A Clustering Based Approach to Detect Cyber Attacks in Internet of Things (IoTs)

[1]Barkha Kumari, [2]Vinay Singh, [3]Mohit Kumar, [4]Shrikant Upadhyay
[1]Usha Martin University, Ranchi, Jharkhand, India, 835103
[2]Usha Martin University, Ranchi, Jharkhand, India, 835103
[3]MIT Art, Design and Technology University, Pune, India
[4]Department of ECE, Cambridge Institute of Technology, Ranchi, Jharkhand, India

[1]singhvarsha35@gmail.com, [2]vinaysinghumu@gmail.com,[3]mohit.kumar@mituniversity.edu.in, [4]shri.kant.yay@gmail.com

Abstract

The Internet of Things application is currently confronted with novel ultimatum as the dimension of sensitive data grows. To secure data and ensure privacy, a judicious IoT system is required. In recent days machine learning has offered a propitious solution for securing sensitive data and devices of Interne of Things application. This paper proposed a framework for securing Internet of Things application. The proposed scheme is mainly developed for real time data monitoring with preserving confidentiality and classification of normal and abnormal traffic. A cluster-based approach has been considered here for decision purposes and real time analysis for securing application. The numerical result of proposed framework outperforms in respect of benchmark models in terms of accuracy, false positive rate and area under the cover.

Keywords- Internet of Things, Clustering, Confidentiality, Monitoring System, IoT gateway

1. Introduction

The internet of things (IoT) can be broadly defined as a pervasive network of a range of connected smart nodes that offer diverse digital services including the collection of environmental and user data. For example, IoT nodes can sense process and communicate information through IoT infrastructure to improve the quality of services and user experience in sectors ranging from healthcare to transportation to power management to military etc. on the flip side, IoT devices and system also be on attack vector where an attacker can seek to obtain information, target other entities, and facilitate nefarious activities. Internet of things devices and system are increasingly targeted by cyber criminals as they become an integral part of our society and ecosystem [1-2].

In recent years, there has been renewed interest in exploring the utility of artificial intelligence (AI) techniques such as machine learning (ML) and deep learning in designing cybersecurity solutions such as network detection, thread intelligence forensic investigation and privacy preserving techniques [3-5] . However, designing efficient and effective AI based IoT attack detection system remains an open research challenge [6-8]. Different types of attacks like DoS, DDoS, MIMA, spoofing etc. has become increasingly sophisticated in recent years.in fact a single DDoS attack is now able to infect a large group of devices.

Attackers install malicious software which can gain control over a group of compromised machines located within the same network. Machine learning (ML) techniques learn the patterns behind attacks in order to detect them before network resource become unavailable. Modern defense systems make use of ML and DL techniques in an effort to effectively respond to complicated cyberattacks [9]. Sensitive data from IoT application generated by various embedded smart devices transmitted to cloud computing via various gateway layer for global data processing [10-13]. This secure IoT application is designed for smart health care, Industrial IoT, smart home, agriculture, smart city etc.

A framework like this allows for early detection of cyberattacks and prevention from it. To securing the sensitive data of IoT application machine learning techniques including clustering-based algorithm are often used to evaluate sensitive data and categorize devices into specific subgroups based on specific data as well as uncover anomalous pattern [14]. Several cyberattacks can be predicted via IoT data analysis. The classification of various cyberattacks such as DoS, DDoS, MIMA, Spoofing, Hello Flood for example might indicate the risk of DoS attack. Clustering Based methods are a powerful analytical tool that can be used to detect cyberattacks in IoT application. The grouping of data from different devices enables the detection of risk or threats. Numerous cyberattacks detection techniques in IoT application have been presented in the literature. However present cyberattacks detection techniques have several constraints that are not able to detect newer types of attacks in real world applications [15-17].

