Cybersecurity Intrusion Detection System Using Machine Learning

Mr.Abhale B. A[1], Mr. Pathare.G. N[2], Miss.Jadhav Samiksha V[3]., Miss.Kaklij Sakshi B[4].,
Miss.Khare Vishakha G[5]., Miss.Gangurde Shreya S[6]
S.N.D College of engineering and research center yeola, Savitribai Phule Pune University

***Abstract*—This project develops an advanced Intrusion Detection System (IDS) by combining real time attack simulations run on Kali Linux with the CIC-IDS2017 dataset. Machine learning models for precise intrusion detection are developed and evaluated using the CIC-IDS2017 dataset, which includes labeled network traffic data covering a range of attack types and typical activities. To supplement this, Kali Linux is used to create more realistic attacks like scanning ports, brute force, and denial-of-service (DoS) in a controlled seFng. To improve the machine learning models, network traffic is recorded and analyzed during these attacks. Combining real- world, live attack data with thorough dataset training improves the system's capacity to identify a variety of malicious activity with low false positives and high accuracy. By successfully detecting changing cyberthreats in dynamic environments, this hybrid approach provides a scalable and adaptable intrusion detection system (IDS) that improves network security.**

***Index Terms*—Intrusion Detection System, machine learning, CIC-IDS2017 dataset, Kali Linux, real-time attack simulation, network security, port scanning, denial of service, brute force attack, traffic analysis, cybersecurity, anomaly detection.**

I. INTRODUCTION

In today's digital age, the rapid expansion of connected systems over the internet has created a fearful surge in cyber threats against networks, organizations, and individuals alike. As companies increasingly grow to depend on interconnected infrastructure, maintaining data confidentiality, integrity, and availability has become an essential issue. Antivirus software and traditional security solutions like firewalls
are mainly meant to secure against recognized threats through predetermined signatures; however, such solutions cannot identify new, advanced, or developing attacks like zero-day exploits, advanced persistent threats (APTs), and polymorphic malware. To deal with these shortcomings, Intrusion Detection Systems (IDS) have become prime mechanisms for detecting unauthorized or irregular activities on a network. IDS solutions scan incoming and outgoing traffic, inspect data packets, and mark them as normal or malicious. In spite of their significance, most conventional IDS deployments are plagued by high false positives and poor support for fresh attack patterns, which decrease their effectiveness in realworld settings. The incorporation of Machine Learning (ML) into IDS systems provides a viable solution to these problems. Through learning from both typical and attack traffic, ML-based IDS models are capable of identifying underlying patterns and detecting anomalies that could portend cyber intrusions. In this study, an efficient Cybersecurity Intrusion Detection System based on Machine Learning is constructed, taking advantage of both the CIC-IDS2017 dataset and real-time attack simulations run in a controlled Kali Linux setting. The blending of synthetic data with actual attack traffic increases the system's capacity to identify a wide range of cyber threats, such as port scanning, denial-of-service (DoS), and brute force attacks. This IDS hybrid seeks to attain high detection rates, low false positives, and real-time adaptability to adaptive attack behaviors. This IDS hybrid seeks to attain high detection rates, low false positives, and real-time adaptability to adaptive attack behaviors. The suggested system hence presents an enhanced, scalable, and more robust cybersecurity infrastructure capable of safeguarding contemporary networks from changing cyberattacks. Furthermore, the proposed IDS integrates automation and continuous learning capabilities, allowing it to adapt dynamically to new threats. By combining realtime data analysis with advanced ML algorithms, the system ensures

faster threat identification, minimizes manual intervention, and enhances the overall resilience of network infrastructures emerging cyberattacks.

