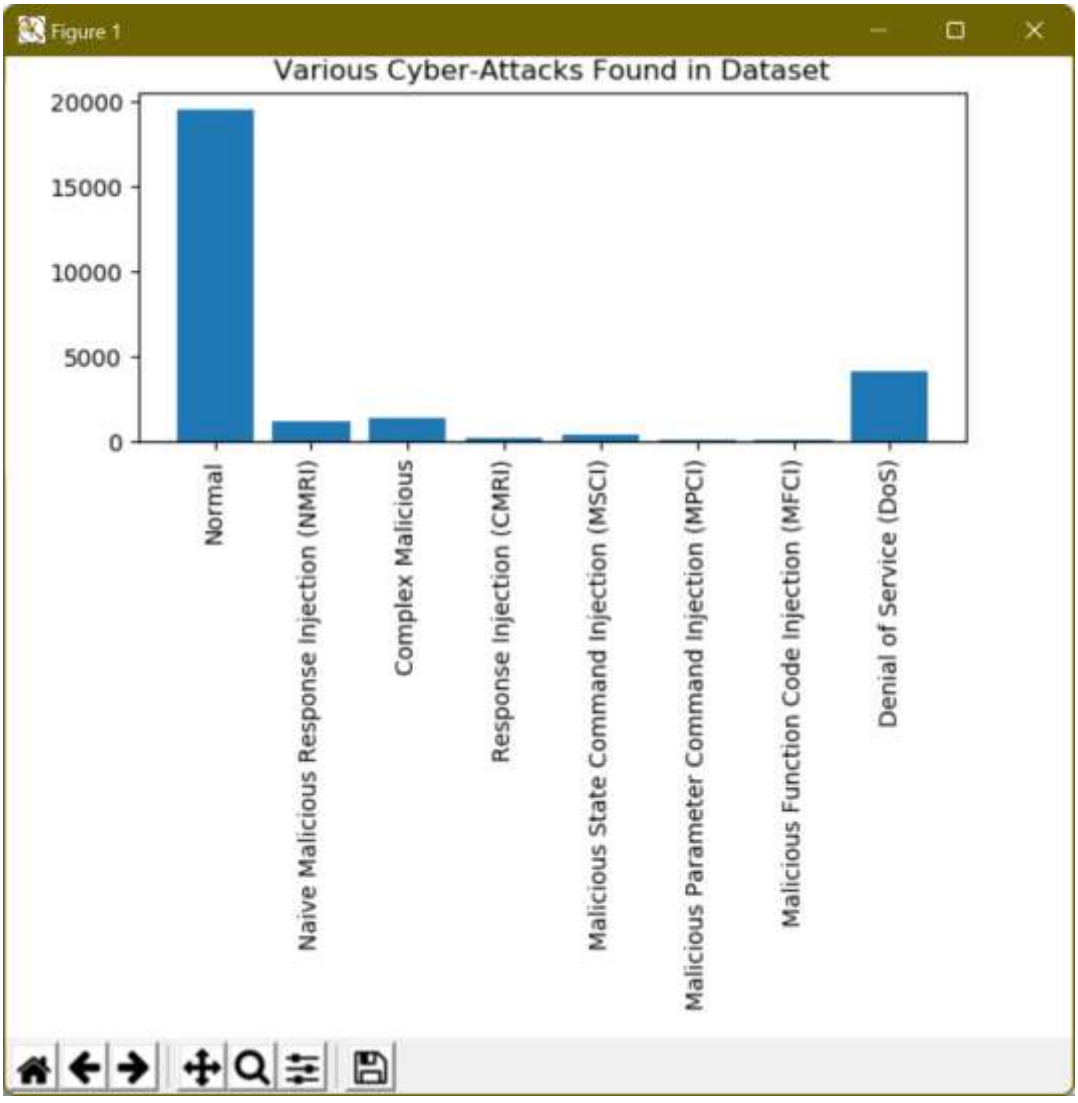
To implement this project, we have used SWAT (secure water production treatment) and this dataset contains IOT request and response signature and associate each dataset with unique attack label and dataset contains below cyber-attack labels