These include latency problem (especially troublesome in time critical application) in detection of cyberattacks and data leakage between source to destination nodes. Also, the traditional security and privacy solution are not suitable for protecting sensitive data internet of things due to the diverse characteristics of the smart devices of IoT application. In this paper we have developed a clustering-based algorithm in ML techniques to detect cyberattacks in IoT application [18-20]. The framework can gather, monitor and analyze sensitive data in real time. The novelty of this framework is that it uses clustering-based techniques to provide analytic services for internet of things monitoring purposes. Machine learning approaches based on clustering are used to evaluate and detect anomaly changes in sensitive data. The experiments demonstrate that our framework can significantly accelerate performance while maintaining a high degree of accuracy in the overall analysis process safely. The remaining part of the paper are structured as follows section 2 describes related works, in section 3, the proposed work Is described, section 4 and 5 describe the implementation and results of the proposed work. finally, the conclusion of the paper is described in section 6.

2. Related work:

Several techniques to solve the security problems in an IoT based application system is to share data among various smart devices connected in an IoT network to execute specific task without human interruption.

Table 1 shows the various techniques to secure the IoT application. The Internet of things (IoT)layer is useful for basic conveyed functionality and can also provide true analytic services for sensible decisions with a localized smart community domain.

Table1: Related work on cyberattacks detection using machine learning in IoT

Sl. No.	References	year	Details & Domains	Objective and Functionality	Future Scope
1.	Alaba et al. [2]	2017	Performed survey on security issues in IoT Application	A taxonomy of contemporary security cancerns in application and architecture context is offered	Only investigate the threats and attacks; security solutions are required to propose
2	T. Saba et al[4]	2021	aim to present a machine learning-based approach for autonomous IoT Security to achieve optimal energy efficiency and reliable transmissions.	proposed protocol optimizes network performance using a model-free Q-learning algorithm and achieves fault-tolerant data transmission. Second, it accomplishes data confidentiality against adversaries using a cryptography-based deterministic algorithm.	More network efficacy and data availability with the smart intrusion detection system is needed
3	M.Saharkhizan et. al [5]	2020	design an approach using advanced deep learning to detect cyber-attacks against IoT systems.	presented a novel ensemble method to detect IoT cyber-attacks over Modbus network traffic.	explore the explain ability of LSTM models to propose a more transparent DL model for detect-ing IoT cyber-attacks, particularly those in adversarial settings (e.g., battlefields). deployment of proposed

					approach to different IoT protocols and transfer the learned Modbus cyber-attacks to other domains
4	A. Hamza et al. [6]	2022	aim to enhance cybersecurity of a large-scale IoT infrastructure-true by formally capturing the expected behavior of the system using the static profile of devices' intended usage, buildings information, and network configurations (pre-deployment) along with dynamic diagnosis (post-deployment) of network activity using machine-learning models	developed a tool that automatically generates a formal model for the intended behaviour of the entire IoT system by combining the MUD profile of devices, Brick schema of buildings, and network configurations. 2.developed anomaly detection models to continuously monitor the activity of IoT traffic flows at the building level and device level.	compatibility of the IoT system behaviours can be systematically checked against three representative organizational policies, prior to deployment.
5	L. Vu et al [7]	2020	propose a novel deep transfer learning (DTL) method that allows to learn from data collected from multiple IoT devices in which not all of them are labelled.	Proposed approach aims to address the problem of ''lack of labelled information'' for the training detection model in ubiquitous IoT devices.	distribute the training process to the multiple IoT nodes by using the federated learning technique to speed up this process. 2. the cur-rent DTL model is developed based on AutoEncoder. In the future, we will attempt to

					extend this model based on other neural networks such as Deep Adaptation Network (DAN), Adversarial Discriminative Domain Adaptation (ADDA), Maximum Classifier Discrepancy (MCD), and Conditional Domain Adversarial Network (CDAN)
6	I. Ullah et al. [8]	2021	design and develop a novel anomaly-based intrusion detection model for IoT networks.	This article proposes and develops an anomaly detection model for IoT networks using a convolutional neural network to detect and classify binary and multiclass anomalies. We provide a technique for detecting anomalous activity in IoT networks by generating a new dataset from an existing one.	plan to investigate further anomaly detection using various deep learning methods, like FFN and RNN, GAN, and contrast the findings to those obtained using a deep convolutional neural network model.
7	N. Abdalgawad et al.[9]	2021	This paper shows that it is possible to use generative deep learning methods like Adversarial Autoencoders	Aim to uses generative deep learning methods like AAE and BiGAN to classify attacks with a very	Required large set of attacks and IoT devices, to use generative deep learning methods like AAE and