II. LITERATURE SURVEY

Sr n o	Paper title	Author name	year
1	Intrusion System Using machine Detection learning Algorithms	Rachid Tahri1*, Youssef Balouki1, Abdessamad Jarrar2, and Abdellatif Lasbahani3 (2022)	2022
2	A Review of Intrusion Detection Systems Using Machine Learning: Attacks, Algorithms, and Challenges	Jose Luis Gutierrez-Garcia1, Eddy Sanchez DelaCruz2, and Maria del Pilar Pozos Parra3	2023
3	Network Intrusion Detection and Comparative Analysis using Ensemble Machine Learning and Feature Selection	Saikat Das, Sajal Saha, Annita Tahsin Priyoti, Etee Kawna Roy, Frederick T. Sheldon, Senior Member, IEEE, Anwar Haque, and Sajjan Shiva, Fellow	2022
4	Ensuring network security with a robust intrusion detection system using ensemble-based machine learning	Md. Alamgir Hossain * , Md. Saiful Islam.	2023
5	Enhancing Network Security through Machine Learning Based Intrusion Detection Learning Systems	Salar Mohammad1, Vrince Vimal2, Dr. Aradhana Sahu3*, Anna Shalini4, Dr. S. Farhad5, Elangovan Muniyandy6, Dr. Ajmeera Kiran7	2024
6	Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction	Md. Alamin Talukder1,7*, Md. Manowarul Islam1, Md Ashraf Uddin2, Khondokar Fida Hasan3, Selina Sharmin1, Salem A. Alyami4 and Mohammad Ali Moni5.	2025
7	Intrusion Detection Systems Based on Machine Learning Algorithms	Sandy Victor Amanoul, Adnan Mohsin Abdulazeez, Diyar Qader Zeebare, Falah ,Y. H. Ahmed	2025
8	Evaluation of Machine Learning Algorithms for Intrusion Detection System	Mohammad1, Vrince Vimal2, Dr. Aradhana Sahu, Anna Shalini4,Dr. S. Farhad5, Elan govan Muniyandy, Dr. Ajmeera Kiran	2024
9	Real-Time Intrusion Detection via Machine Learning Approaches	Erik Murtaj1, Fausto Marcantoni1, Michele Loreti1, Michela Quadrini1,* and Hans Friedrich Witschel	2023

III. METHODOLOGY

The methodology of the proposed Cybersecurity Intrusion Detection System (IDS) is based on a hybrid framework that combines machine learning algorithms with real-time attack simulations to improve accuracy, adaptability, and efficiency in detecting malicious network activities. The approach begins with data collection, where the CIC-IDS2017 dataset is used as the primary source for model training. This dataset, developed by the Canadian Institute for Cybersecurity, contains labeled network representing normal activity and multiple types of attacks such as denial-ofservice (DoS), brute force, infiltration, and web based threats. However, since static datasets alone are not sufficient to capture dynamic real-world behavior, live attack

traffic is also generated in a controlled environment using Kali Linux tools such as Nmap for port scanning, Hydra for
brute force attacks, and Slowloris for denial-of-service simulations. This combination ensures that the data used for training is both diverse and representative of real- world scenarios. After data collection, the next stage involves data preprocessing, which is crucial for ensuring data data quality and model performance. In this stage, the captured traffic is 20 21 an incomplete or cleaned to remove redundant entries, normalized to achieve uniform feature scaling, and encoded for consistency. Relevant features such as packet length, duration, protocol type, source and destination IP addresses, and connection counts are extracted to serve as meaningful inputs to the machine learning models. This refined data is then divided into training and testing sets for model evaluation. The third phase focuses on model training and selection, where multiple machine learning algorithms such as Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Deep Neural Networks (DNN) are implemented. Each model is trained on the processed dataset to learn the distinguishing characteristics of normal and malicious traffic. Among these, the Deep Neural Network shows superior performance due to its ability to capture nonlinear patterns and complex feature relationships. The trained models are then evaluated using statistical metrics such as accuracy, precision, recall, F1 score, and confusion matrix analysis to measure their effectiveness and reliability in detecting different attack categories. In the final stage, the best-performing model is integrated into a real-time detection system. traffic Using network monitoring tools like Wireshark or Tcpdump, the IDS captures live network packets, analyzes them through the trained model, and generates immediate alerts for any detected anomalies. The system is designed to continuously update its knowledge base by learning from new attack data, thus improving its detection capability over time. This hybrid methodology dataset-driven effectively training combines with live simulation, resulting in an adaptive and scalable IDS that offers high detection accuracy and reduced false positives in dynamic cybersecurity environments.

