IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Exploring the Impact of Socioeconomic Factors on Cybercrime Rate Prediction

[1]Deepak Yadav, [2]Dr. Nitesh Kaushik, [3]Nishant Soni, [4]Ankit Saini, [5]Sachin Mahawar

[1,3,4,5] Student, Department of Computer Science and Engineering, Anand International College of Engineering, Jaipur, Rajasthan, India.

[2] Professor, Department of Computer Science and Engineering, Anand International College of Engineering, Jaipur, Rajasthan, India.

Abstract: The continuous provision of services in organizations through the IoT (Internet of Things) has made it a new gateway for cyber-attacks. The risks of software piracy and malware attacks are high, which can lead to economic and reputational damages due to the theft of important information. Cybercrime is a growing concern in the cyber age, and classification methods like support vector machine (SVM) and K-nearest neighbour (KNN) are used to classify cybercrime data. Unsupervised classification methods include K-means clustering, Gaussian mixture model, and cluster quasi-random via fuzzy C-means and fuzzy clustering. Neural networks are used to determine synthetic identity theft. Cybercrime detection uses datasets from CBS open data StatLine, with personal characteristics of crime victims. Different training and testing data are analysed for performance. The best technique is used to identify criminals, and the Gaussian mixture model in the unsupervised method shows enhanced performance. The accuracy of the detection method is 76.56%, while the SVM classifier achieves 89% accuracy. Performance metrics include true positive, false positive, false negative, false alarm rate, detection rate, accuracy, recall, precision, specificity, sensitivity, and Fowlkes-Mallows scores.[1]-[2] The expectation-maximization (EM) algorithm is used to assess the performance of the Gaussian mixture model. In this paper, we present a distributed framework based on deep learning that can detect and classify malicious traffic to ensure the security of IoT systems. We evaluate two different DL models, feed forward neural network and long short-term memory, using two different datasets (NSL-KDD and BoT-IoT) in terms of performance and identification of different kinds of attacks.

Index: Internet of Things, data mining, cyber security, software piracy, malware detection, Attack detection, Cyber-security, Deep learning, Distributed framework, Feed forward neural network, Long short-term memory, K-nearest neighbour, Cybercrime data, K-means clustering, Cyber age, CBS.

I. INTRODUCTION

IoT is the interconnection of actual moving articles "Things" through web implanted with an electronic chip, sensors, and different types of equipment. Every gadget is remarkably distinguished all around the world by Radio Recurrence Identifier (RFID) labels. Cybercrime is a growing concern in the cyber age, necessitating a focus on access control and detecting cyber users. Researchers are using machine learning models to classify data and predict classes based on features. The wafer might duplicate the rationale of the first programming by figuring out techniques and afterward plan a similar rationale in one more kind of source codes [4], [5]. It is an extreme danger to web security, which gives admittance to limitless downloads of pilfered programming, open-source codes and, advances. This work is authorized under an Inventive Lodge Attribution and promotes of pilfered variants [7]. It quickly builds every year and gives significant financial misfortune to the product business. The Business Programming Collusion (BSA) 2016 report expressed that the public programming robbery proportion is roughly 39%, which brings about business harms up to 52.2$ billion consistently. Many sorts of examination have shown that each product contains counterfeited source codes with regards to rationale with a scope of 5% to 20% [8], [9]. The clever programming counterfeiting strategies are expected to get the appropriated source code in pilfered programming. Different counterfeiting discovery strategies are

proposed, i.e., clone recognition, source code similitude distinguishing proof, programming bugs examination, and programming skin pigmentation examination [10], [11].

These procedures chiefly are construction and text-based examination. Malignant way of behaving can be seen by capability calls, capability boundaries' investigation, information stream, guidance follows, and visual examination of codes. This technique is additional tedious because of observing each unique way of behaving of source codes [12]-[14]. The static malware investigation techniques need not bother with the continuous execution of source codes. It could be utilized to catch the design data of malware parallels. The mark-based malware recognizable proof procedures are static based, i.e., control stream diagram, opcode recurrence, n-gram, and string mark. The dismantling devices, i.e., IDA Ace and OllyDbg[15], are applied to reveal the executables prior to carrying out static based calculations. The practical call chart is a static investigation strategy used to remove the underlying examination of codes [16].