			(AAE) and Bidirectional Generative Adversarial Networks (BiGAN) to detect intruders based on an analysis of the network data.	high accuracy.	BiGAN to classify attacks with a very high accuracy in future.
8	E. Muhati et al. [10]	2022	This article proposes a combination of cyberattack projection and cyber-defence agility estimation to dynamically but reliably augur intrusion detection performance.	To apply a machine-learning (ML)-based hidden Markov model (HMM) to predict intrusion detection agility.	To check our models' validity, HMM is applied to a selected data set with a basic assessment of numerically quantified details but scanty awareness of exposed attack surface.
9	MM Hassan et al. [11]	2020	In this article, we propose to improve the trustworthiness of an IIoT network [i.e., supervisory control and data acquisition (SCADA) network] through a reliable and salable cyberattack detection model.	an ensemble-learning model based on the combination of a random subspace (RS) learning method with random tree (RT) is proposed for detecting cyberattacks of SCADA by using the network traffics from the SCADA-based IIoT platform	Improvement required to optimal random feature subsets when there is a very small number of features. Also required improvement in execution time which is slightly more than the execution time of a single RT classifier.
10	A Aljuhani et al [12]	2021	This paper analyses recent studies concerning DDoS detection methods that have adapted single and	paper discusses different DDoS defence systems based on ML techniques that make use of a virtualized environment, including cloud	explored some of the challenges currently facing this field of research and outlined important considerations

			hybrid ML approaches in modern networking environments.	computing, software-defined network, and network functions virtualization environments.	with regard to the design of effective and practical defence systems for combating DDoS attacks.

The proposed secured IoT framework for early detection of cyberattacks is shown in fig 2.