'Normal', 'Naive Malicious Response Injection (NMRI)', 'Complex Malicious', 'Response Injection (CMRI)', 'Malicious State Command Injection (MSCI)', 'Malicious Parameter Command Injection (MPCI)', 'Malicious Function Code Injection (MFCI)', 'Denial of Service (DoS)'

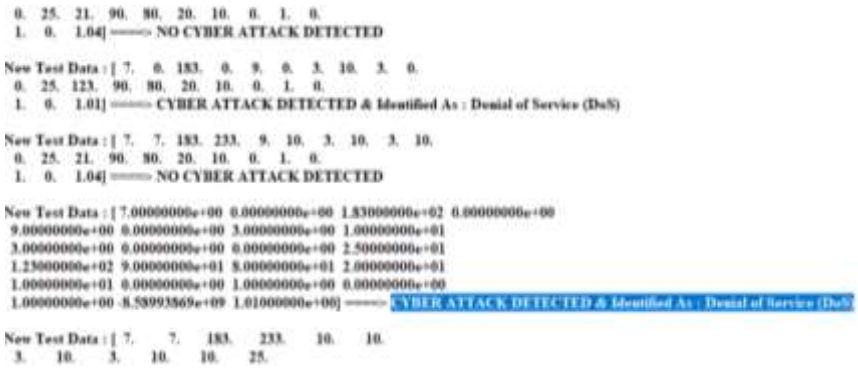
Above are the attacks found in dataset and dataset contains above labels as integer value of its index for example NORMAL label index will be 0 and continues up to 8 class labels. Below screen showing dataset details.

To implement this project, we have designed following modules.

- Upload SWAT Water Dataset: using this module we will upload dataset to application and then read dataset and then find different attacks found in dataset
- Pre-process Dataset: using this module we will replace all missing values with 0 and then apply MIN-MAX scaling algorithm to normalized features values and then split dataset into train and test where application used 80% dataset for training and 20% for testing
- AutoEncoder: using this module we will trained AutoEncoder deep learning algorithm and then extract features from that model.
- Decision Tree with PCA: extracted features from AutoEncoder will get transform using PCA to reduce features size and then retrain with Decision tree. Decision tree will predict label for each record based on dataset signatures
- Proposed DNN model: predicted decision tree label will further train with DNN (deep neural network) algorithm to detect and attribute attacks
- Detection & Attribute Attack Type: using this module we will upload unknown or un-label TEST DATA and then DNN will predict attack type
- Performance Comparison: using this module we will display comparison table of all algorithms which contains metrics like accuracy, precision, recall and FSCORE.



In above screen dataset loaded and in graph x-axis contains ATTACK NAME and y-axis contains count of those attacks found in dataset and we can see 'NORMAL' class contains so many records and other attacks contains very few records so it will raise data imbalance problem which can be solved using AutoEncoder, Decision Tree and DNN. Now close above graph and then click on 'Preprocess Dataset' button to remove missing values and then normalized values with MIN-MAX algorithm.



In above screen in square bracket, we can see TEST data values and after arrow =➔ symbol we can see detected ATTACK TYPE and scroll down above text area to view all detection. In above screen we can see detected various attacks and now click on 'Performance Comparison' to get below comparison table of all algorithms

Algorithm Name	Accuracy	Precision	Recall	FSCORE
AutoEncoder	90.56985294117646	73.26786088413995	74.27556818181817	73.72553080939959
Decision Tree with PCA	90.60661764705881	73.54066712360063	74.72222222222223	74.0846506467183
DNN	99.98161764705881	85.25345622119815	85.71428571428571	85.48009367681499

In above table we can see algorithm names and its metrics values such as accuracy and precision and other.