IV. OBJECTIVE

The primary objective of this research is to design and develop an intelligent, hybrid Intrusion Detection System (IDS) that effectively utilizes Machine Learning (ML) algorithms in combination with realtime network attack simulations to accurately detect and classify malicious network activities. This IDS aims to overcome the inherent limitations of conventional rule-based and signature based detection systems by providing a dynamic, adaptive, and datadriven solution capable of identifying both known and previously unseen attacks. The integration of datasetbased training and live simulation data forms the foundation of this hybrid approach, ensuring that the model is not only theoretically sound but also practical and reliable under real world network conditions. A key objective of the study is to enhance the accuracy, efficiency, and reliability of intrusion detection mechanisms. Existing IDS frameworks frequently suffer from high false- positive rates, which lead to unnecessary alerts and decreased system trustworthiness. By implementing advanced machine learning techniques such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Deep Neural Networks (DNN), this research seeks to identify the most effective model in achieving optimal accuracy, minimal false alarms, and consistent performance across varied network scenarios. The system also aims to utilize data preprocessing techniques such as feature extraction, normalization, and encoding to ensure that the input data is clean, balanced, and representative of real network conditions. Another crucial objective is to create a scalable and adaptive IDS capable of handling large volumes of network traffic in real time. The proposed system will leverage Kali Linux to generate realistic attack simulations such as Denial of Service (DoS), port scanning, and brute force attacks, which will serve as supplemental data to the CIC- IDS2017 dataset. This hybrid dataset ensures greater diversity in training data and allows the model to generalize effectively when exposed to unseen patterns. Furthermore, the system will continuously learn and update its detection rules as new threats emerge, thus ensuring its resilience against evolving cyberattack techniques. In addition to detection capability, this research also focuses on improving the interpretability and usability of IDS systems. Many machine learningbased IDS solutions perform well in detecting anomalies but lack transparency and user friendly

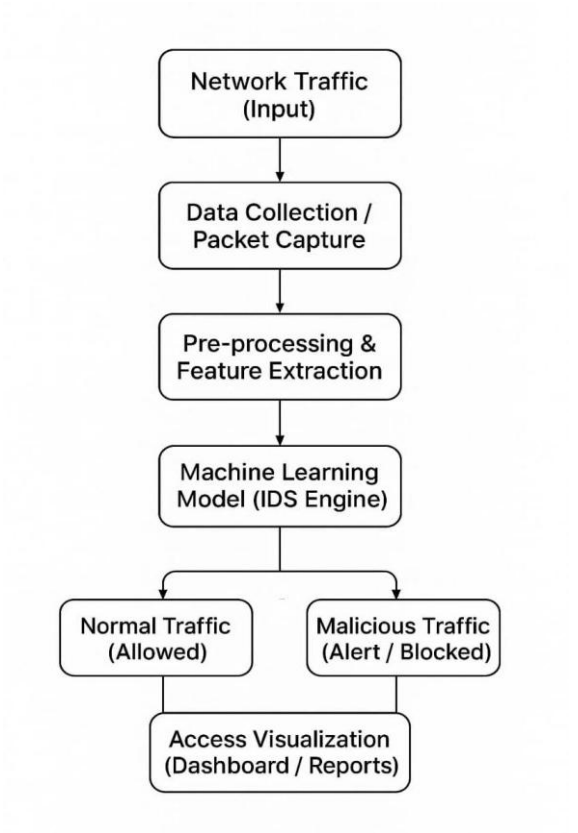
visualization tools. Hence, this study also aims to develop an interactive monitoring dashboard that displays real time network activity, logs, and intrusion alerts in a clear and interpretable format. This interface will enable cybersecurity professionals and network administrators to visualize potential threats, analyze attack vectors, and take proactive security measures efficiently. Finally, this research aspires to contribute toward the advancement of automated cybersecurity intelligence by demonstrating how the integration of machine learning with real time data selflearning, analysis can create a adaptive, and scalable defence mechanism. By combining the precision of ML algorithms with the authenticity of real network traffic, the proposed system significantly is expected to reduce false positives, increase detection accuracy, and enhance the overall robustness of modern network security infrastructures.

