IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

International Conference on Muti-Disciplinary Application & Research Technologies (ICMART-2023)

Geetanjali Institute of Technical Studies

Vol. 10, Special Issue 2, May 2023

System for Monitoring and Notifying Cyber Threats using ML Techniques

Pankaj Kumar Vaishnav[1], Rakshit Kothari[2], Meenal Joshi[3], Dr. Mayank Patel[4], Dr. Narendra Singh Rathore[5]

Assistant Professor, Computer Science and Engineering, Geetanjali Institute of Technical Studies, Udaipur, India[1,2,3]

Professor and Head, Computer Science and Engineering, Geetanjali Institute of Technical Studies, Udaipur, India[4]

Campus Director, Computer Science and Engineering, Geetanjali Institute of Technical Studies, Udaipur, India[5]

Abstract: Cybercriminals are taking advantage of weaknesses that exist anywhere there are computers. Assessing vulnerabilities and suggesting mitigation techniques are priorities for ethical hackers. In the subject of cyber security, there has been a pressing need for the creation of efficient methods. The majority of surveillance approaches now in use are unable to handle the dynamic and intricate nature of cyber-attacks on computer networks. Due to machine learning's efficiency in addressing cyber security challenges, it has lately become a topic of significant relevance. For the most difficult problems in cyber security, such as intrusion detection, malware classification and detection, spam detection, and phishing detection, machine learning approaches have been used. Machine learning may assist identify cyber security risks more effectively than other software-oriented approaches, which lessens the workload on security analysts even if it cannot fully automate a cyber security system. As a consequence, effective adaptive approaches, such as various machine learning techniques, can lead to increased detection rates, decreased false alarm rates, and reasonable computation and communication costs. Our major objective is to demonstrate that the challenge of identifying assaults fundamentally differs from these other applications, making it far more difficult for the intrusion detection community to properly use machine learning.

Keywords*:* machine learning, cybercrime, intrusion detection, surveillance

I. INTRODUCTION

Currently, sophisticated cyber-warfare is being used by political and commercial organizations more frequently to harm, disrupt, or suppress the information content in computer networks. In building network protocols, it is necessary to provide dependability against incursions by strong attackers who may even control a small number of network participants. The groups under their control have the ability to launch both passive (eavesdropping, for example) and aggressive attacks (e.g., jamming, message dropping, corruption, and forging). The technique of actively tracking events in a computer system or network, examining them for indications of potential problems, and frequently blocking unwanted access is known as intrusion detection. This is often done by automatically gathering data from various systems and network sources, then reviewing the data for any potential security issues.

Conventional methods of intrusion detection and prevention, including as firewalls, access control systems, and encryptions, have various shortcomings that prohibit them from providing complete defence against increasingly complex attacks like denial-of-service attacks on networks and systems [1]. Besides that, the majority of systems created using such approaches have significant false positive and false negative detection rates as well as a lack of ongoing adaptation to evolving harmful behaviour. Nonetheless, a number of Machine Learning (ML) approaches have been applied to the intrusion detection issue in the last decade in an effort to increase detection rates and adaptability. These methods are frequently employed to keep the attack knowledge bases complete and up to date. Cyber security and defence against multiple cyber-attacks have recently become urgent concerns. The massive expansion of computer networks and their phenomenal growth are the primary causes of it Several pertinent applications that people or groups utilize for either personal or commercial purposes, particularly after the adoption of the Internet of Things (IoT). With large-scale networks, the cyber-attacks result in severe physical harm and significant financial losses. Even so, current solutions like user authentication, hardware and software firewalls, and data encryption techniques are unable to meet the challenge of rising demand and are unable to thwart the numerous cyber threats to computer networks [3].

Due to the quicker, more severe growth of intrusion systems, these conventional security structures are insufficient as a defence. A barrier simply restricts access between networks; hence it blocks access between networks. Yet, it doesn't provide any signals in the event of an inside assault. Thus, it is evident that precise defensive systems, such as an intrusion identification system (IIS) based on machine learning, must be developed. An intrusion identification system (IIS) is often a hardware or software component that keeps an eye out for unusual behaviour and policy violations on a network or system [3]. An IIS is a system or network used to detect hazards or assaults relating to network security, such as denial-of-service attacks, by identifying inconsistencies and aberrant activity on a network during the running of daily operations (Dos). Detecting, deciding, and controlling unwanted system behaviour

