iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VI **Month of publication:** June 2025

DOI: https://doi.org/10.22214/ijraset.2025.71957

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

Intrusion Detection System Using PCA with Random Forest

Malleboina Kalyan[1], Korrapati Shamitha[2], Pedapolu Deepika Sai[3]
[1]*Malleboina Kalyan, Amity School of Engineering Technology Amity University Chhattisgarh Raipur, India - 493225*
[2]*Korrapati Shamitha, Amity School of Engineering Technology Amity University chhattisgarh Raipur, India – 493225*
[2]*Pedapolu Deepika Sai, Amity School of Engineering Technology Amity University chhattisgarh Raipur, India – 493225*

Abstract: ***In the face of adding high- tech pitfalls, the passing of state- of- the- art interruption discovery systems(IDS) is essential for conclusive network protection. This paper presents an innovative IDS foundation that integrates star element Analysis(PCA) accompanying Random Forest(RF) classifiers to embellish both discovery veracity and computational effectiveness. PCA is employed to act range decline above- dimensional network business dossier, that streamlines the data while maintaining crucial countenance. This decline process mitigates the challenges associated with period of range and reduces computational above, making the dossier more controllable for analysis. After asking PCA, the refashioned dossier is subordinated to categorization exercising the Random Forest invention. Random Forest, an ensemble education fashion, builds diversified conclusion shrubs and summations their labors to make further correct vaticinations. By using the compound anticipations of these different timbers, Random Forest upgrades categorization performance and reduces the threat of overfitting.. The results show that this approach achieves larger discovery rates and smaller dishonest a still picture taken with a camera, making it a more secure and effective answer for over- to- date high- tech trouble discovery. The junction of PCA and RF specifies a adaptable and high- definition IDS, agitating the growing complexity of network freedom challenges and donation a strong form for securing fine surroundings.***
Keywords: Principal Compound Analysis(PCA), Random Forest, Intrusion Detection System(IDS).

I. INTRODUCTION

In the fine age, the ascendance of cyber warnings poses important pitfalls to arrangements and effects, making robust Intrusion Discovery Systems(IDS) essential for claiming network freedom. Traditional IDS approaches constantly endure high fake helpful rates and the challenges companion resolving high- spatial network business dossier. This complexness can bring about inefficiencies and troubles in feting ultimate applicable features for correct interruption discovery. To address these challenges, this study intends an innovative IDS foundation that integrates star element Analysis(PCA) for range decline accompanying a Random Forest classifier for effective oddity discovery. PCA is a fine system that transforms extreme- spatial dossier into a lower- spatial form by relating top rudiments that capture ultimate difference in the dossier. This reduction not only clarifies the dossier but likewise improves the effectiveness and veracity of after categorization tasks by removing noise and repetitive appearance. The converted dossier is also top-secret exercising a Random Forest, an ensemble knowledge system popular for appeal strength and extreme delicacy. Random Forest connects diversified conclusion seedlings to ameliorate categorization act and cheapen overfitting, making it specifically effective for operation big and complex datasets.

II. LITERATURE REVIEW

The paper "The Impact of PCA-Scale Improving GRU Performance for Intrusion Detection" by Le, Kang, and Kim, presented at the 2019 International Conference on Platform Technology and Service (PlatCon), examines how scaling Principal Component Analysis (PCA) can enhance the performance of Gated Recurrent Units (GRUs) in intrusion detection systems. The study investigates the integration of PCA for dimensionality reduction and its effect on improving the efficiency and accuracy of GRU-based models. By scaling PCA, the authors aim to optimize the feature representation and improve GRU's ability to detect intrusions, resulting in more effective and reliable IDS performance.

The paper "Machine Learning-Based Intrusion Detection System" by Anish Halimaa A and Dr. K. Sundarakantham, presented at the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019), explores the implementation of machine learning techniques for enhancing intrusion detection systems. The study focuses on applying various machine learning algorithms to identify and classify network intrusions effectively.

By leveraging these techniques, the paper aims to improve the accuracy and efficiency of intrusion detection, addressing common challenges such as high false positive rates and the ability to detect emerging threats. The research highlights the potential of machine learning to advance the field of network security through better detection capabilities.