These days, IoT networks involve billions of brilliant gadgets that impart with one another requiring negligible human mediation. It is assessed that the number of IoT gadgets will fill in 2022, anticipating 46 billion gadgets before the current year's over. IoT advancements are critical brilliant applications inside the shrewd business and savvy urban communities. In 2019, an organization of honeypots set up by Kaspersky found that notwithstanding the low complexity of the assaults, they stayed unseen by the client until the casualty was enacted as a feature of a botnet. In 2020, the main digital assaults in IoT were worms, bots, and Disperse Disavowal of Administration (DDoS), with upwards of 16 different types [17]. While in 2021, IoT digital goes after dramatically increased by Kaspersky. The ongoing restrictions of IoT gadgets draw in malevolent entertainers towards the IoT biological system. With this new IoT worldview the regular security objectives: confidentiality, respectability, and accessibility (CIA) are not sufficient and bomb in tending to novel dangers [18]. The new security necessities to consider are: responsibility, auditability, protection, and reliability. Conventional network safety frameworks safeguard clients and gadgets through Interruption Location Frameworks (IDS) client frameworks forestalling assaults by recognizing designs that are different from ordinary way of behaving [23]. Cyberattacks share a typical component with picture acknowledgment, since over close to 100% of the new goes after are little freaks of existing ones; similarly, that adjustments of pictures can be identified by little changes in their pixels. In IoT-Mist networks, they are utilized for identifying network strings and assaults [24,25]. Even though IoT network highlights (i.e., its conveyed nature and the restricted figuring abilities of the end-gadgets) requires novel answers for IDS [26].

This work proposes an original solid framework which applies DL ways to deal with focus on a few sorts of assaults in IoT organizations. The commitments of this work are as follows:

- Plan and improvement of a clever circulated DL-based assault identification structure in IoT organizations.
- Preprocessing of the BoT-IoT and NSL-KDD datasets to accomplish a higher exactness of the system.
- Correlation of FFNN and LSTM models to choose the best model for a wide scope of digital assaults.
- DL model tuning utilizing Hyperband to further develop location rates.

MALWARE DETECTION

The regular strategies might address code muddling con-corns, however high computational expense is required with respect to surface component mining utilizing malware representations. These kinds of component extraction strategies do not perform well with broad malware information investigation. Right now, malware is con-tenuously producing, update, and control that makes the identification really testing [29], [30].

The proposed malware discovery technique attempts to answer the accompanying questions:

- How to recognize malware with diminished above?
- How to separate malware highlights with less computational expense?
- How to process enormous malware datasets to get better precision?

II. LITERATURE REVIEW

ML is an AI subset that can solve design problems without the need for explicit programming. It offers advantages such as reliability, controllability, and discernibility in everyday life. [31]ML strategies include numerical models, information acquisitions, heuristic learning, and decision trees for performing choosing, providing stability, controllability, and discernibility. In the clinical field, ML models help clinical experts with early phase indications for finding human diseases. [32]ML has been particularly beneficial in the system administration field, such as offloading plan and activity of communication networks on machines. Effective execution of The clever programming counterfeiting strategies are expected to get the appropriated source code in pilfered programming. Different counterfeiting discovery strategies are proposed, i.e., clone recognition, source code similitude distinguishing proof, programming bugs examination, and programming skin pigmentation examination [10], [11].

These procedures chiefly are construction and text-based examination. Malignant way of behaving can be seen by capability calls, capability boundaries' investigation, information stream, guidance follows, and visual examination of codes. This technique is additional tedious because of observing each unique way of behaving of source codes [12]-[14]. The static malware investigation techniques need not bother with the continuous execution of source codes. It could be utilized to catch the design data of malware parallels. The mark-based malware recognizable proof procedures are static based, i.e., control stream diagram, opcode recurrence, n-gram, and string mark. The dismantling devices, i.e., IDA Ace and OllyDbg[15], are applied to reveal the executables prior to carrying out static based calculations. The practical call chart is a static investigation strategy used to remove the underlying examination of codes [16].