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

International Conference on Muti-Disciplinary Application & Research Technologies (ICMART-2023)

Geetanjali Institute of Technical Studies

Vol. 10, Special Issue 2, May 2023

such unauthorised access, modification, and destruction is also made easier with the use of an intrusion identification system. Depending on the user's perspective, there are many types of intrusion identification systems. These include host-based and network-based IIS, for instance. Often, an IIS must cope with issues including enormous levels of network traffic, extremely unequal data distribution, and the It is challenging to distinguish between normal and aberrant behaviour, and thus makes ongoing adaptation to a continuously changing environment necessary [8]. Generally speaking, the difficulty is effectively capturing and categorizing different behaviours in a computer network. There are commonly two sorts of strategies for categorizing network behaviours: both misuse and anomaly detection. Using signature matching algorithms, abuse detection approaches scan network and system activity for instances of known misuse. This method is efficient at identifying already known attacks. However new assaults are frequently ignored, leading to false negatives. The IIS may create alerts, but responding to each one costs time and resources and makes the system unstable [4]. To solve this issue, IDS must be patient enough to gather warnings and make decisions based on their association, rather than launching the elimination process as soon as the first symptom is identified.

II. ENHANCING THE MODEL FOR ASSURANCE OF CYBER SECURITY

Every time a group of auditors participates in an IT, data security, or compliance audit, there are recurring phases like designing, defining objectives and scope, outlining terms of engagements, carrying out the audit, gathering supporting documentation, assessing risks, reporting the audit findings, and scheduling follow-up tasks. Designing a cyber-security audit isn't totally different than any kind of audit. This however will take a great deal of effort thanks to the quality of the many cyber security domains. Unfortunately, the scope of the internal audits does not include reviewing most cyber capabilities [5]. This particular framework addresses the need for assurance through management reviews, cyber risk assessments, information management and protection, risk analytics, crisis management, and response. It also addresses the development life cycle, security programmer, third-party management, information/asset management, access management, threat/vulnerability management, and threat/vulnerability management [8]. Additionally, Deloitte's framework is compatible with industry frameworks such as those of the World Organization for Standardization (ISO), Committee of Sponsoring Organizations of the Tread way Commission (COSO), National Institute of Standards and Technology (NIST), data Technology Infrastructure Library (ITIL), and (ISO) [8].

III. CURRENT WORKS

It is practically hard to quantify or justify the reasons why cyber security has such a significant influence on the constantly expanding and swiftly expanding sector of cyber security. Allowing harmful threats to run in any location, at any time, or in any context is a long way from being acceptable, and may cause forceful injury. It specifically pertains to the complex network of online consumer and business information that cyber security organizations are struggling to protect and manage. For individuals and families, corporations, governments, and academic institutions that operate inside the framework of the global network or internet, cyber security may be a crucial consideration. We will enhance the state of cyber security with the use of machine learning. The high-tech infrastructure of today, which includes network and cyber security systems, is collecting enormous volumes of data and doing analytics on nearly all the important components of mission-critical systems. While humans continue to provide the essential operational supervision and insightful perspectives for today's infrastructure. The majority of intrusion detection systems, starting with your firewall, concentrate on perimeter attack surface threats. It provides protection for the north-south traffic on your network, but it ignores the lateral spread (east-west) that many modern network threats use to enter your organization's network and remain there undetected [2]. We can be certain of this since studies have shown that only 35% of risks are identified by north-south surveillance [10]. Typically, when an IIS notices suspicious behaviour, the violation is sent to a security information and event management (SIEM) system, where true threats are finally identified among innocuous traffic irregularities or other false alarms. But, more harm might be done the longer it takes to recognize a problem. An IIS is very valuable for network monitoring, but how beneficial it is on what you do with the data it provides. Tools for detection are inefficient at adding a layer of security since they don't prevent or fix problems, thus you need the correct staff and policies to manage them and respond to threats. Intruders can utilise encrypted packets to enter the network since an IIS cannot see inside of them. Your systems remain exposed until the intrusion is found since an IIS won't record these intrusions until they are farther into the network. When encryption is used increasingly often to protect personal data, this is a major worry. An important drawback of an IIS is that it frequently alerts you to false positives. False positives are frequently more common than real dangers in many situations. Your engineers will still have to spend time responding to false positives even if an IIS is set to limit their frequency [13]. Real assaults may go undetected or be overlooked if they don't take care to watch out for the false positives.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

