IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A survey of Anomaly Detection in IoT Networks Using Machine Learning Techniques in Various Sectors

[1]Prof. Jyotsna Nanajkar, [2]Mr. Shubham Thorat, [3]Mr. Gaurav Borse, [4]Miss. Urmila Naravade, [5]Miss. Vaishnavi Pratale

[1]IT Department, ZCOER, Pune, [2]IT Department, ZCOER, Pune, [3]IT Department, ZCOER, Pune, [4]IT Department, ZCOER, Pune, [5]IT Department, ZCOER, Pune

Abstract: To ensure system stability and security, the rapid growth of Internet of Things (IoT) networks has prompted the development of effective anomaly detection systems. In this study, we give a thorough investigation of anomaly detection in IoT networks, with a particular emphasis on the incorporation of machine learning approaches. While bridging the gap between technical difficulties and popular comprehension, we explore into several machine learning approaches such as supervised, semi-supervised, and unsupervised learning. Our goal is to encourage more technological study and appreciation in the ever-changing IoT ecosystem. We also discuss the practical impacts of anomaly detection in sectors like the field of cyber security, healthcare, and industrial applications. As IoT systems evolve, so will the necessity for effective anomaly detection solutions.

***Index Terms* - Anomaly detection, machine learning, security and privacy protection, Internet of Things (IoT), smart devices.**

I. Introduction

Recent advances in the industry contribute to the creation of intelligent smart cities [1], smart devices [2], smart homes [3], smart transportation [4], healthcare [5], agriculture [6], smart grid [7], military [8] and much more [9]. As the number of linked devices grows around the world, various sensors are utilized to collect real-time data from physical things remotely [10]. This information assists us in developing sophisticated decision-making algorithms and properly managing IoT settings. Simultaneously, the widespread use of real-world gadgets increases the potential of cyber security risks [11]. Traditional intrusion detection technologies do not provide guaranteed security in IoT applications because of their limited bandwidth capacity and global connectivity [12,13,14]. As a result, a sophisticated Intrusion Detection System (IDS) is built to defend IoT devices from intrusions [15].

Anomaly detection in IoT (Internet of Things) is a technique that helps identify unusual patterns or events in the data generated by IoT devices. This technique is crucial for ensuring the reliability and security of IoT systems. The data collected by IoT devices can be used for various purposes, such as monitoring the environment, tracking assets, and detecting anomalies in the system.

Anomaly detection techniques in IoT involve different approaches, such as statistical models, machine learning algorithms, and rule-based systems. Statistical models use probability distributions to identify unusual patterns in the data. Machine learning algorithms use supervised or unsupervised learning techniques to identify anomalies in the data. Rule-based systems use predefined rules to detect anomalies based on specific conditions.

One of the challenges of anomaly detection in IoT is the large volume of data generated by IoT devices. This requires efficient data processing and storage techniques to handle the data. Another challenge is the need to develop robust and accurate anomaly detection models that can adapt to the changing conditions of the IoT environment.

Overall, anomaly detection in IoT is an essential technique for ensuring the reliability and security of IoT systems. It has numerous applications in various fields, such as healthcare, transportation, and manufacturing. As IoT systems continue to grow and evolve, the need for effective anomaly detection techniques will only increase.

II. DIFFERENT SECTORS OF AN IOT-BASED NETWORK WHERE ANOMALY DETECTION IS REQUIRED [16]

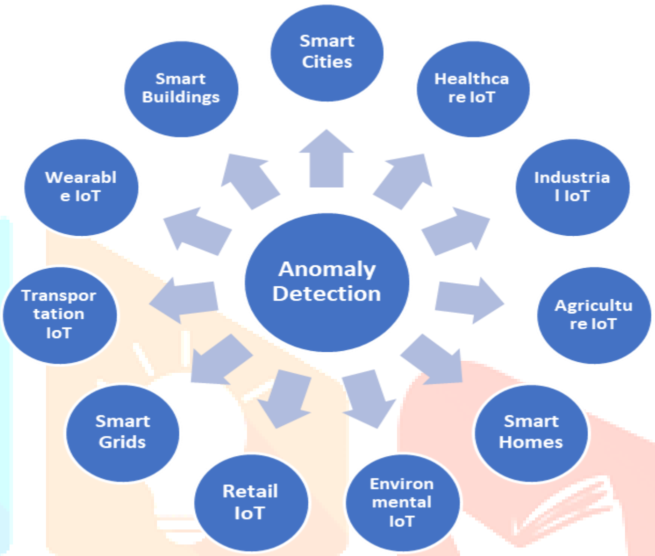

Fig.1: Applications of anomaly detection source: Refer [16]