V. PROBLEM DEFINATIONS

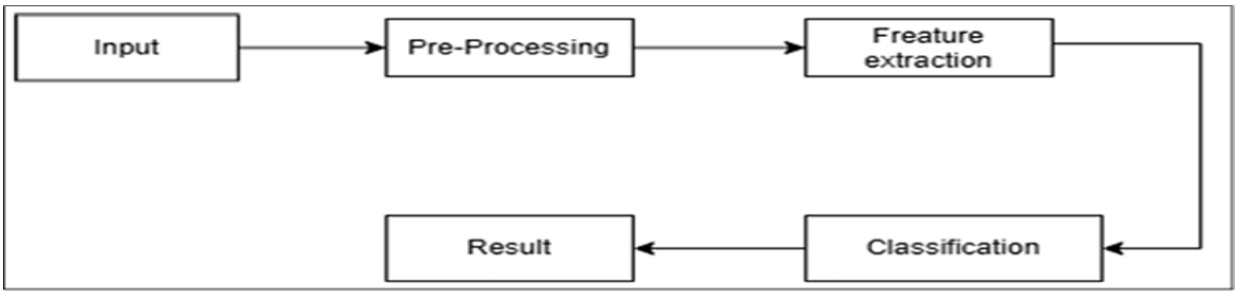
In the current era of rapid digital transformation, the sophistication increased of frequency cyberattacks exponentially. and have Traditional network security systems such as firewalls, antivirus programs, and signature-based Intrusion Detection Systems (IDS) are limited in their ability to detect new or evolving threats. These systems rely on predefined signatures and static rules, making them ineffective against zero-day exploits, polymorphic malware, and advanced persistent threats (APTs). Furthermore, they often suffer from high false- positive rates, which reduce the efficiency of network administrators and delay the response to genuine security breaches. With the growing complexity of modern network infrastructures — including cloud computing environments, IoT devices, and distributed architectures

— the challenge of distinguishing between legitimate and malicious network activities has become more difficult. Existing IDS models trained on static datasets lack adaptability when deployed in real-world scenarios, as they cannot cope with the dynamic nature of live traffic and the continuous evolution of attack patterns. Consequently, these systems fail to maintain accuracy and reliability over time. Another significant issue is the imbalance between false positives and false negatives in existing IDS frameworks. Excessive false alarms can lead to alert fatigue, while undetected intrusions may cause severe damage to organizational assets. Moreover, most IDS models are limited to laboratory conditions and fail to generalize in live network environments where real-time adaptability is essential. This limitation emphasizes the urgent need for a hybrid IDS that integrates machine learning techniques with real-time data analysis and live attack simulation. The problem this research addresses is the lack of an adaptive, intelligent, and data-driven Intrusion Detection System capable of learning from both labeled datasets and real- time attack data. The proposed system leverages the CIC-IDS2017 dataset and Kali Linuxbased live attack simulations to create a hybrid learning environment that mimics realistic network conditions. This approach enables the system to detect known and unknown threats efficiently while minimizing false alarms. By combining the analytical power of machine learning with the practical realism of live simulations, the proposed IDS aims to deliver high detection accuracy, scalability, and robustness — thereby enhancing the overall security posture of modern digital infrastructure.