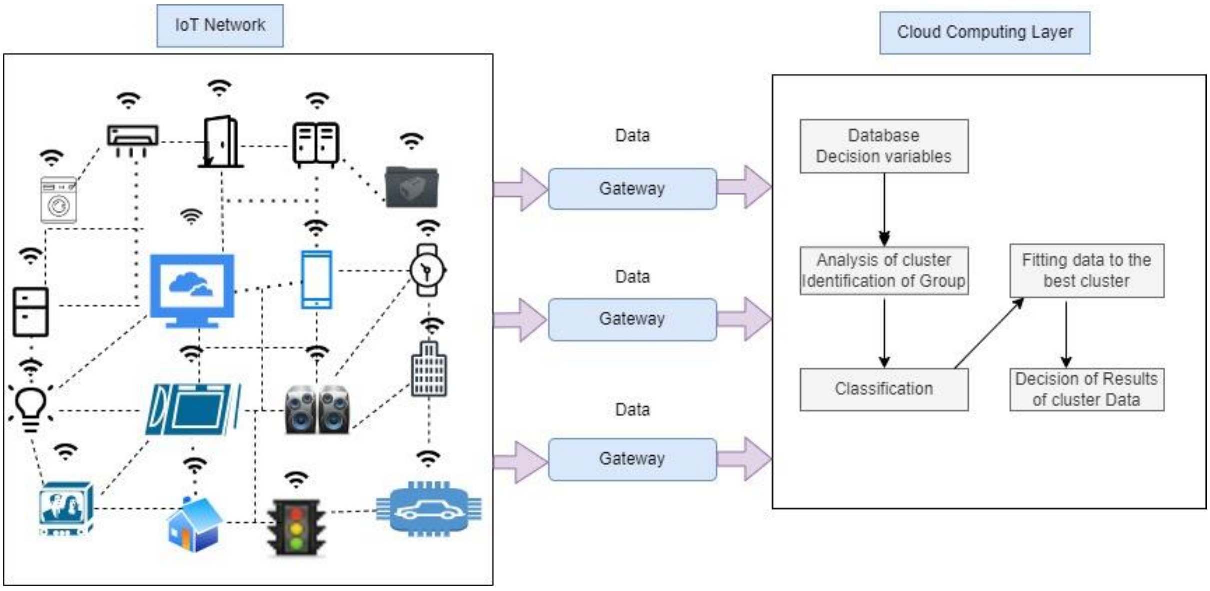

Fig. 1: Proposed Secure IoT Network using Cluster Based Techniques.

3. Proposed Framework

This framework has considered 3-main entities as smart community devices, IoT gateways and cloud servers. Smart community devices include different types of IoT devices such as sensors, gadgets, appliances, and other machines that collect and exchange data over the internet. They are programmed for certain applications and can be embedded into other IoT devices. For example, an IoT devices in your car can identify the traffic ahead and send out a message automatically to the person you are about to meet of your impending delay[21-23]. The IoT gateway connect the smart devices network with cloud servers and endpoint devices. IoT technology works by using built in sensors, software, and communication hardware to collect and send data generated by their usage and environment IoT devices share this data through IoT gateway or platform or another IoT devices. Data is typically sent to the cloud for storage and analysis[24].

Theses IoT gateways are trying in the edge layer for communication as well as quick response. The cloud layer id aggregated and monitoring IoT data using a cluster-based mining technique [25].

The following steps describe the working of the proposed model:

Step 1: The IoT data is collected form different devices after a regular interval (t sec) and sent to the end point devices gateway through ZigBee network.

***Step 2*:** End point devices will create device profile (p1, p2) and sent them to the IoT gateway for processing.

Step 3: The end point devices first identify the difference between two different device profiles.

Consider the profiles are P1 and P2 and corresponding vectors of data parameter are

A= (a1, a2,a3………..an) and B=(b1,b2,b3…………..br).

These two profiles are created in two different time intervals (t sec). calculate the centroid of each profile cluster and the distance between each data object with the centroid will be calculated [26]. The distance between the data objects and two profiles are calculated by measuring the Euclidean distance vector as represented by equation 1.

$$\text{E (A, B)} = \sqrt{\sum_{i=1}^{n}(ai - bi)}^2 \qquad (1)$$

The difference of captured data profile can be calculated using Euclidean norm of the Euclidean distance vector as represented by equation 2.

$$\text{D (P1, P2)} = \sqrt{\sum_{i=1}^{i=4}(Ei\ )}^2 \qquad (2)$$

with the distance vector data is aggregated in two different clusters.

Step 4: Two different clusters have been created on the basis of equation 3.

$T^{d}{}_{\text{value=1-D (P1, P2)/2}}$

where T_{value} is the confidence score

if $T_{\text{value}} \leq$d, then the data is normal and send in cluster 1.

if $T_{\text{value}\geq}$α, then the data is abnormal or threat detected and send in cluster 2.

where α is the actual measured data of the IoT devices.

Step 5: After clustering the data is encrypted, cluster 1 is stored in cloud server and cluster 2 is stored in edge server for cyberattack detection.

Step 6: The Internet of Things application can access the data after authentication form the server. The cluster key is provided after authentication.

This framework can be used for easy cyberattack detection and remote monitoring of the sensitive data of IoT application.

In this framework the IoT data is collected from IoT devices and stored in a cloud server that smart community like Industrial IoT, Smart health care system, smart home, agriculture etc can use in smart environment [27-28].

4. Results & Discussions

In this section the following aspects are analysed experimentally.

