

Accurate Monitoring of Cyber Attacks and User Acknowledgment Using Machine Learning

V. Ramya¹, V. Lavanya², Dr. Srinivasulu Manda³, T. Jayasri⁴, K. Vijaya Kumar⁵

^{1, 2, 3, 4, 5}Department of Computer Science and Engineering

^{1, 2, 3, 4, 5}QIS College of Engineering and Technology, Ongole, Andhra Pradesh, India

¹ramya.v@qiscet.edu.in, ²lavanya.v@qiscet.edu.in, ³srinivasulu.m@qiscet.edu.in

⁴jayasri.t@qiscet.edu.in, ⁵vijaykumar.k@qiscet.edu.in

Corresponding Author Mail: qispublications@qiscet.edu.in

Article Info

Page Number: 460 - 469

Publication Issue:

Vol 70 No. 2 (2021)

Article History

Article Received:

05 September 2021

Revised: 09 October 2021

Accepted: 22 November 2021

Publication: 26 December 2021

Abstract

Cyber-Attackers capture internet users and businesses in order to steal vital information. Attackers gain access to sensitive information on company computers, including login details or credit card details and bank account numbers. Phishing attacks are one type of cyber-attack in which hackers fool internet users into thinking their websites are legitimate in order to steal their non-public information. In malware attacks, attackers secretly install harmful Programs on company servers or user computers through internet, then proceeds to steal every piece of data stored on that server or computer. Attacks by malware are getting increasingly frequent. A network intrusion is an attack where the attacker intends to steal every resource from the network. Heuristic and visual similarity-based approach, whether blacklist or whitelist, **Keywords:** Cyber-attacks, Intrusion detection system, HTTP sites, attackers.

1) INTRODUCTION

Cyber attackers seize businesses and internet users in order to access unauthorized data. Credit card details and bank account numbers are among the sensitive data that hacker's access on company computers. Cybercriminals may use phishing attacks to Internet users believe their websites are genuine, but they actually access unauthorized data. In a malware assault, a hacker install a harmful application covertly on a user's computer or a business server through the internet, then continues to steal every piece of data kept in that computer or server. Malware attacks are occurring more frequently. A network incursion is an attack where the goal is to take all network resources. Techniques based on visual resemblance and heuristics, whether blacklist or RELATED WORKS

Several techniques are applied to detect cyber attacks

- Discovery method
- Visual Similarity Based Approach
- Machine-learning Techniques
- Signature-detection
- Anomaly-methods

Disadvantages of the current system

- Not all phishing sites are guaranteed these qualities, so they are less accurate than list-based strategies
- Once an attacker is aware of the algorithm or attributes used to identify phishing sites, he can get around the heuristic protections and steal sensitive data.
- Why it takes extra time and complexity to compare a suspicious website with the whole valid database store.
- More room for storing trustworthy picture databases.
- The low level of similarity between an animated website and a phishing website results in a high false-negative rate. When a website's backdrop is slightly altered without otherwise straying from the aesthetic appearance of a legitimate site, this strategy fails. The enormous data sets won't allow these approaches to operate effectively.

2) PROPOSED SYSTEM ARCHITECTURE

To decrease the false positives in detecting cyber-attacks, incorporate new features with machine learning algorithms. An attempt was made to find the optimal machine learning algorithms to detect cyber threats more accurately than existing methods. Five machine learning methods to distinguish between real and fraudulent websites: Decision Tree, Support Vector Machine, Random Forest, Logistic Regression, and KNN.

Advantages of the Proposed System

The suggested system is made up of a dataset that includes specific details about various cyber-attacks that are useful for attack detection and forecasting. Since the suggested method uses a variety of machine learning algorithms, the results are chosen based on their level of accuracy. Metric values are generated for each algorithm that might be thought of as anticipated outcomes. Voting Classifier, an ensemble approach, is used to compare the models and provides highly accurate error-free results.

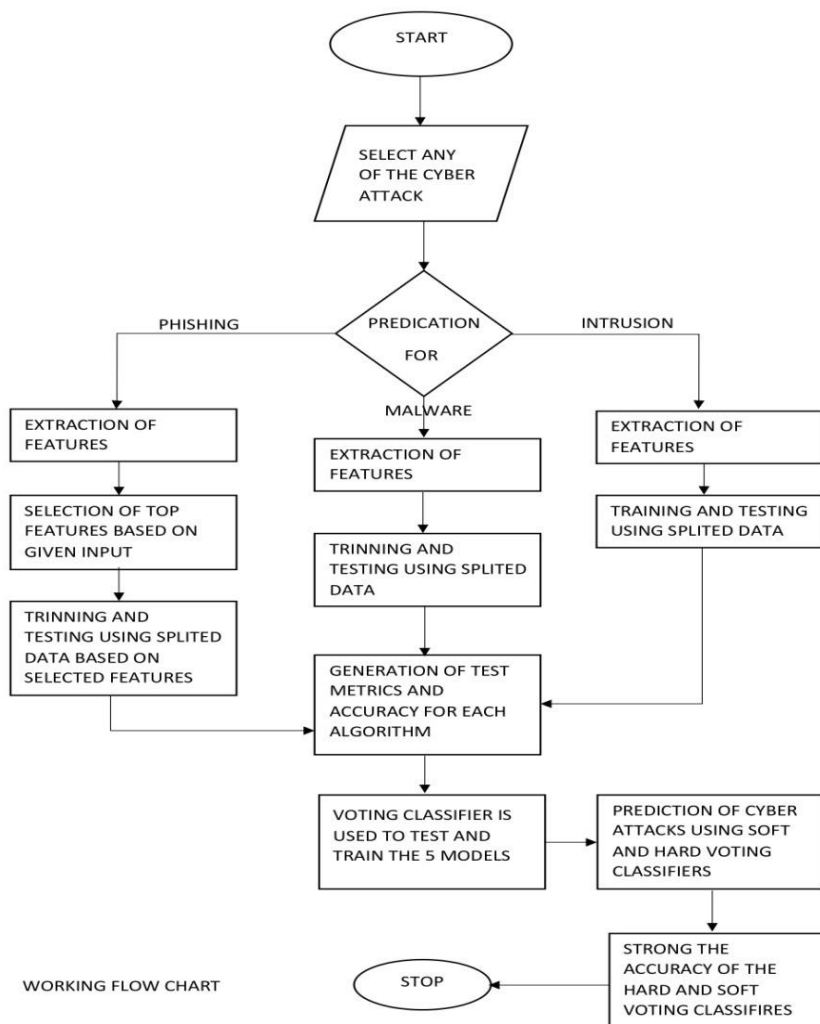


Figure 1: Flowchart

The system carries out the subsequent actions:

Step1: To start the procedure, choose one of the cyber-attacks, and it will proceed.

Step2: If the stated cyber-attack is a phishing attempt, the user must choose how many features to choose.

Step 3: The system uses a given dataset of cyber- attacks to train and test various machine learning algorithms to determine their accuracy.

Step 4: Voting classifier is currently used with 5modelstocompare all models to generate certainty.

DATASET PRE-PROCESSING

Any method of machine learning that involves data pre-processing modifies or encrypts the input to make it simpler for the computer to parse. In other words, the algorithm is now able to swiftly

recognize the features of the input. The term "dataset" refers to a grouping of data objects, which can also be called "records," "points," "vectors," "patterns," "occurrences," "instances," "samples," "observations," or "entities." A number of features that define data items can capture the fundamental characteristics of an entity, such as its mass or the precise instant at which an event occurred. The terms variables, characteristics, fields, attributes, or dimensions are frequently used to refer to feature. There are several The provided dataset contains a large number of features, and we will use the Feature Selection for Phishing dataset to only select the top features from the dataset. Due to the size of the datasets involved in malware detection and intrusion detection, data processing will be delayed. As a result, feature scaling is used to convert the dataset's values.

PHISHING DETECTION

Once data processing and feature selection are complete, To test and train the top features using five different types of algorithms, we take 75% of the dataset for training and 25% for testing. Algorithms generate and store accuracy and standard values.

INTRUSION & MALWARE DETECTION

Datasets are modified so that the values range from 0 to 1 due to the size of the datasets needed to detect intrusions and viruses. The Feature Scaling technique is used to accomplish this. Using 25% of the dataset for testing and 75% for training, the training and testing processes will start once the values have been converted. According to Figure 3, each of the five algorithms generates and stores accuracy and metric values.

VOTING CLASSIFIER

Testing is done using datasets using five different models and compared hard and soft voting based on probabilities.

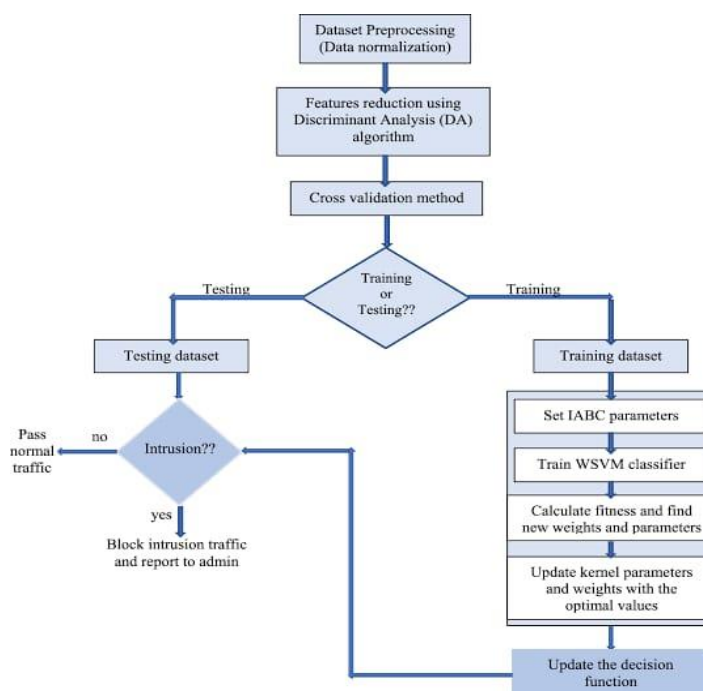


Fig. 2: Methodology of Intrusion Flowchart

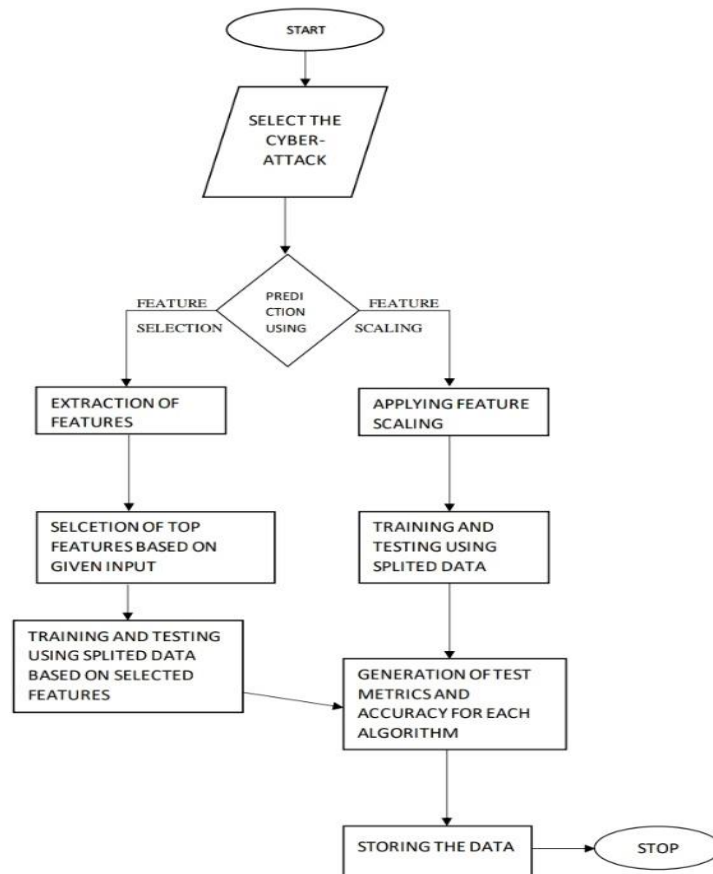


Fig. 3 Flowchart of Malware and Intrusion

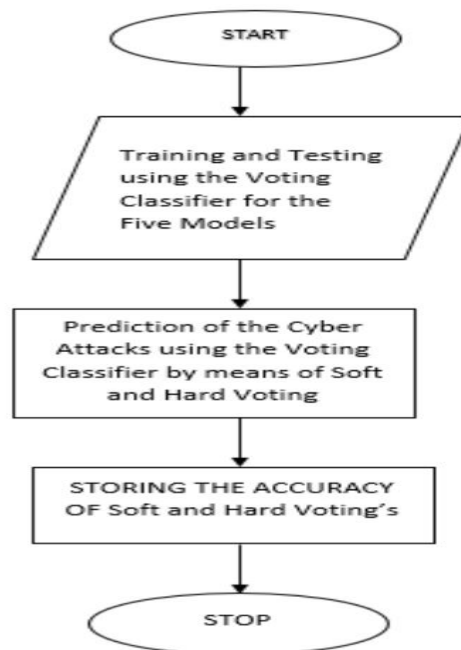


Fig.4 Flowchart of Voting Classifier

3) DISCUSSION AND RESULTS

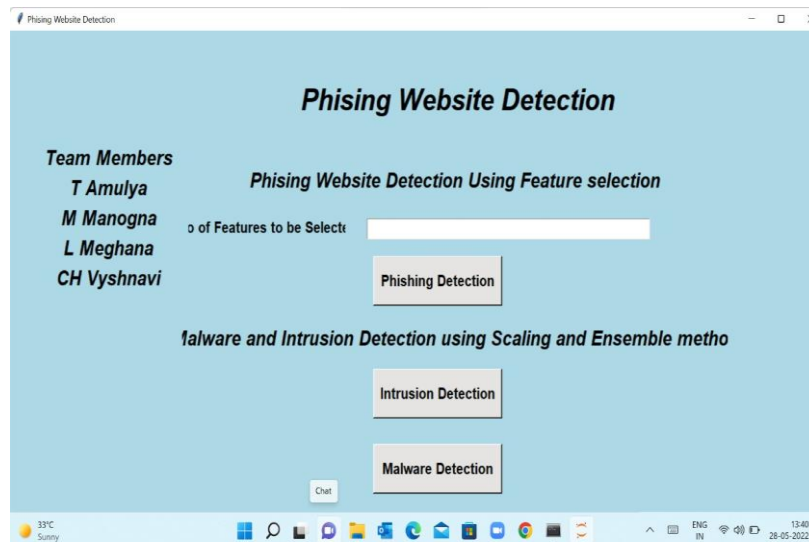


Fig.5: Phishing Website Detection first page

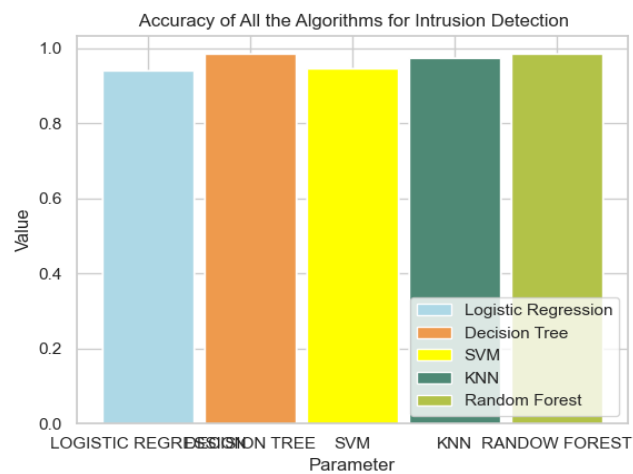


Fig.6 Average accuracy for Intrusion detection

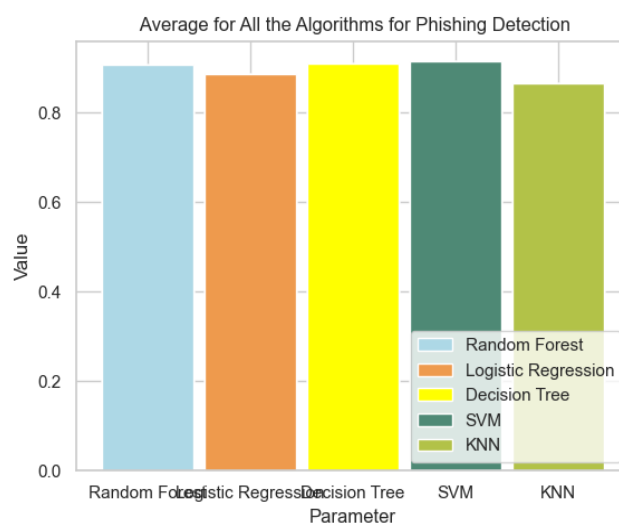


Figure. 7 Average phishing detection accuracy

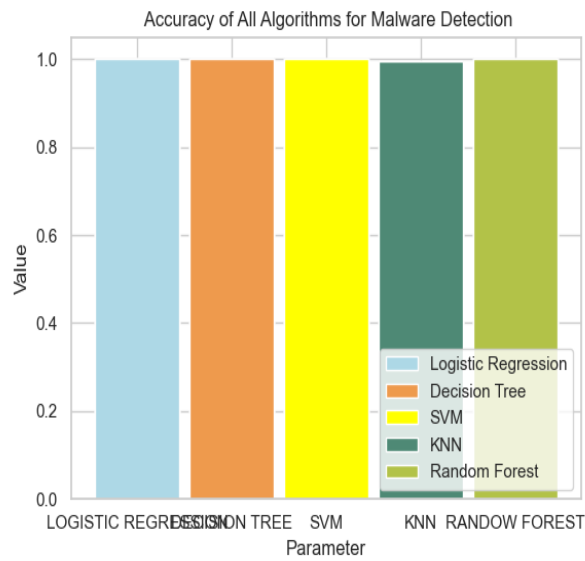


Figure. 8 Average malware detection accuracy

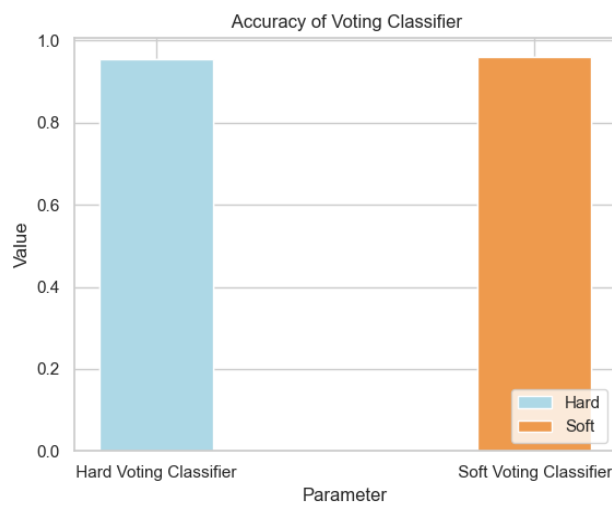


Figure 9. Reliability of Voting classifier,

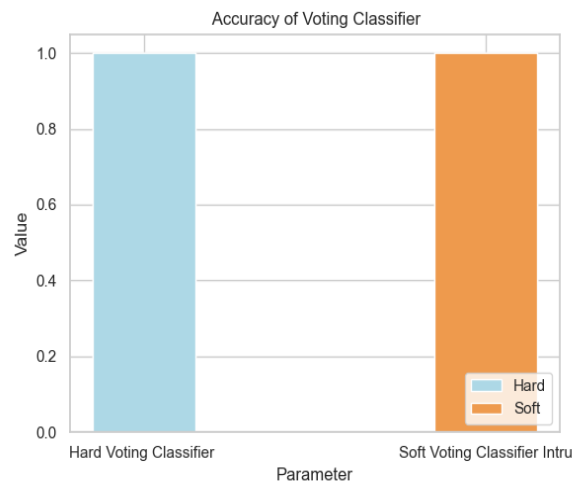


Figure. 10: Accuracy of a voting classifier for detecting phishing

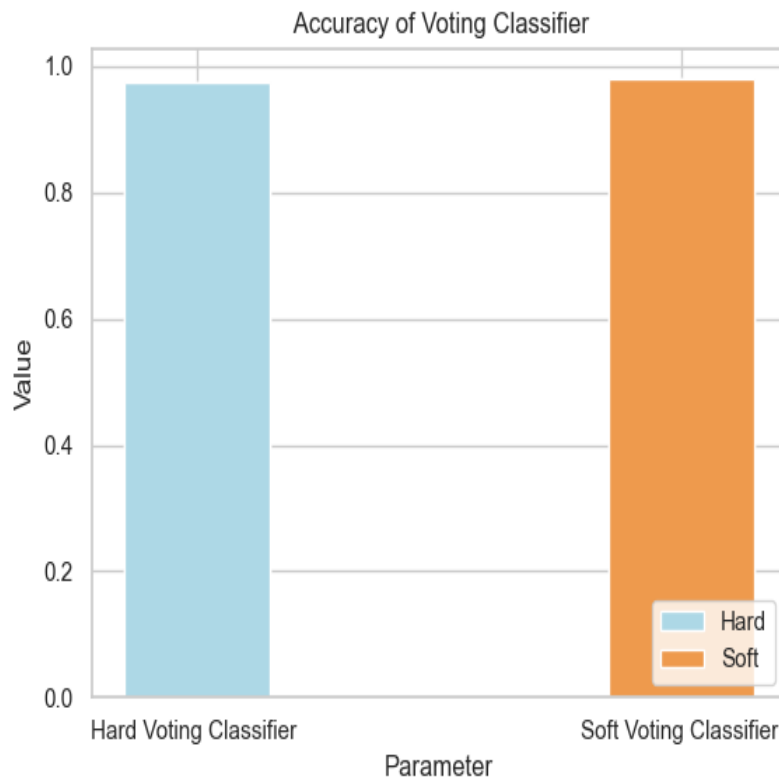


Figure 11. Reliability of intrusion detection

4) FUTURE SCOPE AND CONCLUSION

Cyber crime techniques such as phishing, malware, and intrusions use online services to steal people's personal information. To give the algorithm the best accuracy, the system uses phishing websites, malware detection, and the KDD dataset while applying several categorization techniques. Unsupervised machine learning algorithms identify cyber-attacks more effectively than supervised machine learning algorithms, so in the future we hope to use them to design and host websites.