International Conference on Muti-Disciplinary Application & Research Technologies (ICMART-2023)

Geetanjali Institute of Technical Studies

Vol. 10, Special Issue 2, May 2023

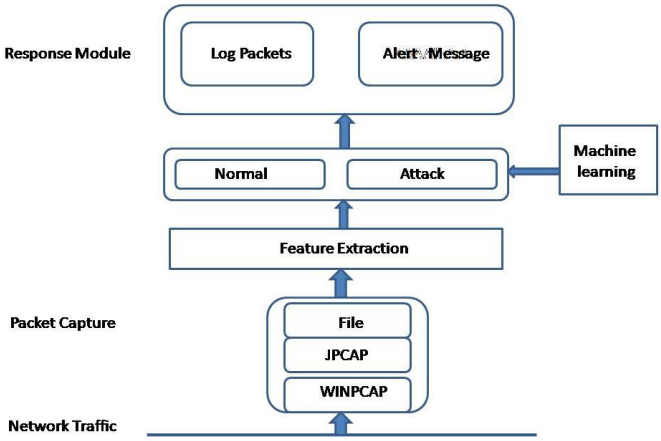

Fig. 1 Architecture of the Proposed System

IV. PROPOSED SYSTEM

A cyber assault may be trained for and detected using machine learning techniques. An email notification can be issued to the security engineers or users as soon as the threat is discovered. Any classification technique may be used to determine whether or not an attack is a DoS/DDoS. Support Vector Machine (SVM), a supervised learning technique that analyses data and finds patterns, is an example of a classification algorithm. Our greatest chance right now is early identification, which will help reduce the possibility of irreversible harm such attacks might do, because we cannot control when, when, or how an attack may come our way and absolute prevention against them cannot yet be guaranteed. To swiftly identify cyber-attacks and lessen their impact, businesses can use existing technologies or create their own. The ideal system would be one that requires little human involvement.

V. DESIGN OF THE SYSTEM

Applications can use machine learning algorithms to recognize and respond to cyber-attacks before they have an impact. Often, a model established by analyzing data sets of security events and determining the pattern of harmful behaviors is used to do this. As a consequence, similar acts are immediately addressed when they are discovered. Indicators of Compromise (IOC) that have already been discovered and documented make up the majority of the training dataset for the models [7], which are then used to create models and systems that can monitor, recognize, and react to threats in real time. Also, we can apply machine learning classification algorithms to recognize the various behaviors of malware in datasets and categories them appropriately thanks to the availability of IOC datasets. This makes it feasible to automate the process of finding and categorizing new viruses by using the learnt patterns. This can speed up the work of security analysts or other automated systems. Determine and categorize a new danger category, and then react to it using data-driven judgements. Applications can use machine learning algorithms to recognize and respond to cyber-attacks before they have an impact. Typically, a model created by evaluating data sets of security events and determining the pattern of harmful behaviors is used to do this. Real-time monitoring, threat detection, and response are all capabilities. Moreover, given the availability of IOC datasets, we may categories datasets using machine learning techniques to recognize the various behaviors of malware. This makes it feasible to automate the process of finding and categorizing new viruses by using the learnt patterns [9]. Security analysts or other automated systems may benefit from having easy access to Determine and categorize a new danger category, and then react to it using data-driven judgments. The volume of data travelling through a network at any particular time is referred to as network traffic. Most network data is contained in network packets, which support the network's load. The primary factor in network traffic measurement, network traffic control, and network traffic simulation is network traffic. In order to guarantee the quality of service in a particular network, network traffic should be organized properly.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

International Conference on Muti-Disciplinary Application & Research Technologies (ICMART-2023)