(1) The performance of numerical value of our proposed model compared with benchmark models.
(2) How the main parameter influences the functioning of the proposed model.

for the clustering purposes, we have used dataset[29] provided by a well-known Chinese internet company, which specializes in cyberspace security and captures web logs of up to 2 TB everyday .In the above datasets ,the attack data mainly include DDoS attack ,SQL injection attack ,Man in the middle attack and other types of attacks.

The performance of the proposed models is evaluated in the following aspects. The proposed model can be used for early detection of threats and remote monitoring for securing sensitive data. Moreover, we choose linear SVM, Naïve Bayes and Decision Tree algorithms as benchmark to evaluate the performance of the proposed clustering classification on given dataset. The proposed method is evaluated via the hold out method for training a model .it is a standard technique to estimate the performance of the machine learning model. The complete dataset is splitted into two subsets. Cluster techniques is trained one of the. Then its classification performance is evaluated on the other subset. The process is repeated several times and the average value of each index is eventually returned as the evaluation result of the hold-out method [30-34].

To quantify the performance of the proposed model and other compared model the following standard measurements are used: accuracy(A), Precision(P), False positive rate(fpr) and F1 score(f1). Here the positive and negative instances refer to the abnormal and normal traffic respectively.

The given four evaluation indexes are explained as follows in the terms of confusion matrix (Table 2).

Accuracy: A=(TP+TN)/(TP+FP+TN+FN).

It denotes the percentage of the correctly classified instances over total instances.

Precision: p=TP/(TP+FP). It indicates the percentage of the correctly classified positive instances over total positive instances which are classified as positive.

False positive rate: fpr =FN/(TP+FN). It explains the percentage of the misclassified positive instances over total positive instances.

F1 score: f1=2*p*r/(p+r). It is the harmonic average of precision and recall rate, where recall rate is defined as r=TP/(TP+FN).

Table 2: Clustering Result (A=Accuracy, fpr=false positive rate , auc=area under the cover)

Data points	Accuracy	False Positive Rate	Area Under the Cover
2000	97.49	0.89	96.45
3000	96.40	0.92	98.45

6000	94.53	6.12	85.34
8000	95.26	7.12	84.12
10000	96.36	8.12	82.28

The accuracy and false positive rate are shown in Fig. 3 and Fig. 4 below.