These days, IoT networks involve billions of brilliant gadgets that impart with one another requiring negligible human mediation. It is assessed that the number of IoT gadgets will fill in 2022, anticipating 46 billion gadgets before the current year's over. IoT advancements are critical brilliant applications inside the shrewd business and savvy urban communities. In 2019, an organization of honeypots set up by Kaspersky found that notwithstanding the low complexity of the assaults, they stayed unseen by the client until the casualty was enacted as a feature of a botnet. In 2020, the main digital assaults in IoT were worms, bots, and Disperse Disavowal of Administration (DDoS), with upwards of 16 different types [17]. While in 2021, IoT digital goes after dramatically increased by Kaspersky. The ongoing restrictions of IoT gadgets draw in malevolent entertainers towards the IoT biological system. With this new IoT worldview the regular security objectives: confidentiality, respectability, and accessibility (CIA) are not sufficient and bomb in tending to novel dangers [18]. The new security necessities to consider are: responsibility, auditability, protection, and reliability. Conventional network safety frameworks safeguard clients and gadgets through Interruption Location Frameworks (IDS) client validation, information encryption, firewalls, and hostile to infection programming. As displayed in Kwon etc. [19].; the utilization of AI (ML) methods to distinguish pernicious organization traffic [20,21], unusual ways of behaving and endeavours in PC frameworks in an IDS is not sufficient. However, exemplary ML needs programmed highlight designing [22], they have low recognition rate, and they are not efficient in identifying little variations of existing assaults. This has prompted consider DL procedures to improve digital protection frameworks.

DL is a ML sub-field that has earned extraordinary respect in numerous areas due to its improvement in precision in complex assignments and late advancements in equipment and programming. DL procedures improve network protection direction, frequency task, and survivability.[34] The EON methodology has been developed as an innovative optical determination fit for dealing with high versatility requirements in managing optical system assets.

The research aims to introduce a speedy use of ML in optical systems administration [35], including a basic instructional exercise on ML techniques and their execution, an overview of momentum research work, and a list of different applications. Both optical communication and optical system administration are considered for generating new cross-layer research directions.

The choice of the suitable DL model.

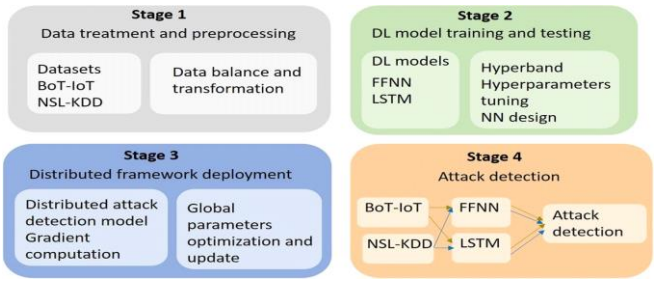


Figure 1. Attack detection framework stages

The two models considered are the FFNN and LSTM because they are utilized for managed learning and they have great sending with exceptionally associated features. The two organizations are assessed in a unified model. Then, at that point, the best performing DL model is utilized in the conveyed system. At last, the proposed disseminated system is introduced, and its presentation surveyed in a sensible scatter environment where information is restricted.

III. CRIME ANALYSIS

- DATA SETS

This paper expects to give a strong structure powerful in the recognition of IoT digital assaults. In this manner, we utilize both the BoT-IoT (a particular IoT digital assaults dataset) and NSL-KDD (general digital assault dataset).

- BIO-IOT DATASET

The BoT-IoT dataset [37] comprises of north of 73 million records of organization action in a mimicked IoT climate, including both ordinary and a few digital assault traffic streams. The preparation and test dataset have 5 result classes reflecting either typical traffic or 4 distinct kinds of assaults: DDoS, Disavowal of Administration (DoS), Keylogging and Information Robbery. DoS and DDoS assaults address pernicious

- DATA PROCESSING

With respect to content of the dataset records, both datasets are lopsided since the quantity of records is far higher for typical traffic than for assaults. Different techniques can be utilized to adjust the information in view of oversampling and under sampling methods. The enormous number of records lead us to consider under sampling techniques. The datasets are adjusted diminishing the quantity of typical traffic records (arbitrarily picked), while keeping up with the quantity of assaults. Subsequently, we integrated a subset of the dataset holding an equality of half assaults and half not assaults. as displayed in Figure 1. Four data sets are executed in distributed storage to process the malware doubles and delicate product pilfered records. The crude organization traffic information is put away in data set 1, and the subsequent data set contains a rundown of prior malware information. Further, the third information base is utilized to store new marks of recognized malware assaults. The wafer stores pilfered programming in data set 4 through IoT gadgets. It functions as a stockpiling place for pilfered duplicates where saltine attempts to spread these duplicates by IoT organization. This enormous measure of information needs high handling, for example time and cost. The principal information base sent crude information to the pre-handling module. It preprocesses the crude information and catches valuable highlights. Pre-handled information is additionally submitted to the recognition mod-ule. The recognition module catches malware and pilfered delicate product assaults by gaining from the marks in data sets two and four. On the off chance that any noxious movement is seen in the organization, the proposed framework cautions the executive for appropriate activity.