5. CONCLUSION

Internet of Things enabled cyber physical systems such as Industrial equipment's and operational IT to send and receive data over internet. This equipment's will have sensors to sense equipment condition and report to centralized server using internet connection. Sometime some malicious users may attack or hack such sensors and then alter their data and this false data will be report to centralized server and false action will be taken. Due to false data many countries equipment and production system got failed and many algorithms was developed to detect attack, but all these algorithms suffer from data imbalance (one class my contains huge records (for example NORMAL records and other class like attack may contains few records which lead to imbalance problem and detection algorithms may failed to predict accurately). To deal with data imbalance existing algorithms were using OVER and UNDER sampling which will generate new records for FEWER class, but this technique improve accuracy but not up to the mark. Therefore, to overcome from this issue, this project introduced an efficient deep learning model without using any under or oversampling algorithms with the usage of auto encoder, decision tree with PCA, and DNN for identifying the attack and classify the type of attack. In addition, the performance evaluation of three models also compared and proven that proposed DNN obtained enhanced accuracy 99.98%.

REFERENCES

[1] Kayad, A.; Paraforos, D.; Marinello, F.; Fountas, S. Latest advances in sensor applications in agriculture. Agriculture 2020, 10, 362.
[2] Elahi, H.; Munir, K.; Eugeni, M.; Atek, S.; Gaudenzi, P. Energy harvesting towards self-powered IoT devices. Energies 2020, 13, 5528.
[3] Ullo, S.L.; Sinha, G.R. Advances in smart environment monitoring systems using IoT and sensors. Sensors 2020, 20, 3113.
[4] Carminati, M.; Sinha, G.R.; Mohdiwale, S.; Ullo, S.L. Miniaturized pervasive sensors for indoor health monitoring in smart cities. Smart Cities 2021, 4, 146–155.
[5] Ullo, S.L.; Addabbo, P.; Di Martire, D.; Sica, S.; Fiscante, N.; Cicala, L.; Angelino, C.V. Application of DInSAR technique to high coherence Sentinel-1 images for dam monitoring and result validation through in situ measurements. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 2019, 12, 875–890.
[6] Ullo, S.L. and Sinha, G.R., 2021. Advances in IoT and smart sensors for remote sensing and agriculture applications. Remote Sensing, 13(13), p.2585.
[7] Sivasuriyan, A., Vijayan, D.S., LeemaRose, A., Revathy, J., Gayathri Monicka, S., Adithya, U.R. and Jebasingh Daniel, J., 2021. Development of smart sensing technology approaches in structural health monitoring of bridge structures. Advances in Materials Science and Engineering, 2021.
[8] Dazhe Zhao, Kaijun Zhang, Yan Meng, Zhaoyang Li, Yucong Pi, Yujun Shi, Jiacheng You, Renkun Wang, Ziyi Dai, Bingpu Zhou, Junwen Zhong, Untethered triboelectric patch for wearable smart sensing and energy harvesting, Nano Energy, Volume 100, 2022, 107500, ISSN 2211-2855, https://doi.org/10.1016/j.nanoen.2022.107500.
[9] M. Bacco, A. Berton, A. Gotta and L. Caviglione, "IEEE 802.15.4 Air-Ground UAV Communications in Smart Farming Scenarios," in IEEE Communications Letters, vol. 22, no. 9, pp. 1910-1913, Sept. 2018, doi: 10.1109/LCOMM.2018.2855211.
[10] A. Verma, S. Prakash, V. Srivastava, A. Kumar and S. C. Mukhopadhyay, "Sensing, Controlling, and IoT Infrastructure in Smart Building: A Review," in IEEE Sensors Journal, vol. 19, no. 20, pp. 9036-9046, 15 Oct.15, 2019, doi: 10.1109/JSEN.2019.2922409.
[11] Z. Hu, Z. Bai, Y. Yang, Z. Zheng, K. Bian and L. Song, "UAV Aided Aerial-Ground IoT for Air Quality Sensing in Smart City: Architecture, Technologies, and Implementation," in IEEE Network, vol. 33, no. 2, pp. 14-22, March/April 2019, doi: 10.1109/MNET.2019.1800214.
[12] Famila, S., Jawahar, A., Sariga, A. et al. Improved artificial bee colony optimization- based clustering algorithm for SMART sensor environments. Peer-to-Peer Netw. Appl. 13, 1071–1079 (2020). https://doi.org/10.1007/s12083-019-00805-4