III. OBJECTIVE

1) Conduct a comprehensive exploration of anomaly detection in IoT networks, with a particular focus on the integration of machine learning techniques.
2) Examine recent research findings, real-world applications, and the challenges faced in implementing these techniques.
3) Cover a range of machine learning methods, including supervised, semi-supervised, and unsupervised learning, while using accessible language to bridge the gap between technical complexities and general comprehension.
4) Highlight practical implications in various domains, such as cyber security, healthcare, and industrial applications, and inspire further exploration and appreciation of technology in the ever-evolving IoT landscape.

IV. COMMON ATTACKS ON IOT NETWORKS

1) Denial of Service (DoS) and Distributed Denial of Service (DDoS) Attacks: Overwhelm IoT devices or networks with a flood of traffic, rendering them unavailable.
2) Botnet Attacks: IoT devices are compromised and used to form botnets for various malicious purposes, including DDoS attacks.
3) Malware Infections: IoT devices can be infected with malware, such as viruses, worms, or ransomware, affecting their operation and potentially spreading to other devices.
4) Physical Attacks: Physical tampering or damage to IoT devices, sensors, or communication infrastructure.
5) Man-in-the-Middle (MitM) Attacks: Intercept communications between IoT devices and networks to eavesdrop, manipulate, or inject malicious data.
6) Eavesdropping and Data Theft: Unauthorized access to data transmitted between IoT devices can lead to sensitive information theft.

7) Device Spoofing: Attackers impersonate IoT devices to gain unauthorized access to the network or data.
8) Injection Attacks: Malicious code or commands are injected into IoT device inputs, potentially causing unintended actions or unauthorized access.
9) Firmware Attacks: Manipulating or replacing device firmware to gain control over IoT devices or networks.
10) Password and Credential Attacks: Brute force attacks, password guessing, or exploiting weak credentials to gain access to IoT devices or networks.
11) Software Vulnerabilities: Exploiting security flaws or vulnerabilities in IoT device software to gain unauthorized access or control.
12) Traffic Analysis: Analyzing network traffic patterns to gather information or infer user behavior and device activities.
13) IoT Device Misconfiguration: Security misconfigurations in IoT devices or networks that leave them vulnerable to attack.
14) Phishing Attacks: Trick users into revealing sensitive information or executing malicious actions through deceptive messages or links.
15) Rogue IoT Devices: Unauthorized IoT devices added to the network without proper security measures, potentially introducing vulnerabilities.
16) Jamming and Interference: Deliberately disrupting IoT device communication through radio frequency interference.
17) Side-Channel Attacks: Exploiting unintended information leaks from IoT devices, such as power consumption, electromagnetic radiation, or timing.
18) Supply Chain Attacks: Tampering with IoT devices or components at various points in the supply chain to compromise their security.
19) Zero-Day Exploits: Exploiting previously unknown vulnerabilities in IoT device software or firmware.
20) IoT Network Sniffing: Unauthorized monitoring of IoT network traffic for data interception or analysis.

V. Anomaly Detection

Finding patterns or occurrences in a dataset that deviate from the norm is known as anomaly detection. While there are many different traditional methods for detecting anomalies, they frequently involve statistical techniques, machine learning algorithms, and rule-based approaches. Now let’s understand the comparison between them.

5.1 Comparison of Traditional techniques used in Anomaly detection

Table 5.1: Traditional techniques for anomaly detection

Technique	Description	Limitations
Signature-Based Detection	Matches known attack patterns (signatures) in data.	- Ineffective against new or unknown attacks. - Requires constant signature updates. - Cannot detect zero-day attacks.
Statistical-Based Detection	Model's normal behavior and flags deviations.	- Sensitivity to variations in normal traffic. - Difficulty handling highly dynamic environments. - Prone to false positives.
Rule-Based Detection	Uses predefined rules to identify anomalies.	- Limited to known attack patterns. - May not adapt well to changing attack techniques. - False positives due to strict rule definitions.
Heuristic-Based Detection	Employs heuristics and expert knowledge for detection.	- Highly dependent on the quality of heuristics. - May not generalize well to evolving threats. - Limited ability to detect novel attack types.
Machine Learning-Based Detection	Utilizes algorithms to learn and detect anomalies.	- Require labeled training data for supervised learning. - Sensitivity to the quality and quantity of data. - Potential bias in models and interpretability issues. - Model over fitting if not tuned properly.