DISTRIBUTED DL-BASED ATTACK DETECTION FRAMEWORK

These days, IoT networks need a circulated arrangement at the edge for digital assault recognition. As a matter of fact, the circulated idea of IoT conditions require re-designing of interruption identification administrations. Conventional unified IDSs have demonstrated to be inef-fective in the avoidance of novel (zero-day) assaults. This segment presents a clever conveyed DL-based assault identification structure. It comprises of four fundamental stages: Information treatment and pre-handling, DL model preparation and testing, appropriated system arrangement and assault discovery

endeavours to upset nor-mal traffic of a server, administration, or organization over-burdening them with a surge of Web traffic. Keylogging address goes after that accumulate data. At long last, information robbery rep-loathes private client data spill.

The BoT-IoT dataset depends on the movement of an organization made by 62 hosts (situated in the organization cover 192.168.100.0/26). Past the situation introduced in the BoT-IoT dataset, in this paper the organization traffic is considered for any organization size since the augmentation of IoT gadgets and the presence of new organizations types (for example 5G) produces numerous conceivable organization sizes and along these lines' situations. Every one of the various assaults of the dataset depend on flooding. This segment first presents the datasets considered, as well as the pre-handling led. To accomplish an expansive arrangement, two distinct datasets are consid-ered: the BoT-IoT dataset that tends to assaults explicit of IoT conditions, and the NSL-KDD dataset to expand the kinds of digital assaults. Both datasets are pre-handled to be subsequently utilized in the NNs. The second piece of this segment resolves a central point of contention for the proposed system, that is

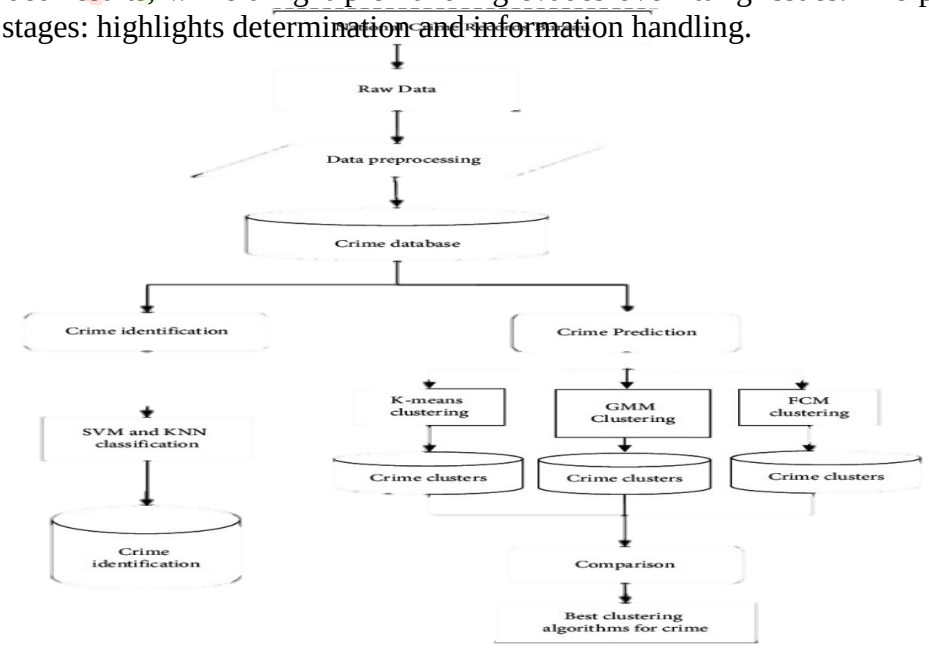
Thusly, the more important elements are those which give data about the rates, the quantity of bundles and kinds of convention utilized by the IoT gadget.

- CLASSIFICATION

The dataset is divided into several groups depending on some specific attributes of the data object. Based on the states and cities, the crime can be grouped. The process of classification involves classifying the crime depending on the different types of crime. Using the K-means algorithm, data having similar attributes can be grouped or clustered.