Geetanjali Institute of Technical Studies

Vol. 10, Special Issue 2, May 2023

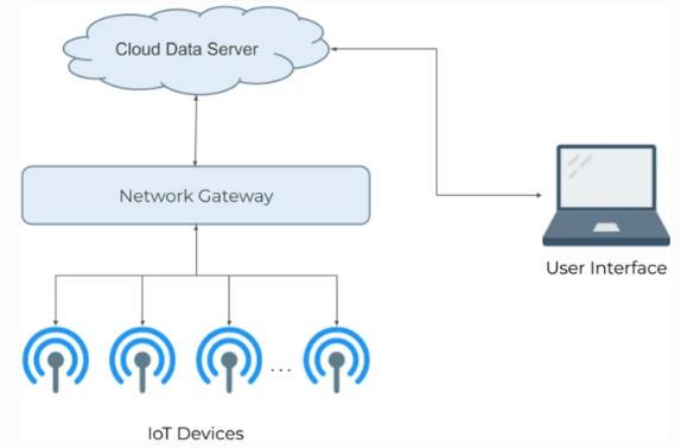

Fig. 2 Typical structure of IoT using Cyber Security

A. System

A system is an orderly group of interdependent components linked together according to a plan to achieve a specific objective. Its main characteristics are organization, interaction, interdependence, integration and a central objective. In a data network, intercepting a data packet while it is travelling through a particular location is known as packet capture [10]. A real-time packet is collected, held for a while so it can be examined, and then it may be downloaded, archived, or deleted. To identify and repair network issues like these, packets are recorded and inspected and resolving problematic network behaviours knowing when a network is congested Identifying data/packet loss Investigative network analysis. The system is constructed using the ANACONDA software, which is the most widely used data science platform in the world and the cornerstone of contemporary machine learning [11]. We established Python's usage for data science, support its active community, and continue to oversee open-source initiatives that enable tomorrow's discoveries. With the help of our enterprise-grade solutions, businesses, universities, and other organizations may use open source to their advantage and conduct ground-breaking research that will improve the world [13]. Delivering a package repository and strong open source tools in a coordinated, cooperative, and version-controlled environment providing the tools for auditing, versioning, and tracking data science activities It has been tested to automate model deployment and training on scalable container-based infrastructure.

B. Analysing the System

System analysis and design are the application of the system approach to problem solving generally using computers. To reconstruct a system the analyst must consider its elements output and inputs, processors, controls, feedback and environment. Another common supervisory machine learning job is characterization. Spam detection in cyber security is effectively carried out by ML-based classifiers, which determine if a certain email message is spam or not. Spam and non-spam communications can be distinguished using the spam filter models. Logistic regression, K-Nearest Neighbors, Support Vector Machine, Naive Bayes, Decision Tree, and Random Forest Classification are examples of machine learning approaches for classification [12]. Deep Learning classification models using Restricted Boltzmann Machines (RBM), Convolution Neural Networks (CNN), Recurrent Neural Networks (RNN), or Long-Short Term Memory (LSTM) cells for feature extraction followed by a densely connected neural network have become more effective in solving challenging tasks as a result of the availability of large collections of historical data with labels. The availability of huge amounts of labelled data is a prerequisite for the use of the supervisory machine learning algorithms mentioned above. The phrase "incidence response" is used to describe how an organization responds to a data breach or cyber-attack, including how it tries to control the fallout from the assault or breach (the "incident"). The ultimate objective is to successfully manage the event such that the harm is minimized, recovery time and expenses are kept to a minimum, collateral damage like brand reputation is maintained to a minimum and a warning message is sent to security analysts. CPU consumption, both overall and per user. Real-world and virtual memory use Amount of free RAM I/O and disc utilization Amount of swap space that is currently available.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

International Conference on Muti-Disciplinary Application & Research Technologies (ICMART-2023)

Geetanjali Institute of Technical Studies

Vol. 10, Special Issue 2, May 2023

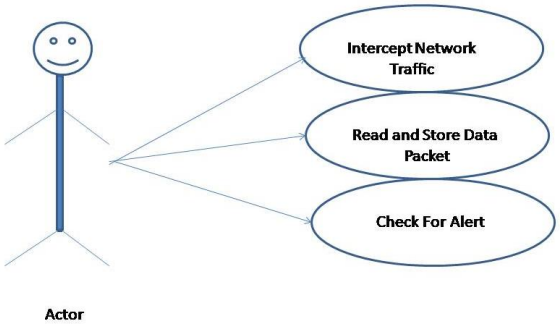

Fig. 3 Use case diagram of the Proposed System

VI. RESULTS AND DISCUSSIONS

Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. We reviewed several influential algorithms for intrusion detection based on various machine learning techniques.

