Exploiting Machine Learning Techniques for Proactive Detection and Prevention of Network Intrusions

Puspraj Kumar Saket
Department of Computer Science, Madhyanchal Professional University, Bhopal (M.P.)

Md. Vaseem Naiyer
Department of Computer Science, Madhyanchal Professional University, Bhopal (M.P.)

Ankit Temurnikar
Department of Computer Science, Madhyanchal Professional University, Bhopal (M.P.)

ABSTRACT
Traditional intrusion detection systems (IDS) are sometimes insufficient when it comes to spotting sophisticated and ever-evolving attack patterns. This is due to the exponential growth in cyber threats as well as the rising complexity of modern networks. An investigation into the application of sophisticated machine learning (ML) methods for the purpose of enabling proactive detection and prevention of network intrusions is presented in this study. The purpose of this research is to improve the accuracy of anomaly detection, decrease the number of false positives, and respond more quickly to threats in real time. This will be accomplished by utilizing supervised, unsupervised, and deep learning models. A detailed study is carried out by utilizing benchmark datasets such as NSL-KDD and CICIDS. This analysis evaluates the performance of several methods, such as Random Forest, Support Vector Machines (SVM), K-Means Clustering, and Long Short-Term Memory (LSTM) networks. Through the identification of zero-day assaults and adaptive threat behaviors, the findings reveal that machine learning-driven intrusion detection systems (IDS) may vastly outperform traditional signature-based systems. In addition to this, the article explores the incorporation of these models into a real-time security framework in order to facilitate automated responses and improve the overall cybersecurity posture. The findings highlight the significant role that machine learning plays in the construction of network intrusion prevention systems that are intelligent, adaptable, and scalable for the future generation of digital networks.

Keywords
Machine Learning, Techniques, Network Intrusions

1. INTRODUCTION
Since the exponential growth of networked systems and the increasing sophistication of cyber threats, cybersecurity has become an enormously essential worry for individuals, organizations, and governments alike in this age of digital technology. This is because cyber-attacks are becoming increasingly sophisticated. There are times when traditional security solutions, such as signature-based intrusion detection systems (IDS), struggle to keep up with the ever-evolving techniques that cyber attackers deploy. This is especially true when it comes to zero-day vulnerabilities and polymorphic malware. Because of these limitations, there is an immediate and compelling need for increased intelligence, flexibility, and proactive approaches to the security of network infrastructure. This is a necessity that must be met immediately. In recent years, machine learning (ML), which is a branch of artificial intelligence (AI), has emerged as a very successful instrument in the realm of cybersecurity. ML is an acronym that stands for automated learning. When it comes to detecting and preventing breaches in real time, the application of machine learning techniques has the potential to dramatically increase their capabilities. This is achieved by studying massive amounts of data pertaining to network traffic and gaining knowledge from patterns of activity that are both normal and harmful while the process is being carried out. Machine learning models are able to adapt to new and unexpected attack vectors by identifying abnormalities and suspicious patterns. This is in contrast to traditional systems, which are dependent on preset rules and signatures. Machine learning models are able to adapt to new attack vectors. The objective of this project is to examine the application of a range of machine learning techniques, including supervised, unsupervised, and deep learning approaches, with the intention of constructing proactive intrusion detection and prevention systems. The fundamental objective is to establish which machine learning algorithms are the most successful in detecting various types of network assaults, lowering the number of false alarms, and enabling a quick reaction to prospective threats. The recommended models are trained and tested using benchmarks that are comprised of datasets such as NSL-KDD, CICIDS2017, and UNSW-NB15. These benchmarks are used for the purpose of training and testing the models. Additionally, the project investigates the implementation of intrusion detection systems (IDS) that are powered by machine learning into real-time security architectures. The difficulties that are linked with scalability, model interpretability, and deployment in dynamic network settings are also included in the scope of this inquiry. This research makes a contribution to the development of cybersecurity systems that are superior in terms of their resilience, intelligence, and autonomy. These systems are able to safeguard digital assets against a wide range of threats in a proactive manner. The expansion of knowledge regarding machine learning in network security and the implementation of this knowledge in network security are the means by which this objective is achieved. The fig 1 illustrates the overall workflow of the intrusion detection model, starting from data preprocessing (handling missing values, scaling, and encoding) to imbalanced data processing and feature reduction. The optimized data is then trained and tested using the SVEDM classifier to accurately identify different types of network attacks.

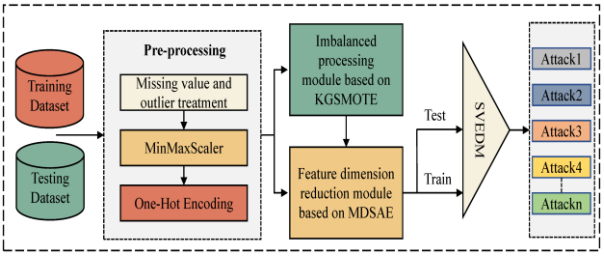

Fig 1: Framework for Intrusion Detection Using Machine Learning

1.1 Importance of Cybersecurity and Threat Detection

Many cyber threats, such as malware, phishing, denial-of-service (DoS) attacks, and unauthorized access, have significantly increased in danger as a direct result of the rapid development of digital infrastructures. These are only some of the cyber hazards that have significantly increased in risk. Cybersecurity is an essential component that must be present in order to ensure the protection of sensitive data, the preservation of privacy, and the continuation of company operations. Threat detection methods, like as intrusion detection systems (IDS), play an important part in the process of identifying potential security breaches at an early stage and taking preventative measures against them. Integrated threat management systems, often known as IDS, are employed by businesses in order to recognise, assess, and respond to threats in real time. This allows for the prevention of significant damage and the reduction of vulnerabilities.

1.2 Role of Machine Learning in IDS

The ability of intrusion detection systems (IDS) to identify new threats has been considerably enhanced by machine learning (ML), which has also resulted in a reduction in the IDS's dependence on previously defined signatures. It is very uncommon for standard intrusion detection systems to experience difficulties when it comes to recognising zero-day attacks and responding to new dangerous situations. approaches such as supervised learning, unsupervised learning, and reinforcement learning are utilised by machine learning-driven intrusion detection systems (IDS) in order to analyse massive amounts of data, identify patterns, and classify actions that are not usual. These activities are accomplished through the utilisation of these approaches. There are a number of methods that have been used in order to improve the accuracy and efficiency of intrusion detection systems (IDS), hence making them more resistant to complicated cyber-attacks. Some of these methods include deep learning, support vector machines (SVM), and ensemble learning. In conclusion, the use of machine learning into intrusion detection systems (IDS) represents a significant advancement in the field of cyber security. In order to safeguard digital ecosystems, it provides threat detection technologies that are not only intelligent but also adaptive and efficient.

1.3 Objectives

1. To analyse the shortcomings of conventional intrusion detection systems and emphasise the necessity of methods based on machine learning.
2. To find and evaluate a variety of ML techniques that work well for detecting intrusions in networks.
3. To use benchmark datasets like NSL-KDD, CICIDS2017, and UNSW-NB15 to assess the efficiency of chosen algorithms.

1.4 SIGNIFICANCE OF THE STUDY

The findings of this research are relevant when evaluated from the viewpoint of the shifting nature of the threat landscape and the limitations of the security mechanisms that are presently in place. The objective of this project is to make a contribution to the development of intelligent security systems that are not only able to recognize existing threats but also have the ability to foresee and react to attacks that might not have been anticipated. It is via the use of machine learning that this will be realized. It is possible that the findings of this study may give cyber security professionals, system architects, and researchers with useful insights that will assist them in the creation of intrusion detection systems that are more resilient and prepared for the future. Additionally, the results may play a vital part in the process of influencing policy decisions, developing security protocols, and strengthening the overall cyber defence strategies of businesses all over the world. This is because the findings include the process of establishing security protocols.

1.5 Fundamentals of Intrusion Detection Systems (IDS)

Within the realm of cybersecurity, Intrusion Detection Systems, which are sometimes referred to as IDS on occasion, are indispensable equipment. Identification of instances of malicious activity or unauthorized access to a network or system is the major objective of these surveillance systems. Intrusion detection systems (IDS) are designed to monitor and analyse the activity of systems, network traffic, and data in order to identify potential threats or security breaches. This serves the purpose of identifying potential threats or breaches. The installation of these systems will result in a number of major consequences, including the protection of sensitive data, the prevention of cyberattacks, and the maintenance of the integrity of network operations.

1.6 Traditional IDS Methods and Their Limitations:

For the most part, classic intrusion detection systems (IDS) rely on pre-defined signatures or criteria in order to identify possible threats. Among these ways is the comparison of the activity of the system or the traffic on the network to a database that contains information about attack patterns that is already known. Traditional intrusion detection systems appear to be effective against known threats; nevertheless, they have difficulty detecting attacks that are not recognised:

- Unknown Threats: Networks are susceptible to zero-day vulnerabilities because signature-based methods are unable to identify new or undiscovered assaults.
- False Positives: False positives and an overburden on security personnel can result from rule-based systems that incorrectly identify benign actions as harmful.
- Scalability Issues: Performance issues and missed threats could occur if traditional IDS is unable to handle a large amount of network traffic.

1.7 Need for Intelligent IDS Solutions:

The traditional intrusion detection systems (IDS) are no longer sufficient to protect against the increasingly sophisticated assaults that are being launched. The need for advanced intrusion detection systems (IDS) that make use of cutting-edge technology such as machine learning and artificial intelligence is a result of this. The ability to learn from data, recognize trends, and adapt to new threats is not dependent on rules for intelligent systems such as this one. Because they make use of machine learning algorithms to recognize irregularities, uncover hidden risks, and significantly cut down on the amount of false positives, intelligent intrusion detection systems (IDS) offer a more effective and adaptive solution to the current cybersecurity challenges than traditional methods.

1.8 Machine Learning Techniques for Intrusion Detection Systems (IDS)

When it comes to identifying potential cyber threats, the effectiveness of intrusion detection systems (IDS) has been

significantly enhanced by machine learning (ML). A wide range of machine learning techniques are utilised in order to identify activity in networks that is deemed to be odd or suspicious. Approaches such as supervised learning, unsupervised learning, and deep learning are included in these methodologies. In addition to providing improved accuracy and flexibility, hybrid approaches, which combine a number of different learning processes, are also available.

1.8.1 Supervised Learning Techniques

The use of labelled datasets, in which both benign and harmful behaviors are preset, is essential to the process of supervised learning. New cases are classified using these methods, which are based on patterns that have been learnt from previous data.

1.8.2 Decision Trees

- A model that is similar to a tree and it divides the data into branches according to the feature requirements.
- Easy to understand and makes optimal use of computing resources.
- Limitation: When dealing with complicated datasets, prone to overfitting than other situations.

1.8.3 Random Forests

- An approach to ensemble learning that mixes a number of different decision trees in order to achieve greater precision.
- The process of averaging the predictions of many trees helps to reduce the danger of over fitting.
- Strength: More robust than a single decision tree.

1.8.4 Support Vector Machines (SVM)

- A classification method that locates the hyperplane that is most effective in distinguishing between benign and potentially harmful activities.
- It performs admirably with both high-dimensional data and the smaller datasets.
- Limitation: Computationally expensive for large datasets.

1.8.5 Artificial Neural Networks (ANN)

- The artificial neural network (ANN) is modelled after the human brain and is made up of layers of neurones that are linked to one another and process information.
- Effective for complex pattern recognition tasks.
- Limitation: Demands a substantial amount of computer resources and is dependent on huge datasets.

1.8.6 Machine Learning (ML)

Machine learning (ML) is a process that enables computers to be educated to have the ability to automatically learn and improve or optimise performance criteria without being explicitly programmed. This is accomplished by utilising previous experience or example data. Machine learning is a technology that can be classified as a subfield of artificial intelligence (AI). For the purpose of making predictions about a variety of classes, the basic aim of a machine learning model is to train a collection of data in accordance with certain qualities of interest. Machine learning encompasses three distinct categories of algorithms: supervised, unsupervised, and reinforcement learning. In general, these three categories are referred to as the three types of algorithms.

1.8.7 Single Classifiers

In the field of machine learning, the term "single machine learning classifier" is used to describe any classifier that is made up of a single classification algorithm. Numerous various intrusion detection systems all make use of a same machine learning classification model. This is done in order to perform their respective functions. A number of different machine learning classifiers, including support vector machines (SVM), artificial neural networks (ANN), decision trees, kernel neural networks (KNN), and Naïve Bayes, have been employed in the various intrusion detection systems that have been investigated in this research.

1.8.8 Hybrid Classifiers

The hybrid classifier is a combination of two or more machine learning techniques, and its primary objective is to enhance the overall performance of the aggregated or resulting classifier in an intrusion detection system. The utilization of a hybrid approach inside the intrusion detection system (IDS) is carried out with the purpose of enhancing the effectiveness of the system. There is no question that hybrid systems are far more effective than a single machine learning categorizing intrusion detection system (IDS). This has been shown beyond a reasonable doubt. For the purpose of representing the initial level of hybrid classifiers, either unsupervised machine learning methods or supervised machine learning algorithms may be utilized.

1.8.9 Ensemble Classifier

The Ensemble Classifier is a collection of several different machine learning classifiers, which are sometimes referred to as weak learners in some contexts. These classifiers integrate their individual findings in some fashion in order to provide a prediction performance that is more effective and efficient in relation to the choice that is reached by consensus. Consequently, ensemble classifiers provide improved performance by integrating the results of a number of weak learners. This results in an improved overall performance. Several research investigations that made use of ensemble techniques exhibit a high level of complexity in terms of accuracy and prediction performance. These studies were conducted by a variety of researchers. The process of jointly creating ensembles may be accomplished by the use of a variety of techniques, including but not limited to the following: bagging and random forest, majority voting, randomness injection, feature selection ensemble, and error-correcting output coding.

2. REVIEW AND COMPARISON OF RELATED WORKS

Alkasassbeh and Almseidin utilized three different categorization strategies in order to address the issue of low accuracy in intrusion detection systems (IDS) that make use of artificial neural networks with fuzzy clustering in order to deal with attacks that occur seldom. The heterogeneous training data was separated into homogeneous subsets so that they could minimise the complexity of each training set and improve the accuracy of the training. J48 trees was the method that produced the best accuracy among those that were used in the study that was proposed. Multilayer Perceptron (MLP) and Bayes network came in second and third, respectively. One of the most significant challenges they have in their work is that they are unable to employ feature selection to get rid of any characteristics that are unnecessary, redundant, or undesired. An intrusion detection system that is based on single machine learning classifiers was developed by utilising the KDD-NSL

dataset in conjunction with techniques from decision trees and random forests. A return accuracy of 95.323% is achieved by the random classifier, which is superior to the other available choice. The difficulties of poor detection and false positive rate were not resolved by the application of the strategy that was provided. To determine whether or not a network has been compromised, Ponthapalli and colleagues used a single machine learning model in the research that they proposed. This study makes use of a variety of methodologies, including decision trees, logistic regression, random forests, and support vector machines.