- DATA PRE-PROCESSING

This segment presents the primary phase of the structure. Information pre-handling is per-shaped before the DL models are prepared. The suitable pre-handling of net-work traffic permits DL models proficiently foresee non-one-sided results, while a right pre-handling evades overfitting issues. The pre-handling is partitioned in two stages: highlights determination and information handling.



- DATA COLLECTION

Police records contain vast amounts of crime data, which are recorded annually by the National Crime Bureau of Records. However, these records are often unprocessed and contain incorrect or missing values. Preprocessing, which involves cleansing and preprocessing, is crucial to rectify these issues and ensure the proper form of the collected data.

REFERENCES

1. F. Ullah, J. Wang, M. Farhan, M. Habib, and S. Khalid, ‘‘Software plagiarism detection in multiprogramming languages using machine learningapproach,’’ Concurrency Comput., Pract. Exper., to be published.
2. D.-K. Chae, J. Ha, S.-W. Kim, B. Kang, and E. G. Im, ‘‘Software plagiarismdetection: A graph-based approach,’’ in Proc. 22nd ACM Int. Conf. Inf.Knowl. Manage., Nov. 2013, pp. 1577–1580.
3. Y. Akbulut and O. Dönmez, ‘‘Predictors of digital piracy among Turkishundergraduate students,’’ Telematics Inform., vol. 35, no. 5, pp. 1324–1334,2018.
4. M. ShanmughaSundaram and S. Subramani, ‘‘A measurement of similarityto identify identical code clones,’’ Int. Arab J. Inf. Technol., vol. 12,pp. 735–740,Dec. 2015.
5. C. Ragkhitwetsagul, ‘‘Measuring code similarity in large-scaled codeCorpora,’’ in Proc. IEEE Int. Conf. Softw. Maintenance Evol. (ICSME),Oct. 2016, pp. 626–630.
6. S. Imran, M. U. G. Khan, M. Idrees, I. Muneer, and M. M. Iqbal,‘‘An enhanced framework for extrinsic plagiarism avoidance for researcharticle,’’ Tech. J., vol. 23, no. 01, pp. 84–92, 2018.
7. A. Shabtai, R. Moskovitch, Y. Elovici, and C. Glezer, ‘‘Detection ofmalicious code by applying machine learning classifiers on static features:A
8. M. Egele, T. Scholte, E. Kirda, and C. Kruegel, ‘‘A survey on automateddynamic malware-analysis techniques and tools,’’ ACM Comput. Surv.,vol. 44, no. 2, p. 6, Feb. 2012.
9. I. Ghafir, J. Saleem, M. Hammoudeh, H. Faour, V. Prenosil, S. Jaf,S. Jabbar, and T. Baker, ‘‘Security threats to critical infrastructure: Thehuman factor,’’ J. Supercomput., vol. 74, no. 10, pp. 4986–5002, Oct. 2018.
10. I. Raz, ‘‘Introduction to reverse engineering,’’ Cs. tau. ac. il, 2011.
11. E. Gandotra, D. Bansal, and S. Sofat, ‘‘Malware analysis and classification:A survey,’’ J. Inf. Secur., vol. 5, no. 2, p. 56, 2014.
12. Lab, W.T.: Internet security report: WatchGuard’s threat lab analyzes the latest malware and internet attacks. Technical report (Q3 2020). https://www.watchguard.com/wgrd-resource-center/security-report-q3-2020. Accessed 17 Mar 2021
13. Cherdantseva, Y., Hilton, J.: A reference model of information assurance security. In: 2013 International Conference on Availability, Reliability and Security, pp. 546–555 (2013). https://doi.org/10.1109/ARES.2013.72
14. Kwon, D., Kim, H., Kim, J., Suh, S., Kim, I., Kim, J.: A survey of deep learning-based network anomaly detection. Clust. Comput. 22(5), 949–961 (2019)
15. Anderson, J., Carbonell, J., Mitchell, T., Michalski, R., Amarel, S., Tecuci, T., Kodratof, Y.: Machine Learning: An Artifcial Intelligence Approach. M. Kaufmann, Los Altos, CA (1983)
16. Kilincer, I., Ertam, F., Sengur, A.: Machine learning methods for cyber security intrusion detection: datasets and comparative study. Comput. Netw. 188, 107840 (2021). https://doi.org/10.1016/j.comnet.2021.107840

17. Fadlullah, Z.M., Tang, F., Mao, B., Kato, N., Akashi, O., Inoue, T., Mizutani, K.: State-of-theart deep learning: evolving machine intelligence toward tomorrow's intelligent network trafc control systems.