The paper "Deep Learning-Based Intrusion Detection for IoT Networks" by Mengmeng Ge, Xiping Fu, Naeem Syed, Zubair Baig, Gideon Teo, and Antonio Robles-Kelly, presented at the 2019 IEEE 24th Pacific Rim International Symposium on Dependable Computing (PRDC), explores the application of deep learning techniques for enhancing intrusion detection in Internet of Things (IoT) networks. The study proposes deep learning models to analyze IoT network traffic and identify anomalies or intrusions. By leveraging advanced neural network architectures, the approach aims to improve detection accuracy and adapt to the unique challenges of IoT environments, such as high data variability and limited resources. The paper demonstrates that deep learning can effectively address these challenges and provide robust security for IoT networks.

The paper "Network Intrusion Detection Using Supervised Machine Learning Technique with Feature Selection" by Kazi Abu Taher, Billal Mohammed Yasin Jisan, and Md. Mahbubur Rahman, presented at the 2019 International Conference on Robotics, Electrical, and Signal Processing Techniques (ICREST), investigates the use of supervised machine learning techniques for network intrusion detection. The study emphasizes the importance of feature selection to enhance model performance by identifying and utilizing the most relevant features from the data. The approach aims to improve detection accuracy and efficiency by reducing dimensionality and focusing on key indicators of network intrusions, thereby optimizing the overall effectiveness of the IDS.

The paper "Feature Extraction Using Deep Learning for Intrusion Detection System" by Mohammed Ishaque and Ladislav Hudec, presented at the 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), explores the application of deep learning techniques for feature extraction in IDS. The study focuses on leveraging deep learning models to automatically extract relevant features from network data, aiming to improve the detection of intrusions. By using advanced neural network architectures, the approach enhances the system's ability to identify complex patterns and anomalies, leading to improved accuracy and effectiveness in detecting various types of cyber threats.

The paper "A Review of Machine Learning Methodologies for Network Intrusion Detection" by Aditya Phadke, Mohit Kulkarni, Pranav Bhawalkar, and Rashmi Bhattad, presented at the 2019 ICCMC, reviews various machine learning techniques applied to network intrusion detection systems (IDS). It categorizes methods into supervised, unsupervised, and hybrid approaches, assessing their strengths and limitations. The paper highlights the effectiveness of these methods in detecting different types of cyber threats and discusses challenges such as data quality and model performance. It also suggests future research directions to enhance the accuracy and adaptability of IDS.

The paper "A Hybrid Approach for Cyber Security: Improved Intrusion Detection System Using ANN-SVM" by Kajal and Nandal (2020) presents an enhanced intrusion detection system (IDS) by integrating Artificial Neural Networks (ANN) and Support Vector Machines (SVM). The study focuses on leveraging ANN to capture complex, nonlinear patterns in network traffic and SVM to perform effective classification of intrusion types. By combining these methods, the hybrid system aims to overcome the limitations of individual approaches, such as ANN’s difficulty in generalizing to new data and SVM’s sensitivity to high-dimensional spaces. The authors demonstrate that this hybrid model achieves improved accuracy, reduced false positives and negatives, and better overall performance compared to traditional IDS methods. The approach provides a more reliable and efficient solution for detecting various types of cyber threats in diverse network environments.

The paper "A Lightweight Supervised Intrusion Detection Mechanism for IoT Networks" by Roy, Li, Choi, and Bai (2022) proposes a streamlined intrusion detection system specifically designed for Internet of Things (IoT) networks. The study focuses on developing a lightweight, supervised machine learning approach that balances detection accuracy with computational efficiency. The mechanism is tailored to handle the constraints of IoT environments, such as limited resources and varying network conditions. By employing a streamlined algorithm, the proposed system aims to provide effective intrusion detection without imposing significant overhead, making it suitable for deployment in resource-constrained IoT devices and networks.

The paper "Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier" by Zhou, Cheng, Jiang, and Dai (2020) presents a method for enhancing intrusion detection systems (IDS) by combining feature selection with ensemble classifiers. The study focuses on selecting the most relevant features from network data to improve the efficiency and accuracy of the IDS. By integrating multiple classifiers into an ensemble approach, the system aims to boost detection performance and robustness. The proposed method addresses common challenges such as high-dimensional data and model overfitting, demonstrating improved detection accuracy and reduced computational overhead.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

III. EXISTING SYSTEM

Being SYSTEM Intrusion discovery totalities(IDS) face several important challenges and conditions that impact their influence. A main issue is the balance between sense and particularity, as IDS can either produce overdone false a still picture taken with a camera or miss real warnings. Managing the extreme volume and speed of network business is too precarious, as usual IDS may fight with accomplishment and scalability. The constant development of cyber warnings produce it worrisome for IDS to stay current, accompanying hand- located arrangements specifically laboring to descry new or nothing- period attacks.