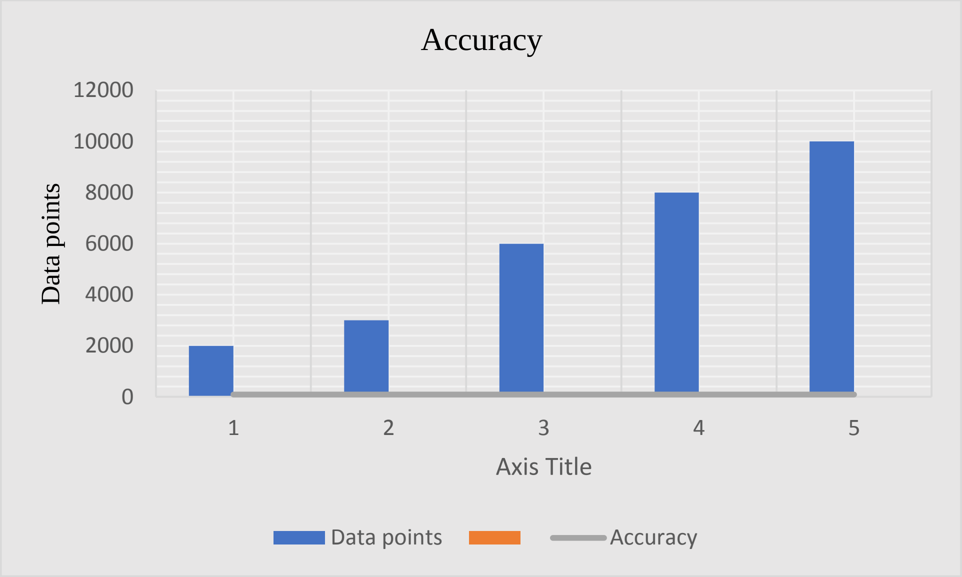

Fig. 2. Rate of Accuracy for different data points

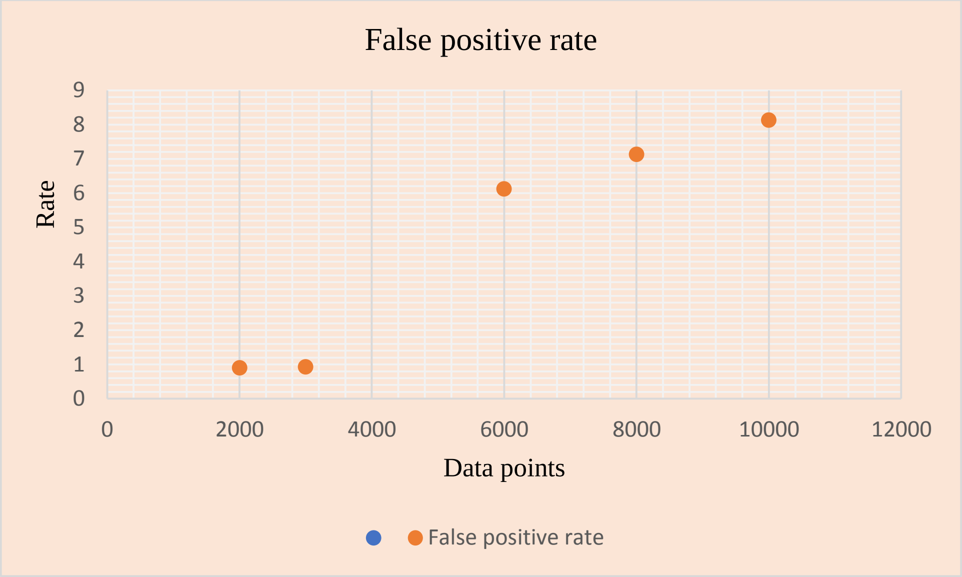

Fig. 3: False rate

5. Conclusion & Future Scope

In this paper, a novel clustering approach was proposed for cyberattack detection in IoT network. To preserve data and ensure privacy, a secure IoT application is required. The proposed framework is mainly designed for real time data monitoring with maintaining confidentiality and classify abnormal and normal traffic using clustering approach. The clustering-based technique is adopted to analyse IoT data for monitoring and identify normal and abnormal traffic in a secure manner. A quick decision can be made from the cluster data. The numerical results showed that the proposed model outperformed the benchmark models in term of accuracy, false positive rate and area under cover. In such a setting, research is heavily focused on the three key issues data monitoring, classification of normal and abnormal traffic and data confidentiality. In future research, the privacy and more attack detection of such sensitive data can be considered.

References

[1] Saharkhizan, M., Azmoodeh, A., Dehghantanha, A., Choo, K. K. R., & Parizi, R. M. (2020). An ensemble of deep recurrent neural networks for detecting IoT cyber-attacks using network traffic. IEEE Internet of Things Journal, 7(9), 8852-8859.

[2] Aljuhani, A. (2021). Machine learning approaches for combating distributed denial of service attacks in modern networking environments. IEEE Access, 9, 42236-42264.

[3] Alabdulatif, A., Khalil, I., Yi, X., & Guizani, M. (2019). Secure edge of things for smart healthcare surveillance framework. IEEE Access, 7, 31010-31021.

[4] Saba, T., Haseeb, K., Shah, A. A., Rehman, A., Tariq, U., & Mehmood, Z. (2021). A machine-learning-based approach for autonomous IoT security. IT Professional, 23(3), 69-75.

[5] Saharkhizan, M., Azmoodeh, A., Dehghantanha, A., Choo, K. K. R., & Parizi, R. M. (2020). An ensemble of deep recurrent neural networks for detecting IoT cyber-attacks using network traffic. IEEE Internet of Things Journal, 7(9), 8852-8859.