Source: https://en.wikipedia.org/wiki/Anomaly_detection

VI. Related Work

Various approaches and methodologies have been investigated in related works regarding anomaly detection. The foundation for anomaly detection has been established by conventional approaches, which include statistical techniques like time-series analysis, Gaussian distribution modeling, and z-score analysis. At the same time, machine learning algorithms such as classification, clustering, isolation forests, and one-class support vector machines (SVMs) have become popular because they provide different insights into recognizing patterns that deviate from the norm. Rule-based systems have also been developed, which use expert-defined rules or predefined thresholds to identify abnormalities. Together, these varied approaches in related work provide a strong landscape for anomaly detection, with each offering special advantages and things to keep in mind for different kinds of data and use cases.

6.1 Published papers reviewed based on the Network IDS with different publicly available datasets

Table 6.1: Comparison between related research papers

Reference paper	Author	Publishing Year	Publishing Journal	Dataset Used	Method Used	Output	Anomaly/ Multiclass/other
17	Jun Yang	2020	MDPI	Own dataset by Authors	Sparse anomaly perception	Acts on sparse anomalies. Fast convergence and low prediction error.	Binary
18	Jin Wang	2020	Journal of Physics: Conference Series	Logs Parsing dataset-logevent2vec	Uses NLP and ML approach:NB,RF,Neural Network	Accuracy of time series class between 82 to 97%	Multiclass
19	Ahmed Zekry	2021	IEEE	Live Data	Convlstm model	98%	Multiclass
20	Ying Cui	2019	IEEE	Live Data	ML clustering	-	Multiclass
21	Rongbin Xu	2020	Elsevier (sciencedirect)	Live Data (Future work)	(I-LSTM)	-	-
22	Aymen Yahyaoui	2021	IEEE	NSL-KDD public dataset and Castalia 3.2 simulator for WSN	Designed own framework	99% accuracy	Reliable Event and Anomaly Detection Framework (READ-iot for short). The designed framework supports outlier's management in iot
23	Awajan Albara	2023	MDPI	Developed dataset by Authors	DNN	93.74% accuracy	Multiclass
24	Rayeesa Malik	2022	Hindawi	Ton_iot	DBN	86.3% accuracy	Binary
25	Alqahtani AS	2022	Springer	Developed dataset by Authors, NSL-KDD	FSO-LSTM	98.92% accuracy overall	Multiclass

26	Sharma Bhawana	2023	Elsevier	UNSW-NB15	GAN-DNN	91% accuracy	Binary

A wide range of machine learning (ML) and deep learning (DL) algorithms are used in anomaly detection; these algorithms are all intended to find anomalies and deviations in data. Conventional machine learning algorithms, like k-means clustering and density-based techniques like DBSCAN, group data points or define dense regions in the feature space in an effort to find outliers. Collaborative techniques such as isolation forests and one-class support vector machines (SVMs) are highly effective in identifying patterns of typical behavior and separating out anomalies. Entering the DL space, auto encoders – a type of neural network – reconstruct typical data and identify examples that substantially depart from this learned representation. Long short-term memory (LSTM) networks and recurrent neural networks (RNNs) are particularly good at identifying irregularities in sequential data, like time series. These below are some algorithms that are commonly use:

1) Isolation Forest: Efficiently isolates anomalies by randomly selecting features and partitioning data. It expects anomalies to require fewer splits to isolate. Real-time applications include network intrusion detection and fraud detection in financial transactions.
2) One-Class SVM: Separates data into normal and abnormal classes by constructing a boundary around normal data points. Any data point outside this boundary is considered an anomaly. It's used for detecting anomalies in sensor data from industrial equipment and real-time fault detection in machinery.
3) Auto encoders: Neural networks used for data reconstruction. They're trained to reconstruct normal data and flag anomalies when the reconstruction error is high. Real-time applications involve quality control in manufacturing and detecting anomalies in network traffic.
4) CNN (Convolutional Neural Network): These networks specialize in image and spatial data analysis. Adapted for anomaly detection by learning spatial features from normal data. They're applied in intrusion detection for video surveillance and real-time defect identification in image processing.
5) RNN (Recurrent Neural Network): Suitable for sequential data analysis, capturing temporal dependencies. Anomalies are detected when the predicted sequences deviate significantly from the actual data. Real-time applications include detecting anomalies in stock prices and identifying abnormal behavior in network traffic.