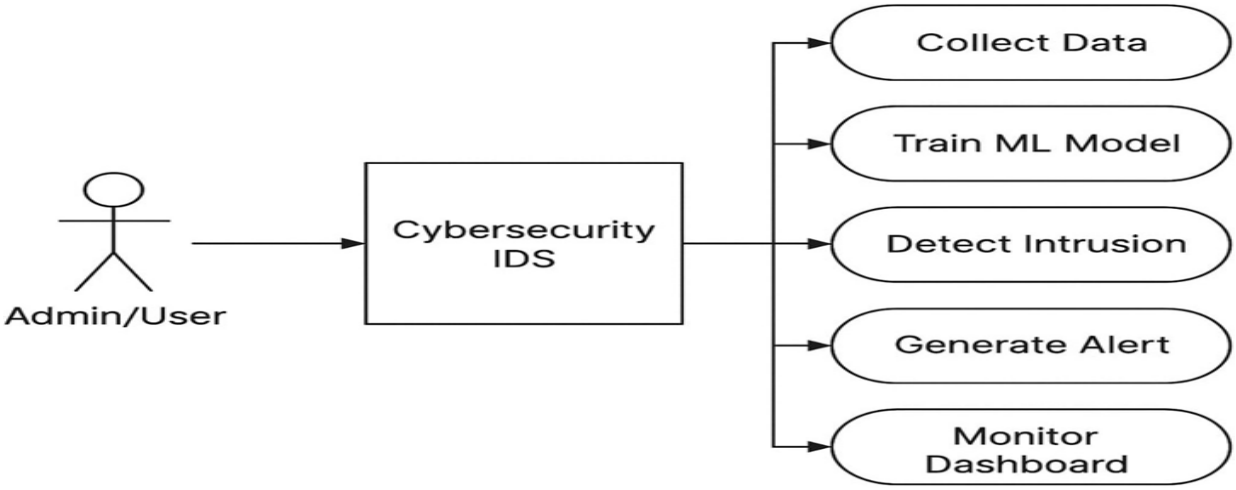
FLOW CHART



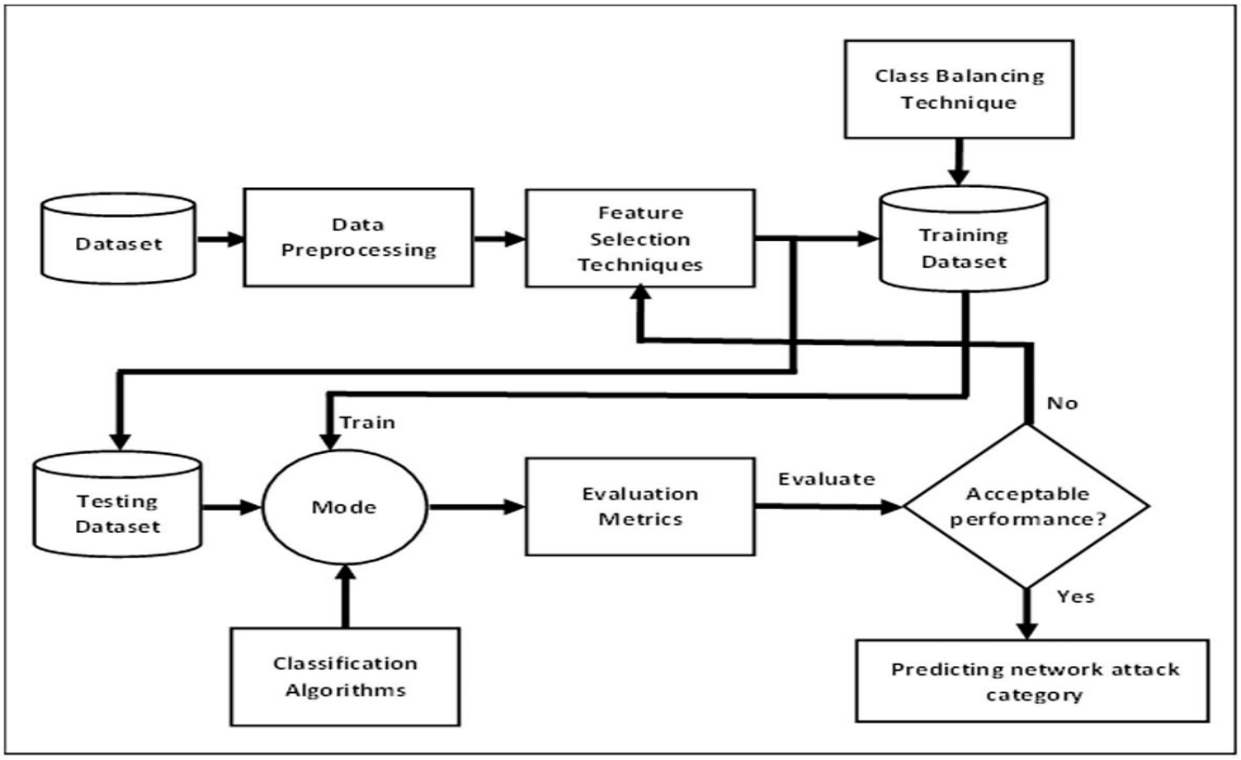
DataFlow Diagram:



Class Diagram:



Architecture:



VI. FUCTIONAL REQUIREMENTS

The system must be capable of collecting network traffic data from multiple sources, including pre-existing datasets and live simulations. The CIC-IDS2017 dataset serves as the foundational dataset containing various types of labeled network traffic. In addition, the IDS must collect real-time attack data generated using Kali Linux tools such as Nmap, Hydra, and Slowloris. This ensures that the training and testing data represent both controlled and realistic network conditions, improving the overall performance of the detection model.

• Before analysis, the raw data collected from network traffic must undergo preprocessing to ensure quality and uniformity. This involves data cleaning, normalization, and transformation into a structured form suitable for analysis. The IDS should extract relevant features such as protocol type, source and destination IP, packet size, connection duration, and flow rate. These features are essential for distinguishing between normal and abnormal traffic behavior during the training and classification process.

• The system must implement machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Deep Neural Networks (DNN) to train models capable classifying of detecting and network intrusions. The models must be trained on the processed dataset to learn the characteristics of both benign and malicious network traffic. The IDS should also provide the ability to test the trained models, evaluate performance metrics (accuracy, precision, recall, F1 score), and optimize hyperparameters for better detection outcomes.

• The system must continuously monitor live network traffic and classify it as normal or malicious in real time. Once the model is deployed, it should be capable of detecting different types of attacks such as Denial of Service (DoS), Brute Force, and Port Scanning. The system must analyze each incoming packet or connection using the trained model and identify deviations from normal behaviour, thus ensuring rapid and accurate intrusion detection.

• The IDS must have an automatic alert system that notifies administrators when an intrusion or suspicious activity is detected. These alerts should include details about the type of attack, the affected IP addresses, and the time of detection. The notification system can use email, SMS, or dashboard alerts to inform network administrators, allowing them to take immediate corrective actions to prevent further damage

. • Every detection event must be logged in the system for auditing and analysis purposes. The IDS should maintain comprehensive logs of network traffic, detected intrusions, timestamps, and actions taken. The reporting module should generate detailed reports summarizing network behavior, attack statistics, and model performance. This functionality helps in performance evaluation, compliance, and continuous improvement of the system

. • The IDS must implement a secure login system to ensure that only authorized users can access the system dashboard and configuration settings. Role-based access control should be used to differentiate between administrators, analysts, and regular users. This helps in maintaining data confidentiality and prevents unauthorized modifications or misuse of system data.

• A graphical user interface (GUI) or web based dashboard must be provided for users to visualize network activity, monitor ongoing processes, and view analytical results in real time. The dashboard should include charts, graphs, and tables representing the distribution of traffic, detected attacks, and system performance metrics. This helps users make informed decisions quickly based on real-time data insights.

VII. NON FUCTIONAL REQUIREMENTS

The IDS must achieve a high detection rate with a low rate of false positives and false negatives. Since false alarms can lead to unnecessary actions and missed detections can compromise network security, the system must ensure reliable classification through robust training, cross- validation, and model optimization. Reliability also includes consistent performance across different datasets and environments.

• The system must be optimized to handle large volumes of data in real time without significant delays. Efficient processing ensures that attacks are detected promptly before they cause damage. The IDS should maintain minimal latency in analyzing traffic and generating alerts, even under high-load conditions.

• The proposed IDS should be scalable to handle increasing data volumes and diverse network environments. It must support deployment across

different architectures including enterpriselevel networks, cloud systems, and distributed infrastructures — without affecting performance. The system design should allow horizontal and vertical scaling as the number of connected devices and network traffic increases.

- The system must ensure the confidentiality, integrity, and security of captured data, including network packets, logs, and user credentials. Encryption protocols should be used for data transmission and storage, while access controls should prevent unauthorized access. The IDS should comply with relevant cybersecurity standards and privacy regulations, ensuring ethical handling of sensitive information.
- The IDS must be adaptable to different network configurations and capable of learning from new attack patterns. The system's modular design should make it easy to update, retrain, or modify components without disrupting overall functionality. Maintenance should involve periodic performance evaluation, model retraining, and updating of attack signatures or training data. • The system must provide an intuitive and userfriendly interface that allows users to navigate easily.

VIII. CONCLUSION

The proposed Cybersecurity Intrusion Detection System using Machine Learning presents a hybrid, intelligent, and adaptive approach to safeguarding networks from modern cyber threats. By combining machine learning algorithms with real-time attack simulations, the system addresses the limitations of traditional rule-based IDS solutions that struggle to detect new or evolving attacks. The integration of the CIC-IDS2017 dataset with Kali Linux–based live attack data provides the IDS with both theoretical depth and practical exposure, significantly enhancing its detection capability and reducing false alarm rates. The methodology followed in this research encompassing data collection, preprocessing, model training, and evaluation — demonstrates the potential of machine learning techniques such as Random Forest, SVM, KNN, and DNN in accurately identifying and classifying various network intrusions. Among these, deep learning based models have shown superior performance, enabling the system to detect complex attack patterns that conventional IDS might overlook. The inclusion of real-time traffic analysis and adaptive retraining mechanisms ensures that the system remains effective against emerging cyber threats, continuously learning from new data to improve its performance over time. The results of this study indicate that a hybrid IDS leveraging both supervised learning and live simulation can significantly enhance the reliability and responsiveness of network defense systems. Beyond technical achievements, this research also contributes to the growing field of cyber defense automation, where intelligent algorithms can make real-time security decisions without human intervention. In conclusion, the developed IDS not only strengthens the detection and prevention of cyberattacks but also establishes a scalable and efficient framework for future advancements in network security. Future work may include integrating deep reinforcement learning, scalability, and cloudbased automated response systems, thereby evolving this IDS into a comprehensive, self-learning cybersecurity solution capable of defending against everchanging digital threats.

REFERENCES

[1] A Review of Intrusion Detection Systems Using Machine Learning: Attacks, Algorithms, and Challenges Jose Luis Gutierrez-Garcia1, Eddy Sanchez DelaCruz2, and Maria del Pilar Pozos Parra3(2023).

[2] Intrusion System Using Machine Detection Learning Algorithms Rachid Tahri1*, Youssef Balouki1, Abdessamad Jarrar2, and Abdellatif Lasbahani3 (2022).

[3] Network Intrusion Detection and Comparative Analysis using Ensemble Machine Learning and Feature Selection Saikat Das, Sajal Saha, Annita Tahsin Priyoti, Etee Kawna Roy, Frederick T. Sheldon, Senior Member, IEEE, Anwar Haque, and Sajjan Shiva, Fellow, IEEE (2022).

[4] Ensuring network security with a robust intrusion detection system using ensemble-based machine learning Md. Alamgir Hossain *, Md. Saiful Islam (2023).

[5] Enhancing Network Security through Machine Intrusion Detection Learning Systems Based Salar Mohammad1, Vrince Vimal2, Dr. Aradhana Sahu3*, Anna Shalini4, Dr. S. Farhad5,

Elangovan Muniyandy6, Dr. Ajmeera Kiran7 .(2024).

[6] Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction Md. Alamin Talukder1,7*, Md. Manowarul Islam1, Md Ashraf Uddin2, Khondokar Fida Hasan3, Selina Sharmin1, Salem A. Alyami4 and Mohammad Ali Moni5.

[7] Balanced MultiClass Network Intrusion Detection Using Machine Learning FARAZ AHMAD KHAN SAIFULLAH SAIF, WASIM KHAN, ASGHARAL I SHAH NIZAL ALSHAMMR Y, MUHAM MADOSAM AMALIK, AND ZAHID ULLAH 5(2024)