[6] Hamza, A., Gharakheili, H. H., Pering, T., & Sivaraman, V. (2022). Combining Device Behavioral Models and Building Schema for Cybersecurity of Large-Scale IoT Infrastructure. IEEE Internet of Things Journal, 9(23), 24174-24185.

[7] Vu, L., Nguyen, Q. U., Nguyen, D. N., Hoang, D. T., & Dutkiewicz, E. (2020). Deep transfer learning for IoT attack detection. IEEE Access, 8, 107335-107344.

[8] Ullah, I., & Mahmoud, Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. IEEE Access, 9, 103906-103926.

[9] Abdalgawad, N., Sajun, A., Kaddoura, Y., Zualkernan, I. A., & Aloul, F. (2021). Generative deep learning to detect cyberattacks for the IoT-23 dataset. IEEE Access, 10, 6430-6441.

[10] Muhati, E., & Rawat, D. B. (2021). Hidden-Markov-model-enabled prediction and visualization of cyber agility in IoT era. IEEE Internet of Things Journal, 9(12), 9117-9127.

[11] Hassan, M. M., Gumaei, A., Huda, S., & Almogren, A. (2020). Increasing the trustworthiness in the industrial IoT networks through a reliable cyberattack detection model. IEEE Transactions on Industrial Informatics, 16(9), 6154-6162.

[12] Aljuhani, A. (2021). Machine learning approaches for combating distributed denial of service attacks in modern networking environments. IEEE Access, 9, 42236-42264.

[13] Kumar, M., Mukherjee, P., Verma, S., Kaur, M., Singh, S., Kobielnik, M., ... & Ijaz, M. F. (2022). BBNSF: Blockchain-Based Novel Secure Framework Using RP2-RSA and ASR-ANN Technique for IoT Enabled Healthcare Systems. Sensors, 22(23), 9448

[14] Kumar, M., Kumar, A., Verma, S., Bhattacharya, P., Ghimire, D., Kim, S. H., & Hosen, A. S. (2023). Healthcare Internet of Things (H-IoT): Current Trends, Future Prospects, Applications, Challenges, and Security Issues. Electronics, 12(9), 2050.

[15] Kumar, M., Verma, S., Kumar, A., Ijaz, M. F., & Rawat, D. B. (2022). ANAF-IoMT: A Novel Architectural Framework for IoMT-Enabled Smart Healthcare System by Enhancing Security Based on RECC-VC. IEEE Transactions on Industrial Informatics, 18(12), 8936-8943.

[16] Kumar, M., Mukherjee, P., Verma, K., Verma, S., & Rawat, D. B. (2021). Improved deep convolutional neural network based malicious node detection and energy-efficient data transmission in wireless sensor networks. IEEE Transactions on Network Science and Engineering, 9(5), 3272-3281

[17] Kumar, M., Mukherjee, P., Verma, S., Shafi, J., Wozniak, M., & Ijaz, M. F. (2023). A smart privacy preserving framework for industrial IoT using hybrid meta-heuristic algorithm. Scientific Reports, 13(1), 1-17

[18] Pratap, A., Kumar, A., & Kumar, M. (2021, May). Analyzing the Need of Edge Computing for Internet of Things (IoT). In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020 (pp. 203-212). Singapore: Springer Singapore

[19] Gupta, A., Kumar, M., Rangra, A., Tiwari, V. K., & Saxena, P. (2012). Network intrusion detection types and analysis of their tools. Department of Computer Science and Information Technology, Jaypee University of Information Technology, India.

[20] Upadhyay, S., Kumar, M., Kumar, A., Ghafoor, K. Z., & Manoharan, S. (2022). SmHeSol (IoT-BC): smart healthcare solution for future development using speech feature extraction integration approach with IoT and blockchain. Journal of Sensors, 2022.