VII. FEATURE WORK

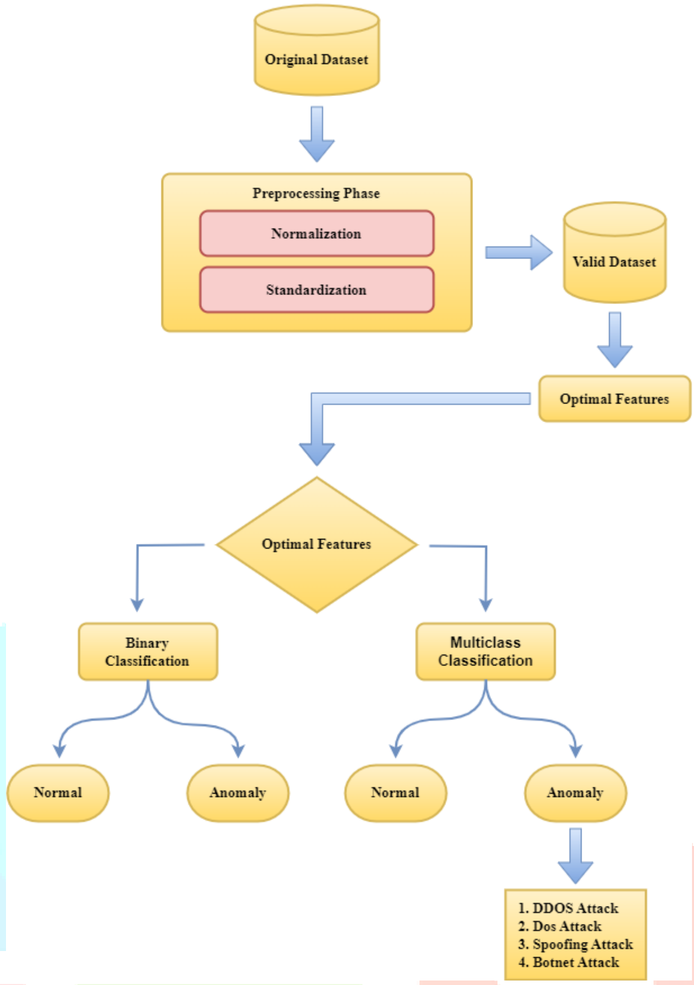

Fig.2: Proposed Framework

Compared to conventional techniques, ML and DL-based anomaly detection greatly improves security in IoT networks. In sharp contrast to rule-based systems, these sophisticated techniques autonomously learn from a variety of data sources and adjust in real-time to changing attack patterns. Their capacity to identify complex patterns make it possible to identify sophisticated attacks and unidentified threats, even ones that lack known signatures. Sophisticated analysis is offered by ML and DL systems, which lower false positives and negatives and increase security effectiveness. Through constant learning, they protect connected devices and systems from a range of threats, such as malware, data breaches, and intrusion attempts, maintaining a strong security posture in IoT environments. Researcher of this paper mainly focus on developing the ML model which suits to the given dataset train it and increase the frequency of that anomaly's detection in IoT network.