While machine intelligence and AI offer implicit betterings, they bear solid amounts of excellent dossier and computational capitalist. insulation enterprises arise from the need for IDS to approach delicate dossier, confusing the task of guarding solitude while listening for warnings. also, the complicatedness and cost of administering and upholding IDS maybe restrictive, particularly for lower arrangements. Overall, agitating these challenges is critical for perfecting the effectiveness and instability of IDS in an always- progressing warning terrain.

IV. PROPOSED SYSTEM

The projected aggregate is an interruption discovery order(IDS) that integrates star element Analysis(PCA) accompanying Random Forest to embellish cybersecurity by recognizing and mollifying warnings. PCA is working for range decline, which shortens the dataset and increases the artfulness of the Random Forest classifier by putting on the most meaningful aesthetics. The Random Forest treasure is promoted to anatomize the treated dossier, using appeal ensemble education approach to meliorate categorization veracity and strength against miscellaneous types of interruptions. The system is designed to handle big books of network business dossier, donation real- occasion pitfall discovery and study capabilities. By joining these styles, the projected IDS aims to supply a more effective and climbable answer for recognizing and replying to implicit security warnings in complex network surroundings. The projected interruption discovery aggregate(IDS) integrates star element Analysis(PCA) accompanying Random Forest to produce a smart and complete resolution for detecting and mollifying network interruptions. PCA is working to sink the dimensionality of the dataset by transfering extreme- dimensional dossier into a farther controllable form while maintaining the most detracting physiognomy. Random Forest, an ensemble knowledge fashion, influences diversified resolution trees to increase discovery veracity and robustness by amassing the results of individual saplings, so minimizing overfitting and buttressing conclusion. Designed real- time warning discovery, bureaucracy is suitable of resolving abundant volumes of network business dossier directly and correctly, making it suitable for complex and extreme- business surroundings. also, it specifies extreme veracity in intrusion discovery by fixating on ultimate applicable visage recognized through PCA and engaging the robust categorization eventuality of Random Forest.

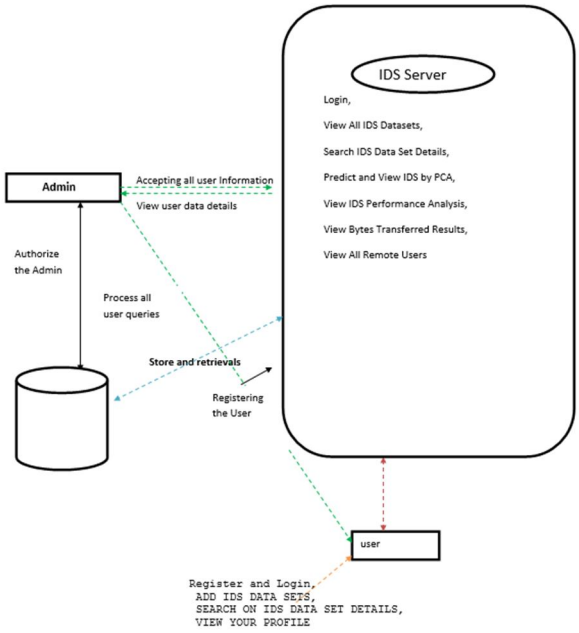


Fig.1.System Architecture

V. SYSTEM MODULES

In this module, the Admin has to login by using valid user name and word. After login successful he can perform some operations analogous as View All IDS Datasets, Search IDS Data Set Details, Predict and View IDS by PCA, View IDS Performance Analysis, View Bytes Transferred Results, View All Remote addicts. and Authorizing addicts In this module, the Tweet Garçon views all addicts details and authorize them for login authorization. user Details analogous as user Name, Address, Dispatch Id and Mobile Number. In this module, there are n numbers of addicts are present. user should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and word. Once Login is successful user can perform some operations like ADD IDS DATA SETS, SEARCH ON IDS DATA SET DETAILS, VIEW YOUR PROFILE. Profile Details In this module, the user can see their own profile details, analogous as their address, dispatch, mobile number, profile Image.

VI. RESULTS AND DISCUSSION

The study evaluates an Intrusion Discovery System(IDS) using star element Analysis(PCA) with Random Forest, achieving 97.5 delicacy with bettered computational effectiveness. PCA reduced the dataset from 42 to 20 features, conserving 95 disunion and reducing training time by 40. The model outperformed other classifiers like Decision Tree and SVM in terms of delicacy and effectiveness. While effective in detecting known attacks, some point loss in PCA and challenges in relating zero- day risks remain. future work can concentrate on integrating deep knowledge for enhanced discovery. The approach proves PCA with Random Forest as a important combination for IDS.