REFERENCES

1. Somaiya Vidyaviharet.all, Phishing Website Detection using Machine Learning, IJCA, Volume 181-No.23, October 2018.
2. Sadeh N, Tomasic A, Fette I. Learning to detect phishing emails. Proceedings of the 16th international conference on World Wide Web. 2007: p. 649-656
3. Andr Bergholz, Gerhard Paa, Frank Reichartz, Siehyun Strobili, and Schlo Birlinghoven. Improved phishing detection using model-based features. In Fifth Conference on Email and Anti-Spam, CEAS, 2008
4. UCI Machine Learning Repository." <http://archive.ics.uci.edu/ml/>, 2012.
5. H. A. Chip man, E. I. George, and R. E. McCulloch. BART: Bayesian Additive Regression Trees. Journal of the Royal Statistical Society, 2006. Ser.B, Revised.
6. S. Nawafleh, W. Hadi (2012). Multi-class associative classification to predicting phishing websites. International Journal of Academic Research Part A; 2012;4(6), 302-306J. Clerk

- Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
7. P. Tiwari, R. Singh International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 4 Issue 12, December-2015.
 8. J. P. Marques de Sa. Pattern Recognition: Concepts, Methods and Applications. Springer, 2001.
 9. D. Michie, D. J. Spiegelhalter, and C. C. Taylor. Machine Learning, Neural and Statistical Classification. Ellis Horwood, 1994.
 10. L. Breiman. Random forests. Machine Learning, 45(1):5–32, October 2001
 11. Mrs. Sayantani Ghosh, Mr. Sudipta Roy, Prof. Samir K. Bandyopadhyay, “A tutorial review on Text Mining Algorithms”.
 12. Mauro Ribeiro, Katarina Grolinger, and Miriam AM Capretz. 2015. Mlaas: Machine learning as a service. In Proceedings of the 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). IEEE, Miami, FL, USA, 896–902.
 13. Ahmed Salem, Yang Zhang, Mathias Humbert, Pascal Berrang, Mario Fritz, and Michael Backes. 2019. ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models. In Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS). IEEE, San Diego, California, USA, 1–15.
 14. Ferdinando S Samaria and Andy C Harter. 1994. Parameterisation of a stochastic model for human face identification. In Proceedings of the Second IEEE Workshop on Applications of Computer Vision. IEEE, Sarasota, FL, USA, 138–142.
 15. AMAZON ML SERVICES. 2019. Amazon aws Machine Learning. <https://aws.amazon.com/machine-learning/>
 16. Snehkumar Shahani, Jibi Abraham, and R Venkateswaran. 2017. Distributed Data Aggregation with Privacy Preservation at Endpoint. In Proceedings of the IEEE International Conference on Management of Data. IEEE, Chennai, India, 1–9.
 17. Richard Shay, Saranga Komanduri, Adam L Durity, Phillip Seyoung Huh, Michelle L Mazurek, Sean M Segreti, Blase Ur, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2014. Can long passwords be secure and usable?. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, Toronto, ON, Canada, 2927–2936.
 18. Reza Shokri and Vitaly Shmatikov. 2015. Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS). ACM, Denver, Colorado, USA, 1310–1321.
 19. Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP). IEEE, San Jose, CA, USA, 3–18.
 20. Amit Kumar Sikder, Hidayet Aksu, and A Selcuk Uluagac. 2017. 6thsense: A context-aware sensor-based attack detector for smart devices. In Proceedings of the 26th USENIX Security Symposium (USENIX Security 17). USENIX Association, Vancouver, BC, Canada, 397–414.
 21. Tom W Smith, Peter Marsden, Michael Hout, and Jibum Kim. 2012. The General social surveys. Technical Report. National Opinion Research Center at the University of Chicago
 22. Raphael Spreitzer, Felix Kirchengast, Daniel Gruss, and Stefan Mangard. 2018. ProcHarvester: Fully automated analysis of procfs side-channel leaks on Android. In Proceedings of the 2018 on Asia Conference on Computer and Communications Security (AsiaCCS). ACM, Incheon, Republic of Korea, 749–763.

23. NedimSrndic and Pavel Laskov. 2014. Practical evasion of a learning-based classifier: A case study. In Proceedings of the 2014 IEEE Symposium on Security and Privacy (SP). IEEE, San Jose, CA, USA, 197-211
24. Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural networks* 32 (2012), 323–332.
25. Jingchao Sun, XiaocongJin, Yimin Chen, Jinxue Zhang, Yanchao Zhang, and Rui Zhang. 2016. VISIBLE: Video-assisted keystroke inference from tablet backside motion. In Proceedings of the 23rd Annual Network and Distributed System Security Symposium (NDSS). IEEE, San Diego, CA, USA.
26. P Ramprakash, M Sakthivadivel, N Krishnaraj, J Ramprasath. "Host-based Intrusion Detection System using Sequence of System Calls" *International Journal of Engineering and Management Research*, Vandana Publications, Volume 4, Issue 2, 241-247, 2014
27. N Krishnaraj, S Smys."A multihoming ACO-MDV routing for maximum power efficiency in an IoT environment" *Wireless Personal Communications* 109 (1), 243-256, 2019.
28. N Krishnaraj, R Bhuvanesh Kumar, D Rajeshwar, T Sanjay Kumar, Implementation of energy aware modified distance vector routing protocol for energy efficiency in wireless sensor networks, 2020 *International Conference on Inventive Computation Technologies (ICICT)*,201-204
29. Ibrahim, S. Jafar Ali, and M. Thangamani. "Enhanced singular value decomposition for prediction of drugs and diseases with hepatocellular carcinoma based on multi-source bat algorithm based random walk." *Measurement* 141 (2019): 176-183. <https://doi.org/10.1016/j.measurement.2019.02.056>
30. Ibrahim, Jafar Ali S., S. Rajasekar, Varsha, M. Karunakaran, K. Kasirajan, Kalyan NS Chakravarthy, V. Kumar, and K. J. Kaur. "Recent advances in performance and effect of Zr doping with ZnO thin film sensor in ammonia vapour sensing." *GLOBAL NEST JOURNAL* 23, no. 4 (2021): 526-531. <https://doi.org/10.30955/gnj.004020> , https://journal.gnest.org/publication/gnest_04020
31. N.S. Kalyan Chakravarthy, B. Karthikeyan, K. Alhaf Malik, D.Bujji Babbu,. K. Nithya S.Jafar Ali Ibrahim , Survey of Cooperative Routing Algorithms in Wireless Sensor Networks, *Journal of Annals of the Romanian Society for Cell Biology* ,5316-5320, 2021
32. Rajmohan, G, Chinnappan, CV, John William, AD, Chandrakrishnan Balakrishnan, S, Anand Muthu, B, Manogaran, G. Revamping land coverage analysis using aerial satellite image mapping. *Trans Emerging Tel Tech.* 2021; 32:e3927. <https://doi.org/10.1002/ett.3927>
33. Vignesh, C.C., Sivaparthipan, C.B., Daniel, J.A. et al. Adjacent Node based Energetic Association Factor Routing Protocol in Wireless Sensor Networks. *Wireless Pers Commun* 119, 3255–3270 (2021). <https://doi.org/10.1007/s11277-021-08397-0>.
34. C Chandru Vignesh, S Karthik, Predicting the position of adjacent nodes with QoS in mobile ad hoc networks, *Journal of Multimedia Tools and Applications*, Springer US, Vol 79, 8445-8457,2020