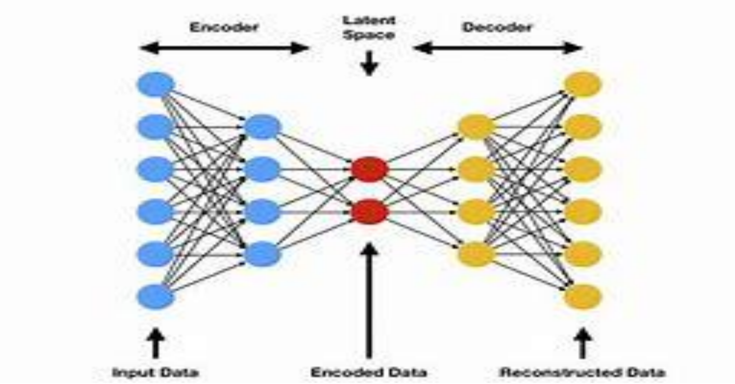

Fig.3: Proposed framework of auto encoder

1) Data Preparation:
   - Gather a dataset that includes both normal data and anomalous data. Ensure the dataset is representative of the real-world conditions you want to detect anomalies in.
2) Data Preprocessing:
   - Normalize or scale the data to have consistent ranges.
   - Split the data into a training set (containing only normal data) and a test set (containing both normal and anomalous data).
3) Autoencoder Architecture:
   - Design an Autoencoder architecture with an encoder and a decoder. The encoder reduces the data's dimensionality, while the decoder attempts to reconstruct the original data.
4) Training:
   - Train the Autoencoder on the training set, where the objective is to minimize the reconstruction error. The Autoencoder learns to encode and decode the normal data accurately.
5) Reconstruction and Error Calculation:
   - Use the trained Autoencoder to reconstruct both the normal and anomalous data points in the test set.
   - Calculate the reconstruction error for each data point by measuring the difference between the original and reconstructed data.
6) Threshold Setting:
   - Define a threshold for the reconstruction error above which a data point is considered an anomaly. The threshold can be set based on statistical methods or domain knowledge.
7) Anomaly Detection:
   - Compare the reconstruction errors of the test data points to the threshold.
   - If the reconstruction error exceeds the threshold, the data point is labelled as an anomaly. Otherwise, it's considered normal.
8) Evaluation:
   - Evaluate the performance of the Autoencoder-based anomaly detection by calculating metrics like precision, recall, F1-score, and the Receiver Operating Characteristic (ROC) curve.
9) Fine-Tuning:
   - Adjust the model and threshold as needed to achieve the desired balance between false positives and false negatives.
10) Real-Time Deployment:
   - Implement the trained Autoencoder model in a real-time system for continuous anomaly detection in streaming data.

REFERENCES

[1] Kyriazis, D., Varvarigou, T., White, D., Rossi, A., Cooper, J. (2013). Sustainable smart city IoT applications: Heat and electricity management & Eco-conscious cruise control for public transportation. IEEE 14th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), 2013, 1-5. [https://doi.org/10.1109/WoWMoM.2013.6583500]

[2] Chen, S., et al. (2014). A vision of IoT: Applications, challenges, and opportunities with China perspective. IEEE Internet of Things Journal, 1(4), 349–359. [https://ieeexplore.ieee.org/document/6851114]

[3] Malche, T., Maheshwary, P. (2017). Internet of Things (IoT) for building a smart home system. 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE.

[4] Zantalis, F., et al. (2019). A review of machine learning and IoT in smart transportation. Future Internet, 11(4), 94.

[5] Mezghani, E., Exposito, E., Drira, K. (2017). A model-driven methodology for the design of autonomic and cognitive IoT-based systems: Application to healthcare. IEEE Transactions on Emerging Topics in Computing, 1(3), 224–234.

[6] Zhao, J., et al. (2010). The study and application of IoT technology in agriculture. 2010 3rd International Conference on Computer Science and Information Technology, Vol. 2. IEEE.

[7] Ou, Q., et al. (2012). Application of the internet of things in smart grid power transmission. 2012 Third FTRA International Conference on Mobile, Ubiquitous, and Intelligent Computing. IEEE.

[8] Bichi, B. Y., Islam, S. U., Kademi, A. M., Ahmad, I. (2022). An energy-aware application module for the fog-based internet of military things. Discover Internet of Things, 2(1), 4.

[9] Khanna, A., Kaur, S. (2020). Internet of Things (IoT), Applications and Challenges: A Comprehensive Review. Wireless Personal Communications, 114, 1687–762. [https://doi.org/10.1007/s11277-020-07446-4]

[10] Tang, M., Alazab, M., Luo, Y. (2019). Big data for cybersecurity: vulnerability disclosure trends and dependencies. IEEE Transactions on Big Data, 5(3), 317–329.

[11] Zerihun, B. M., Olwal, T. O., Hassen, M. R. (2022). Design and Analysis of IoT-Based Modern Agriculture Monitoring System for Real-Time Data Collection. In: Computer Vision and Machine Learning in Agriculture, Vol. 2. Springer Singapore, 73–82.