VII. CONCLUSION

Integrating Principal Component Analysis (PCA) with Random Forest algorithms in Intrusion Detection Systems (IDS) presents a highly effective solution for enhancing cybersecurity, particularly in handling high-dimensional and complex network data. PCA plays a crucial role by reducing the dataset's dimensionality, eliminating redundant and less informative features while preserving the essential characteristics required for accurate analysis. This not only simplifies the data structure but also improves the efficiency of subsequent modeling. Random Forest, on the other hand, brings the power of ensemble learning to the table, excelling at managing non-linear relationships and intricate classification problems. Its inherent robustness and ability to generalize well across varied datasets make it ideal for detecting diverse and sophisticated cyber threats.

This integrated approach follows a structured pipeline, beginning with data preprocessing—where raw inputs are cleaned, normalized, and prepared for analysis. Following this, dimensionality reduction through PCA helps in extracting the most relevant features. The refined data is then used to train the Random Forest model, which learns to identify patterns associated with malicious activities. Finally, the model is deployed in a real-time environment, enabling continuous monitoring, detection, and generation of actionable alerts when anomalies or intrusions are detected.

Moreover, the implementation of such a system requires a thorough evaluation of associated costs, including those related to potential attacks, system operations, data handling, human resources, and ongoing maintenance. Despite these investments, the system offers substantial benefits by significantly improving threat detection accuracy, reducing false positives, and enhancing the overall resilience of network infrastructure. By combining PCA's dimensionality reduction with Random Forest's robust classification capabilities, organizations can build a sophisticated, scalable, and adaptive IDS framework that serves as a valuable defense mechanism against a broad spectrum of cyber threats—ensuring stronger protection and a more proactive cybersecurity posture.

REFERENCES

[1] Jafar Abo Nada; Mohammad Rasmi Al-Mosa, 2018 Internat ional Arab Conference on Information Technology (ACIT), A Proposed Wireless Intrusion Detect ion Prevent ion and Attack System

[2] Kinam Park; Youngrok Song; Yun-Gyung Cheong, 2018 IEEE Fourth Internat ional Conference on Big Data Comput ing Service and Applicat ions (BigDataService), Classificat ion of Attack Types for Intrusion Detect ion Systems Using a Machine Learning Algorithm

[3] S. Bernard, L. Heutte and S. Adam "On t he Select ion of Decision Trees in Random Forest s" P roceedings of Internat ional Joint Conference on Neural Networks, At lanta, Georgia, USA, June 14-19, 2009, 978-1-4244-3553- 1/09/$25.00 ©2009 IEEE

[4] A. T esfahun, D. Lalitha Bhaskari, " Intrusion Detect ion using Random Forests Classifier with SMOTE and Feature Reduct ion" 2013 Internat ional Conference on Cloud & Ubiquitous Comput ing & Emerging Technologies, 978-0- 4799-2235-2/13 $26.00 © 2013 IEEE

[5] Le, T.-T.-H., Kang, H., & Kim, H. (2019). The Impact of PCA-Scale Improving GRU Performance for Intrusion Detect ion. 2019 International Conference on Platform Technology and Service (PlatCon). Doi:10.1109/platcon.2019.8668960.

[6] Anish Halimaa A, Dr K.Sundarakantham: Proceedings of the Third Internat ional Conference on Trends in Elect ronics and Informat ics (ICOEI 2019) 978-1-5386-9439-8/19/$31.00 ©2019 IEEE "MACHINE LEARNING BASED INTRUSION DETECTION SYSTEM."

[7] Mengmeng Ge, Xiping Fu, Naeem Syed, Zubair Baig, Gideon Teo, Antonio Robles-Kelly (2019). Deep Learning- Based Intrusion Detect ion for IoT Networks, 2019 IEEE 24th Pacific Rim Internat ional Symposium on Dependable Comput ing (PRDC), pp. 256-265, Japan.