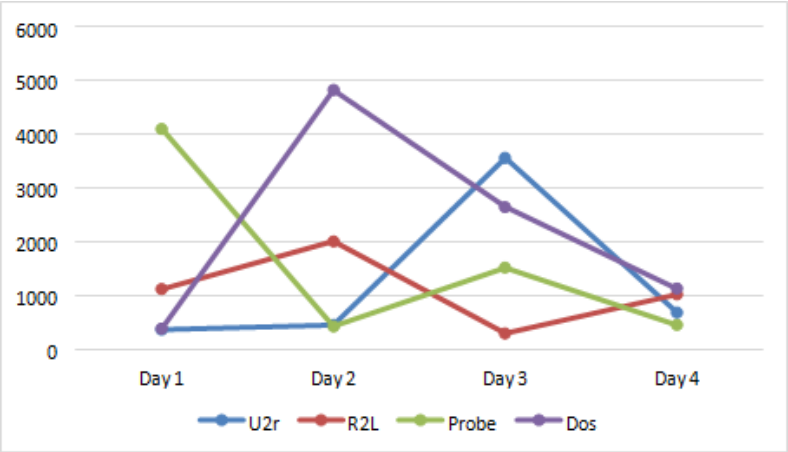

Fig. 4 Graphical analysis of attacks occurred on each attack Characteristics of ML techniques

Characteristics of ML techniques makes it possible to design IIS that have high detection rates and low false positive rates while the system quickly adapts itself to changing malicious behaviors. IIS using many Machine Learning Techniques like Random Forest, Decision tree and logistic regression to perform better in various metrics. The IDS should provide the most effective solutions based on the requirements. One thing is sure, any company failing to adopt these techniques now or in the immediate future risk compromising data or worse servers.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

International Conference on Muti-Disciplinary Application & Research Technologies (ICMART-2023)

Geetanjali Institute of Technical Studies

Vol. 10, Special Issue 2, May 2023

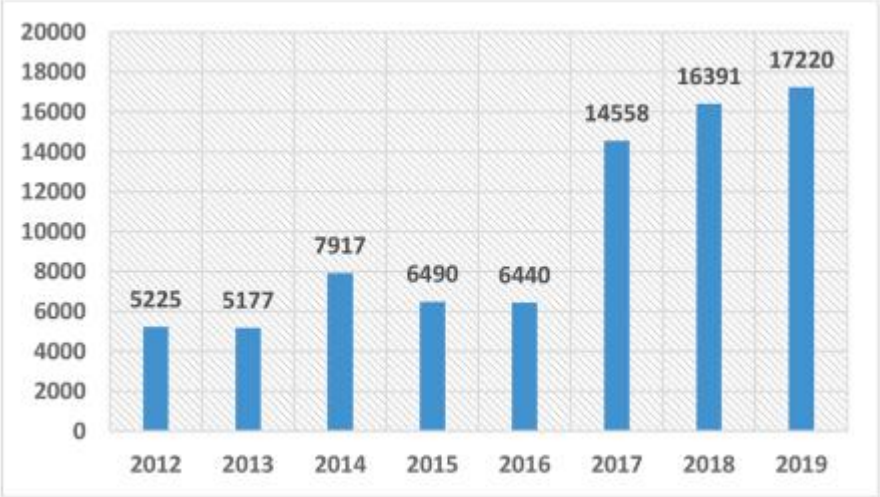

Fig. 5 Vulnerabilities in Cyber-security (in years)

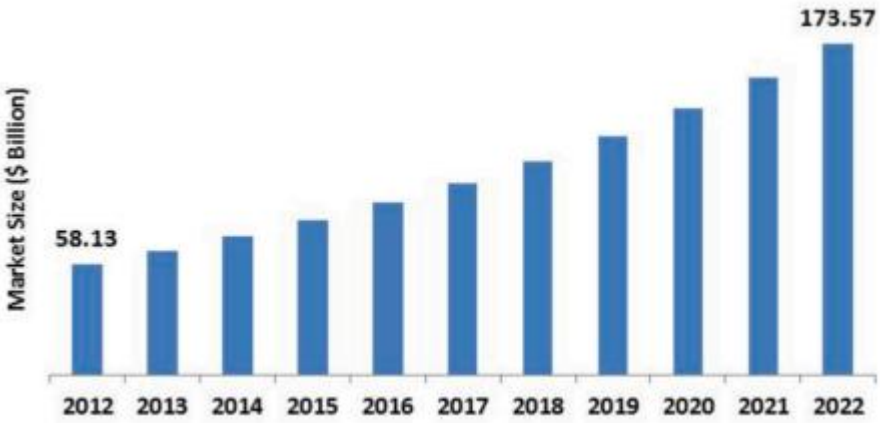

Fig. 5 Market in Cyber security

VII. CONCLUSION

Most techniques used in today's IDS are not able to deal with the dynamic and complex nature of cyber-attacks on computer networks. Hence, efficient adaptive methods like various techniques of machine learning can result in higher detection rates, lower false alarm rates and reasonable computation and communication costs. We reviewed several influential algorithms for intrusion detection based on various machine learning techniques. Characteristics of ML techniques makes it possible to design IIS that have high detection rates and low false positive rates while the system quickly adapts itself to changing malicious behaviors. IIS using many Machine Learning Techniques like Random Forest, Decision tree and logistic regression to perform better in various metrics. The IDS should provide the most effective solutions based on the requirements. One thing is sure, any company failing to adopt these techniques now or in the immediate future risk compromising data or worse servers.