[12] Dong, B., Wang, X. (2016). Comparison deep learning method to traditional methods using for network intrusion detection. 2016 8th IEEE International Conference on Communication Software and Networks (ICCSN). IEEE.

[13] Kruegel, C., Valeur, F., Vigna, G. (2004). Intrusion detection and correlation: challenges and solutions, Vol. 14. Springer Science & Business Media.

[14] Benkhelifa, E., Welsh, T., Hamouda, W. (2018). A critical review of practices and challenges in intrusion detection systems for IoT: Toward universal and resilient systems. IEEE Communications Surveys & Tutorials, 20(4), 3496–3509.

[15] Kocher, G., Kumar, G. (2021). Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges. Soft Computing, 25(15), 9731–9763.

[16] Ahmed, B. S., Chatterjee, A. (2022). A study of IoT anomaly detection methods and applications. [https://www.sciencedirect.com/science/article/pii/S2542660522000622]

[17] Yang, J., et al. (2020). Electricity prediction under edge devices based on sparse anomaly perception. Journal of Physics: Conference Series, 1659(1), 012015. [http://dx.doi.org/10.1088/1742-6596/1659/1/012015]

[18] Wang, J., et al. (2020). LogEvent2vec: LogEvent-to-vector based anomaly detection for large-scale logs in the internet of things. Sensors, 20(9). [http://dx.doi.org/10.3390/s20092451]

[19] Zekry, A., et al. (2021). Anomaly detection using IoT sensor-assisted ConvLSTM models for connected vehicles. 2021 IEEE 93rd Vehicular Technology Conference, VTC2021-Spring, 1-6. [http://dx.doi.org/10.1109/VTC2021-Spring51267.2021.9449086]

[20] Cui, Y., et al. (2019). Spatio-temporal correlation-based anomaly detection and identification method for IoT sensors. 2019 International Conference on Control, Automation and Information Sciences, ICCAIS, 1-6. [http://dx.doi.org/10.1109/ICCAIS46528.2019.9074607]

[21] Xu, R., et al. (2020). Improved long short-term memory-based anomaly detection with concept drift adaptive method for supporting IoT services. Future Generation Computer Systems, 112, 228–242. [http://dx.doi.org/10.1016/j.future.2020.05.035]

[22] Yahyaoui, A., et al. (2021). READ-IoT: Reliable event and anomaly detection framework for the internet of things. IEEE Access, 9, 24168–24186. [http://dx.doi.org/10.1109/ACCESS.2021.3056149]

[23] Awajan, A. (2023). A Novel Deep Learning-Based Intrusion Detection System for IoT Networks. Computers, 12(2), 34.

[24] Malik, R., Singh, Y., Zakir, A. S., Anand, P., Pradeep, K. S., Tewabe, C. W. (2022). An Improved Deep Belief Network IDS on IoT-Based Network for Traffic Systems. Journal of Advanced Transportation. [https://doi.org/10.1155/2022/7892130]

[25] Alqahtani, A. S. (2022). FSO-LSTM IDS: Hybrid optimized and ensembled deep-learning network-based intrusion detection system for smart networks. Journal of Supercomputing, 78, 9438–955. [https://doi.org/10.1007/s11227-021-04285-3]

[26] Sharma, B., et al. (2023). Anomaly-based network intrusion detection for IoT attacks using deep learning technique. Computers & Electrical Engineering, 107, 108626.

[27] Dang, T.-B., Le, D.-T., Nguyen, T.-D., Kim, M., Choo, H. (2021). Monotone split and conquer for anomaly detection in IoT sensory data. IEEE Internet of Things Journal, 1. [http://dx.doi.org/10.1109/JIOT.2021.3073705]

[28] Hou, R., Pan, M., Zhao, Y., Yang, Y. (2019). Image anomaly detection for IoT equipment based on deep learning. Journal of Visual Communication and Image Representation, 64, 102599. [http://dx.doi.org/10.1016/j.jvcir.2019.102599]

[29] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M., Baik, S.W. (2021). CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimedia Tools and Applications, 80(11), 16979–16995. [http://dx.doi.org/10.1007/s11042-020-09480-w]

[30] Basu, S. 1997. The Investment Performance of Common Stocks in Relation to their Price to Earnings Ratio: A Test of the Efficient Markets Hypothesis. Journal of Finance, 33(3): 663-682.