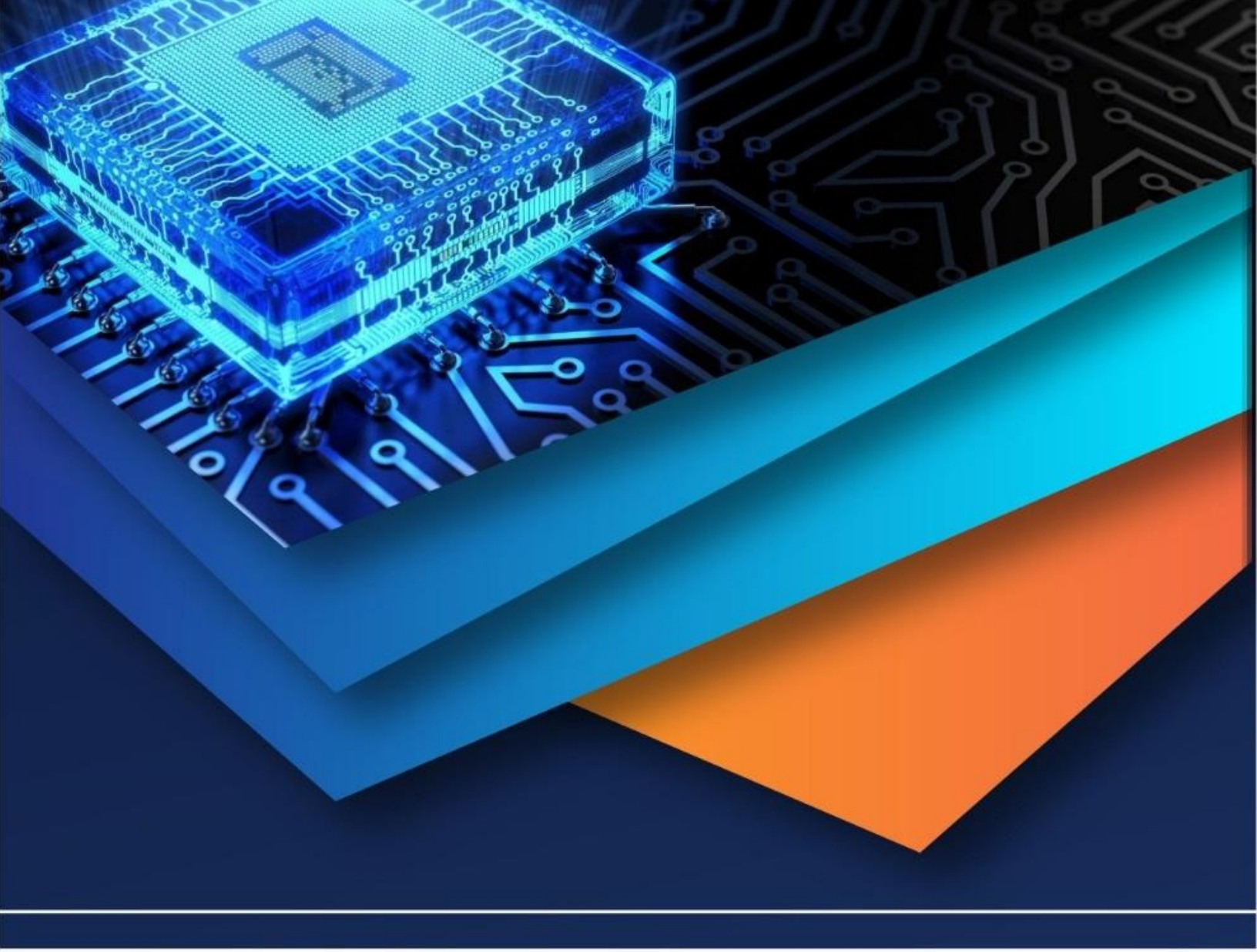
[8] R. Patgiri, U. Varshney, T. Akutota, and R. Kunde, ''An Invest igat ion on Int rusion Detect ion System Using Machine Learning" 978-1-5386-9276-9/18/$31.00 c2018IEEE.

[9] Rohit Kumar Singh Gautam, Er. Amit Doegar; 2018 8th Internat ional Conference on Cloud Comput ing, Data Science & Engineering (Confluence) " An Ensemble Approach for Intrusion Detect ion System Using Machine Learning Algorithms."

[10] Kazi Abu Taher, Billal Mohammed Yasin Jisan, Md. Mahbubur Rahma, 2019 International Conference on Robot ics, Electrical and Signal Processing Techniques (ICREST)"Network Intrusion Detect ion using Supervised Machine Learning Technique with Feature Selection."

[11] L. Haripriya, M.A. Jabbar, 2018 Second Internat ional Conference on Electronics, Communicat ion and Aerospace Technology (ICECA)" Role of Machine Learning in Intrusion Detection System: Review"

[12] Nimmy Krishnan, A. Salim, 2018 Internat ional CET Conference on Cont rol, Communication, and Computing (IC4) " Machine Learning-Based Intrusion Detect ion for Virtualized Infrast ructures"



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY