**Automating Cyber Security and Management of Cyber Threats in Network through ML Techniques: A Smart and Efficient Approach to Deal with Cyber Threats.**

Vaibhav Gupta, Vaibhav Sharma,, Tushar Gupta and Dr. Rohit Rastogi

[vaibhav.19b101024@abes.ac.in, vaibhav.19b101046@abes.ac.in, tushar.19b101125](mailto:vaibhav.19b101024@abes.ac.in,%20vaibhav.19b101046@abes.ac.in,%20tushar.19b101125), rohitrastogi.shantikunj@gmail.com

**ABSTRACT**

***Cyber Security has become a major concern in this digital era. Since, the cyber-attacks and their types are increasing at an immense rate, it is not humanly possible to monitor, identify and take actions against the attacks. With the current automation systems majorly relying on supervised learning algorithms where they have already seen the type of attacks to monitor and manage the attacks, these systems have been rendered inefficient by zeroday attacks. It is essential that the research team start realising the immense potential of AI and utilise it to its full potential in the field of Cyber Security. If correctly applied, Artificial Intelligence can help to detect and deal with the cyber-attacks more efficiently and can help protect users that are not very security conscious and are not aware about the dangers of these security breaches.***

***The authors have decided to utilise machine learning algorithms like Decision Trees and Random Forest to detect zero day attacks as well as handle other common Cyber-attacks, the authors’ team have designed an Artificial Intelligence based cybersecurity system***

***that detects the attacks based on the anomalies rising in the network usage, port usage and packet analysis.***

***Keyword--*** *Supervised Learning, Unsupervised Learning, KDD (Knowledge Discovery in Database), phishing, smishing, DDoS*

**MOTIVATION**

Today's world of web connection, online fraud and cybercrime has increased, so there is the need for software which can prevent them. It is the main motivation behind the project. As cyber criminals are coming up with new methods of crime which are proving harder to be detected, so to make the project more robust and prevent it from becoming outdated, the team of authors chose the field of anomaly detection which is not based on some field pattern detection but uses AI to detect the fraud. With increase in usage of mobile phones, most of the frauds are now shifting towards them, so research members chose to develop the application for mobile phones.

**SCOPE OF STUDY**

As the world gets smarter, the rate of Cyber Attacks also increases. So, there is a large need for Cyber Attack Protection Systems. The devices such as mobile phones, laptops etc. sometimes show unusual behaviour so, to detect the cause of that unusual behaviour one need to install a cyber protection system. These types of systems are also helpful in understanding various different anomalies that occur in our devices. This type of system also protects various devices from hacking. These systems are also used to enhance Cyber Security.

**TOPIC ORGANISATION**

**ROLES OF AUTHOR**

Rohit Rastogi acted as team leader and worked closely with all co-authors. He prepared the topic introduction and background study; also contributed to the development of project development with his vast availability of knowledge. He also prepared the structure of the manuscript and ensured the quality of the content along with all co-authors. Tushar worked keenly on dataset and analysis of the dataset. Vaibhav Sharma and Vaibhav Gupta helped in developing the essential machine model for the project.

**1 INTRODUCTION**

As one knows, with the increase in time the devices such as mobile phones, laptops, televisions, washing machines, etc. are becoming smarter. But as the devices become smarter the threat of Intrusions increases. The resources such as mic, storage, location services etc. make devices more prone to Intrusions like hacking because hackers use these features as a means to gain access to the device so, there is a great need of systems such as Intrusion Detection System.

**1.1 Cyber Attacks and Global Scenario**

Cyberattacks have the potential to disrupt an organisation or individual’s business or ruin their credibility and reputation thus causing huge direct and indirect losses to them. With the recent situation of Coronavirus and work from home model there has been a significant rise in risks to cybersecurity. In a recent [survey](https://www.techrepublic.com/article/remote-working-were-stressed-and-distracted-and-making-these-security-errors/) done by a company named Tessian, 43% of people made a mistake at work that had security threats, while 47% of people working in the IT sector have said they have clicked on a phishing email at work. There is also a lack of cybersecurity expertise and hence it is difficult to track down offenders and bring them to justice. The significance of cybersecurity will likely increase in the future with organisations becoming aware and proactive regarding the threats (S & P Global market Intelligence, 2020) [10]; (Rastogi, R., et al. 2020)[8]; (Rastogi, R. 2018)[5].

**1.2 Usage of AI and ML in Smart Phone Issue Detection and Network Detection**

Artificial Intelligence and Machine learning models are very effective in analysing large amount of data finding relevant insights from them, as mobile phone itself creates large amount of data on daily operations and on increasing usage of phone is also creating many processing data so applying Artificial intelligence on them will provide us with insights needed for anomaly detection an attack predictions (Poremba, S., 2021) [4]; (Rastogi, R., et al., 2015)[9]

Increasing malware is at a staggering rate of 588 per min and keeps on increasing at a faster rate so it becomes difficult to detect them due to this large quantity. In this scenario Artificial Intelligence plays a crucial role in analysing large data and giving out faster responses before significant damage happened (Poremba, S., 2021) [4].

**2 LITERATURE REVIEW**

Cybersecurity using AI and ML techniques is an emerging area. Recent focus has shifted from traditional cybersecurity measures to use of AI and ML to automate cybersecurity. A lot of research has gone into this matter and there have been lots of breakthroughs in the subject but still there is much to discover and realise the true potential of AI and ML in cybersecurity.

Sarker, I. et al., (2021) and his team has elaborated the utilisation of cybersecurity models to deal with cyber threats. They have mentioned and elaborated various types of cyber-attacks possible and various defence strategies that can be implemented to deal with the cyber threats. They have also mentioned various algorithms like KNN, Clustering, etc. and their usage in the field of cybersecurity. They have also elaborated on AI based modelling which involves algorithms like Deep Neural Networks to detect anomalies and deal with cyber threats. They have concluded on the note that AI based Cybersecurity is the need of the hour and is required to deal with future cyber threats. Although it has its shortcomings right now, it can be improved with further research (Sarker, I. H., et al., 2021)[11].

Subham, K. Gupta.et.al (2022) and his team has worked on Hybrid Optimization and deep learning based intrusion detection systems for IOT based applications, which could have adverse effects if attacks were to be performed on them. This is for the security of smart city projects which are vulnerable to network based attacks. The detention of attacks in this sector needs to be carried out in real time and cannot be delayed due to training time of the models. So the project by the team gave a model which could detect intrusions accurately with reduced training time. So the proposed hybrid optimization and DL-centric improved IDS utilising DLHNN s such system. This included following modules:- Pre-processing module, grouping module, feature selection module, separation module and attack detection module. The model has achieved 99.52% accuracy with 362 sec training time which made the model practically applicable in the IoT sector. In collusion the model proposed is 3% faster than existing model in predicting low frequency attacks (Gupta, S.K. et al., 2022)[3].

Alabadi, M., et al., (2020) and his team has written a conference paper on Anomaly Detection for Intrusion Detection Based on Convolution Neural Network. They describe various types of cyber-attacks in this paper. They started their topic introduction by describing the word anomalies. They also describe Machine Learning and Deep Learning. They applied CNN in their project and they compared the results with the results which they got by applying Support Vector Machine (SVM), Random Forest (RF), and Hidden Markov Model (HMM) Models and Bayesian Network (BN). The framework they use is described in four phases - Input Phase, Input Pre-processing Phase, Classification Phase, Output phase. After describing the framework they used they describe some future works related to their research. At last they completed their research paper by giving a conclusion to their paper (Alabadi, M. et al., 2020) [2].

**Methodology**

Authors have decided to use Decision Tree and Random Forest Machine learning algorithms and thus compare the results.

**Data-sets**

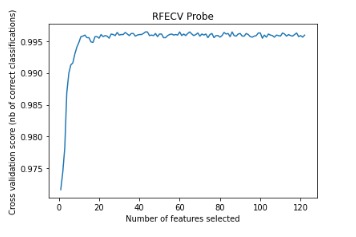
1. **NSL-KDD:**

Refined dataset than KDD cup 1999 by removing redundant rows

**Results and Discussions**

Authors have observed after the analysis of the test and train data packets present in NSL KDD dataset that there are four main types of attacks namely:

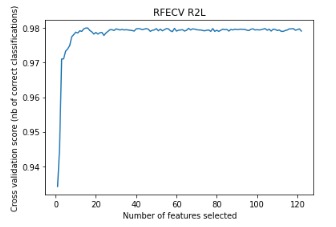
* DoS (Denial of Service)
* Probe
* R2L (Remote to Local)
* U2R (User to Root)



***Figure 23. RFECV Graph of Decision Tree Algorithm for Probe Attacks***

This is a RFECV graph of Decision Tree Algorithm for Probe attack type. On the x axis it has no. of features being selected by RFECV over a period of time and on the y axis it has Cross Validation Score which is basically the no. of correct classifications that is being done by Decision Tree Algorithm.

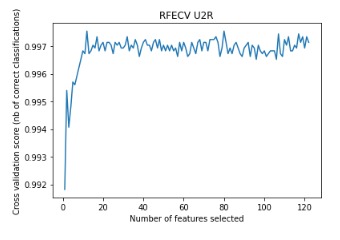
On analysis of the graph it can be seen that It relatively attains a constant at around 30 features and continues to have a near constant Cross Validation Score (Please refer fig 23).



***Figure 24. RFECV Graph of Decision Tree Algorithm for R2L Attacks***

This is a RFECV graph of Decision Tree Algorithm for R2L (remote to local) attack type. On the x axis it has no. of features being selected by RFECV over a period of time and on the y axis it has Cross Validation Score which is basically the no. of correct classifications that is being done by Decision Tree Algorithm.

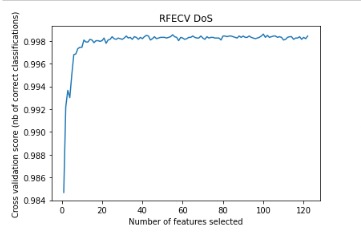
On analysis of the graph it can be seen that Cross Validation Score with the no. of features being selected. It obtains a maxima around 20 no. of features and then experiences a local minima at around 25. It attains a relative constant at around 27 features and continues to have a near constant Cross Validation Score (Please refer fig 24).



***Figure 25. RFECV Graph of Decision Tree Algorithm for U2L Attacks***

This is a RFECV graph of Decision Tree Algorithm for U2L (user to local) attack type. On the x axis it has no. of features being selected by RFECV over a period of time and on the y axis it has Cross Validation Score which is basically the no. of correct classification that is being done by Decision Tree Algorithm (Please refer fig 25).

On the analysis of this graph the contrast can be seen as compared to the other graphs obtained. In this graph it can be seen that there isn’t a clear levelling of rather it has multiple local maximas and minimas, but in the end the score does increase as the number of features selected increase.



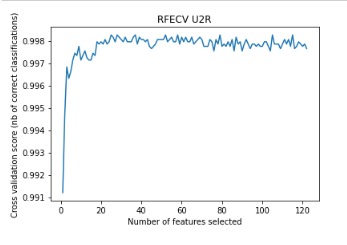
***Figure 26. RFECV Graph of Random Forest Algorithm for DoS Attacks***

This is a RFECV graph of Random Forest Algorithm for DoS attack. On the x-axis it has no of features being selected by Recursive Feature Elimination by Cross Validation. A steep rise in the accuracy (score) of the graph can be observed the above graph. Y axis represents the cross validation score. The graph has various local maximas and minimas but it almost levels out at about .998 score (Please refer fig 26).

RFECV score of Random Forest Algorithm for the probe attack type. Cross Validation score is represented on y axis while the no. of features being selected is represented on x axis.

The graph has no. of steep climbs as the no. of features being selected increase and then levels off at about .996 value. It is very remarkable how it becomes a near constant after a a while (Please refer fig 27).

RFECV graph for R2L (Remote to Local) type of attack using Random Forest Algorithm. On the x axis there is no. of features being selected and then on y axis there is Cross Validation score. The graph here is very smooth in nature as compared to other graphs and has a near linear rise in its value before levelling off at about the score of .98 (Please refer fig 28).



***Figure 29. RFECV Graph of Random Forest Algorithm for U2R Attacks***

This is a RFECV graph for U2R (User to Root) attack type plotted for Random Forest Algorithm. On the x-axis there is no. of features being selected while on the y axis there is Cross Validation score.

On the analysis of this graph the contrast can be seen as compared to the other graphs obtained. In this graph it can be seen that there isn’t a clear levelling off rather it has multiple local maximas and minimas, but in the end the score ends up not quite close to the maximum value it attained but still at an acceptable level (Please refer fig 29).

**5 NOVELTIES**

* Analysing major attack types from dataset
* Model-fitting various algorithms to check their accuracy
* Reducing False positive rates
* Analysis on predictions made by model

**7 FUTURE RESEARCH DIRECTIONS AND LIMITATIONS**

**Limitations**

**NSL KDD** even though is benchmark dataset is an older dataset and thus the models need to be trained on recent data.

The models also need to be trained on real time data for efficiency and robustness.

**7.2 Future Directions**

More research should be done at a less noisy place so that the result is error free. Intensive research needs to be put into some more machine learning algorithms are used so that the result is more accurate. For further research results to be more accurate, there is a need for a system to detect false alarms so that the efficiency of the system increases.

**8 CONCLUSIONS**

The Decision Tree Algorithm as well as The Random Forest Algorithm perform well on NSL KDD dataset. NSL KDD dataset is considered as benchmark dataset for intrusion detection systems and has 4 major attacks types namely DoS (Denial of Service) , Probe, U2R (User to Root), R2L (Remote to Local).

The authors have tried to select the best algorithms for the intrusion detection system. The system is designed to be able to detect four different types of attacks. The four types of attacks are namely DoS (Denial of Service), Probe, U2R (User to Root), R2L (Remote to Local).

Authors have tried to use 2 different types of algorithms to detect and classify attacks namely:

Decision Tree Algorithm, Random Forest Algorithm

Although the Random Forest Algorithm is on par with Decision Trees in terms of accuracy , it lacks the necessary efficiency in terms of time and thus can’t be deployed on the real time data set.

**9 ACKNOWLEDGEMENTS**

The Author Team pays their respects to God. They then thank each of their parents for everything they have done for them. They then express their sincere appreciation to ABES Engineering College, Ghaziabad for providing help and assistance for lab equipment and research. They further offer their sincere gratitude to seniors of ABES Engineering College, Ghaziabad to help and guide them whenever they require them. We also offer our special thanks to ABES management, Director ABESEC, HOD ECE, HOD CSE and all other senior Coordinators for their extraordinary support to our team. In the end, we pay our due respects and thanks to each and everyone involved, direct or indirect in this work.

**REFERENCES**

[1]Agarwal, A., Goel, D., Tyagi, A., Aggarwal, A., Rastogi, R. (2018). A Smarter Approach for Better Lifestyle in Indian Societies. In: Saeed, K., Chaki, N., Pati, B., Bakshi, S., Mohapatra, D. (eds) Progress in Advanced Computing and Intelligent Engineering. Advances in Intelligent Systems and Computing, vol563. Springer, Singapore. <https://doi.org/10.1007/978-981-10-6872-0_33>

[2]Alabadi, M. and Celik, Y., (2020). Anomaly Detection for Cyber-Security Based on Convolution Neural Network : A survey, International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1-14, doi:10.1109/HORA49412.2020.9152899.

[3]Gupta, S.K., Tripathi, M., Grover, J., (2022). Hybrid optimization and deep learning based intrusion detection system. *ELSEVIER, Computers and Electrical Engineering 100*, 107876

https://online.champlain.edu/blog/top-cybersecurity-skills-in-high-demand

[4][Poremba](https://securityboulevard.com/author/sue-poremba/), S., (June, 2021). Adding AI/ML to Mobile Security, Security Boulevard, accessed on 2 April., 2022.

[**https://securityboulevard.com/2021/06/adding-ai-ml-to-mobile-security/**](https://securityboulevard.com/2021/06/adding-ai-ml-to-mobile-security/)

[5]Rastogi, R. (2018). Statistical Analysis for Effect of Positive Thinking on Stress Management and Creative Problem Solving for Adolescents

[6]Rastogi, R. *et al.* (2020). Intelligent Analysis for Personality Detection on Various Indicators by Clinical Reliable Psychological TTH and Stress Surveys. In: Das, A., Nayak, J., Naik, B., Pati, S., Pelusi, D. (eds) Computational Intelligence in Pattern Recognition. Advances in Intelligent Systems and Computing, vol 999. Springer, Singapore.<https://doi.org/10.1007/978-981-13-9042-5_12>

[7]Rastogi, R. *et al.* (2020). Intelligent Personality Analysis on Indicators in IoT-MMBD-Enabled Environment. In: Tanwar, S., Tyagi, S., Kumar, N. (eds) Multimedia Big Data Computing for IoT Applications. Intelligent Systems Reference Library, vol 163. Springer, Singapore.<https://doi.org/10.1007/978-981-13-8759-3_7>

[8]Rastogi, R., Saxena, M., Chaturvedi, D.K., Gupta, M., Rastogi, M., Rustagi, D., Gaur, V., Kohli, V., Srivastava, P., Jain, M., & Kumar, P., (2020). AI-Based Analysis for Novel Covid-19 and Its Treatment Through Yajna and Mantra Science, eds. Moumtzoglou, A., & Kyriakou, A., IJRQEH, Vol.9, issue 4, article 6, PP: 75-98. URL:-<https://www.igi-global.com/journal/international-journal-reliable-quality-healthcare/44660>

[9]Rastogi, R.,Mishra R.,Sharma S.,Nigam A.and Arya P., "Security of data transmission using logic gates and crypt analysis," (2015) 2nd International Conference on Computing for Sustainable Global Development (INDIACom)6986 v, 2015, pp. 101-105.

[10]S&P Global Market Intelligence, (Oct. 2020). 451 Research’s Voice of the Enterprise: Digital Pulse, Coronavirus Flash Survey.

**h**[**ttps://go.451research.com/MC-2020-10-CoronavirusFlashSurveyOct2020\_LandingPage.html?utm\_source=covidms**](https://go.451research.com/MC-2020-10-CoronavirusFlashSurveyOct2020_LandingPage.html?utm_source=covidms) **accessed on 2 April, 2022.**

[11]Sarker, I.H., Furhad, M.H. & Nowrozy, R. (2021). AI-Driven Cybersecurity: An Overview, Security Intelligence Modelling and Research Directions. *SN COMPUT. SCI. 2,* 173. https://doi.org/10.1007/s42979-021-00557-0

[12]Singh, R. Gupta and R. Rastogi, "A novel approach for vehicle tracking system for traffic jam problem," (2015) 2nd International Conference on 16Computing for Sustainable Global Development (INDIACom), 2015, pp. 9-174.

[13]Smart device cyber security: An introduction, (Nov. 03, 2021). Get Cyber Safe, Accessed at 2 April, 2022,

**url:**[**https://www.getcybersafe.gc.ca/en/blogs/smart-device-cyber-security-introduction**](https://www.getcybersafe.gc.ca/en/blogs/smart-device-cyber-security-introduction)

[14]Top 7 Mobile Security Threats in 2020, (2020), Kasperskey, Accessed on 2 April, 2022, https://www.kaspersky.co.in/resource-center/threats/top-seven-mobile-security-threats-smart-phones-tablets-and-mobile-internet-devices-what-the-future-has-in-store

A**uthors’ Profile**

**Dr. Rohit Rastogi**, Associate Professor, Department of Computer Science and Engineering, ABES Engineering College.

**Mr. Vaibhav Gupta**, Department of Computer Science and Engineering, ABES Engineering College.

**Mr. Vaibhav Sharma**, Department of Computer Science and Engineering, ABES Engineering College.

**Mr. Tushar Gupta**, Department of Computer Science and Engineering, ABES Engineering College