REFERENCES

[1] H. Dai, "Imbalanced Protein Data Classification Using Ensemble FTM-SVM", IEEE Transactions on NanoBioscience, vol. 14, no. 4, pp. 350-359, 2015.

[2] V. López, A. Fernández, S. García, V. Palade and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics", Information Sciences, vol. 250, pp. 113-141, 2013.M.

Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in Proc. ECOC'00, 2000, paper 11.3.4, p. 109.

[3] T. Menzies, J. Greenwald and A. Frank, "Data Mining Static Code Attributes to Learn Defect Predictors", IEEE Transactions on Software Engineering, vol. 33, no. 1, pp. 2-13, 2007. (2002) The IEEE website. [Online].. Available: http://www.ieee.org/

[4] K. C. Giri, M. Patel, A. Sinhal and D. Gautam, "A Novel Paradigm of Melanoma Diagnosis Using Machine Learning and Information Theory," 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2019, pp. 1-7, doi: 10.1109/ICACCE46606.2019.9079975.

[5] V.K. Maurya, R.M. Mehra and A. Mehra, "Design and Analysis of Energy Efficient OPAMP for Rectifier in MicroScale Energy Harvesting (Solar Energy)", In: Satapathy, S., Bhatt, Y., Joshi, A., Mishra, D. (eds) Proceedings of the International Congress on Information and Communication Technology. Advances in Intelligent Systems and Computing, vol 439. Springer, Singapore.

[6] G. Huang, S. Song, J. Gupta and C. Wu, "Semi-Supervised and Unsupervised Extreme Learning Machines", IEEE Transactions on Cybernetics, vol. 44, no. 12, pp. 2405-2417, 2014.

[7] D. Kothari, M. Patel and A. K. Sharma, "Implementation of Grey Scale Normalization in Machine Learning & Artificial Intelligence for Bioinformatics using Convolutional Neural Networks," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 1071-1074, https://doi.org/10.1109/ICICT50816.2021.9358549

[8] R. Kothari, N. Choudhary and K. Jain, ``CP-ABE Scheme with Decryption Keys of Constant Size Using ECC with Expressive Threshold Access Structure``, Studies in Autonomic, Data-driven and Industrial Computing, pp. 15-36, 2021.

[9] G. Kaur and H. Singh, "Data Mining Techniques for Text Mining", Indian Journal of Science and Technology, vol. 9, no. 44, 2016.

[10] Sen, S., Patel, M., Sharma, A.K. (2021). Software Development Life Cycle Performance Analysis. In: Mathur, R., Gupta, C.P., Katewa, V., Jat, D.S., Yadav, N. (eds) Emerging Trends in Data Driven Computing and Communications. Studies in Autonomic, Data-driven and Industrial Computing. Springer, Singapore. https://doi.org/10.1007/978-981-16-3915-9_27

[11] F. Khan, R. Kothari, and M. Patel, "Advancements in Blockchain Technology With the Use of Quantum Blockchain and Non-Fungible Tokens." In Advancements in Quantum Blockchain With Real-Time Applications, pp. 199-225. IGI Global, 2022.

[12] F. Khan, R. Kothari, M. Patel and N. Banoth, "Enhancing Non-Fungible Tokens for the Evolution of Blockchain Technology," 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2022, pp. 1148-1153, doi: https://doi.org/10.1109/ICSCDS53736.2022.9760849

[13] H. Li, F. Chung and S. Wang, "A SVM based classification method for homogeneous data", Applied Soft Computing, vol. 36, pp. 228-235, 2015.

[14] H. Parvin, B. Minaei-Bidgoli and H. Alinejad-Rokny, "A New Imbalanced Learning and Dictions Tree Method for Breast Cancer Diagnosis", Journal of Bionanoscience, vol. 7, no. 6, pp. 673-678, 2013.