[21] Kumar, M., Mittal, S., & AKHTAR, A. K. (2020). A NSGA-II Based Energy Efficient Routing Algorithm for Wireless Sensor Networks. Journal of Information Science & Engineering, 36(4).

[22] A. Kumari, N. Gandotra. P. Joshi, S. Upadhyay, “Impact of Node Movement on MANET Using Different Routing Protocol for QoS Improvement under Different Scenario”, International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248-9738 Volume- 1, Issue- 3, pp. 95-101, 2012.

[23] P. Joshi, A. Kumar, S. Upadhyay, S. Vijay et. al., “Impact of Various Mobility Model and Judgment for Selecting Mode of Network in Different Mobility Situation for Mobile Ad-Hoc Network (MANET)”, IEEE International Conference on "Emerging Technology Trends in Electronics, Communication and Networking (ET2ECN- 2012)At: Gujarat, Volume: 978-1-4673-1628-6, 2012.

[24] S. Upadhay, SK. Sharma, A. Upadhyay, “Analysis of Different Classifier Using Feature Extraction in Speaker Identification and Verification under Adverse Acoustic Condition for

Different Scenario", International Journal of Innovations in Engineering and Technology (IJIET), Volume 6 Issue 4ISSN: 2319 – 1058, pp. 425-434, April 2016.

[25] S. Upadhay, SK. Sharma, A. Upadhyay, "Analysis of Feature Extraction Behavior for Speaker Identification and Verification in Adverse Acoustic Condition", International Journal of Control Theory and Applications, IJCTA 10 (9) ISSN: 0974-5572, pp. 25-31, 2017.

[26] MC Saxena, F Banu, A Shrivastava, M Thyagaraj & S. Upadhyay, "Comprehensive Analysis of Energy Efficient Secure Routing Protocol for Sensor Network", Volume 62, Part 7, pp. 5003-5007, 2022, Material Today.

[27] A. Upadhyay, SK Sharma, S. Upadhyay "Face Identification & Verification Using Hidden Markov Model with Maximum Score Approach", Indian Journal of Science & Technology, Vol. (10) 47, pp.1-6, December 2016.

[28] A. Upadhyay, SK Sharma, S. Upadhyay, "Robust Feature Extraction using Embedded HMM for Face Identification & Verification" International Journal of Applied Engineering Research, Vol. 12, No. 24, pp. 15729-15777, 2017.

[29] S. Upadhyay, A. Upadhyay et. al "A Study and Analysis of Suitable Channel Access Protocol for Mobile Ad-hoc Network Considering Different Application" International Journal of Computer Engineering and Application (IJCEA), Vol.-5, Issue-3, pp. 93-106, March 2014.

[30] A. Kumari, N. Gandotra et. al. "Impact of Mobility on the Performance of Wireless Ad hoc Networks Scenario using Distance Vector Routing Protocol", International Journal of Computer Application (IJCA), Volume-57, No. - 12, pp. 14-21, November 2012.

[31] S. Upadhyay, SK Sharma & A. Upadhyay, "Speaker Identification and Verification Using Different Model for Text-Dependent", International Journal of Applied Engineering Research, Vol. 12, No. 08, pp. 1633-1638, 2017.

[32] A. Londhe, P VRD Rao, S. Upadhyay & R. Jain, "Extracting behavior identification features for monitoring and managing speech dependent smart mental illness healthcare systems", Computational Intelligence and Neuroscience, Vol. 2022, pp. 1-14, May 2022.

[33] S. Upadhyay, M. Kumar, A. Kumar, R. Kranti, "Feature Extraction Approach for Speaker Verification to Support Healthcare System using Blockchain for Data Privacy", Computational and Mathematical Methods in Medicine, Vol. 2022, pp. 1-12, July 2022.

[34] S. Upadhyay, M. Kumar, A. Upadhyay, "Digital Image Identification and Verification using Maximum and Preliminary Score Approach with Watermarking for Security Enhancement and Validation", Journal of Electronics, Vol. 12, Issue 7, pp. 1-15, 2023.