18. IEEE Commun. Surv. Tutor. 19(4), 2432–2455 (2017). https://doi.org/10.1109/COMST.2017.2707140

19. Tsimenidis, S., Lagkas, T., Rantos, K.: Deep learning in IoT intrusion detection. J. Netw. Syst.Manag. (2022). https://doi.org/10.1007/s10922-021-09621-9

20. L. Nataraj, S. Karthikeyan, G. Jacob, and B. S. Manjunath, ''Malwareimages: Visualization and automatic classification,'' in Proc. 8th Int. Symp.Vis. Cyber Secur., Jul. 2011, p. 4.

21. Diro, A.A., Chilamkurti, N.: Distributed attack detection scheme using deep learning approach for Internet of Things. Futur.

22. A. Makandar and A. Patrot, ''Malware class recognition using image processing techniques,'' in Proc. Int. Conf.Data Manage., Anal. Innov. (ICDMAI), Feb. 2017, pp. 76–80

23. H.-I. Lim, H. Park, S. Choi, and T. Han, ''A method for detecting the theftof Java programs through analysis of the control flow information,'' Inf.Softw. Technol., vol. 51, no. 9, pp. 1338–1350, 2009.

24. J. Yasaswi, S. Kailash, A. Chilupuri, S. Purini, and C. V. Jawahar,''Unsupervised learning based approach for plagiarism detection in programming assignments,'' in Proc. 10th Innov.Softw. Eng. Conf., Feb. 2017,pp. 117–121.

25. V. Kashyap, D. B. Brown, B. Liblit, D. Melski, and T. Reps,''Source forager: A search engine for similar source code,'' Jun. 2017,arXiv:1706.02769. [Online]. Available: https://arxiv.org/abs/1706.02769

26. F. Zhang, D. Wu, P. Liu, and S. Zhu, ''Program logic based softwareplagiarism detection,'' in Proc. IEEE 25th Int. Symp. Softw. Rel. Eng.,Nov. 2014, pp. 66–77.

27. Rege M., Mbah R. B. K. Machine Learning for Cyber Defense and Attack. Proceedings of the Data Analytics: The Seventh International Conference on Data Analytics; November 2018; Athens, Greece. pp. 73–78. [Google Scholar] [Ref list]

28. Bharati A., Sarvanaguru R. A. Crime prediction and analysis using machine learning. *International Research Journal of Engineering and Technology* . 2018;5(9):1037–1042. [Google Scholar] [Ref list]

29. Marsland S. *Machine Learning: An Algorithmic Perspective"* . FL, USA: CRC Press; 2015. pp. 1–430. [Google Scholar] [Ref list]

30. Bkassiny M., Li Y., Jayaweera S. K. A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys & Tutorials* . 2013;15(3):1136–1159.doi: 10.1109/surv.2012.100412.00017. [CrossRef] [Google Scholar] [Ref list]

31. Mukherjee. *Optical WDM Networks"* . Berlin, Grermany: Springer Science & Business Media; 2017. [Google Scholar] [Ref list]

32. https://doi.org/10.1162/neco.1997.9.8.1735

33. Gener. Comput. Syst. 82, 761–768 (2018). https://doi.org/10.1016/j.future.2017.08.043

34. Roopak, M., Yun Tian, G., Chambers, J.: Deep learning models for cyber security in IoT networks. In: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), pp. 452–457 (2019). https://doi.org/10.1109/CCWC.2019.8666588

35. Shalaka, M., Pawar, P.M., Muthalagu, R.: Efcient Intelligent Intrusion Detection System for Heterogeneous Internet of Things (HetIoT). J. Netw. Syst. Manag. (2023).

36. Moore, Intellectual Property and Information Control: PhilosophicFoundations and Contemporary Issues. Abingdon, U.K.: Routledge, 2017. S. Malabarba, P. Devanbu, and A. Stearns, ''MoHCA-Java: A tool for C++to Java conversion support,'' in Proc. Int. Conf. Softw. Eng., May 1999,pp. 650–653.