KDD-NSL was the dataset that was employed for the inquiry. In light of the findings, the random forest classifier is the approach that should be taken when developing an intrusion detection system. Additionally, they discovered that the random forest classifier is the one that operates the quickest as well. A single dataset is the only one that the work that is being suggested is capable of processing very successfully. In an ensemble-based approach to intrusion detection systems, Marzia Z. used a voting classifier to aggregate the outcomes of a large number of supervised and unsupervised machine learning algorithms. The current intrusion detection systems have been improved in terms of both their accuracy and their speed as a direct result of our efforts. In spite of the fact that KDDCup '99 is the most employable dataset, it is somewhat old; hence, they decided to go with the Kyoto2006+ dataset, which is more promising. They are able to perform their tasks with a certain degree of precision as a consequence of this; nevertheless, in a limited number of situations, the memory of the outcome is very poor, which suggests that there are large false negative rates (FPR). A real-time hybrid intrusion detection system was presented by Dutt I. and colleagues. This method makes use of the abuse methodology to identify frequent attacks and the anomaly approach to find novel approaches to intrusion detection.

The high detection rate that this work achieved was due to the fact that the anomaly detection approach was able to recognise patterns of intrusions that were able to avoid being identified as assaults by the abuse detection system. The accuracy of the model increased gradually each day, reaching a substantial value of 92.65% on the last day of the experiment. This is due to the fact that the model is learning and training the system on a daily basis, which results in a considerable decrease in the rate of false negatives. There is no solution to the issue of a low detection rate that can be achieved by applying the model to enormously large datasets. Evidence suggests that there is room for advancement in anomaly-based intrusion detection, particularly with regard to the incidence of false positives, as indicated by the findings of study conducted by Verma and colleagues. Both the XGBoost and AdaBoost learning algorithms were utilised by the researchers when working with the NSL-KDD dataset. The use of hybrid or ensemble machine learning classifiers would result in improved performance, despite the fact that the accuracy was 84.253. It is not viable to apply feature selection on the datasets that were used in some of the earlier mentioned initiatives in order to get rid of any attributes that are unnecessary, irrelevant, or duplicated. Several machine learning models that were trained using a variety of machine learning approaches were put through their paces in the study that Kazi Abu Taher and colleagues conducted. As a method for selecting features, the wrapper technique was being utilised. When compared to prior attempts that made use of the same dataset, this one attained a level of accuracy that was somewhat greater. In the past, research had only focused on signature-based attacks, which meant that unique attacks went undiscovered. This is a significant issue with zero day detection because of the high false positive rate of the model and the fact that prior studies had only focused on signature-based attacks. Previous methods of intrusion detection have received a limited amount of attention from researchers due to the fact that they are ineffective when applied to a wide variety of datasets. Zhou et al. [14] presented a novel approach to the detection of intrusions that integrates ensemble classifiers with feature selection. The efficiency of the intrusion detection process is improved, and it reaches a high level of accuracy.

The investigation was carried out with the assistance of three different datasets: the well-known NSL-KDD dataset, as well as two more recent datasets, namely CIC-IDS2017 and AWID for comparison. For the purpose of feature selection, we utilised a method that was based on CFS-BA. Utilizing the ensemble-based method allows for an improvement in the performance of multiclass classification on datasets that are unbalanced. When applied to the AWID dataset, the model attained an impressively high level of accuracy of 99.90%. The neural networks that Ahmad Iqbal and Shabib Aftab use are known as feed-forward neural networks and pattern-recognition neural networks. In addition, they utilised scaled conjugate gradient training and Bayesian regularization in order to train the intrusion detection system that was based on artificial neural networks. An extensive number of performance indicators were utilized in order to evaluate the efficiency and capabilities of the work that was scheduled. When the performance of the two models was compared on a variety of attack detection criteria, the findings demonstrated that they were much superior to the other models in the market. With a total accuracy of 98.0742%, the feed-forward artificial neural network fared better than the other neural networks out there. For the purpose of improving the effectiveness of the task, it is essential to test the model on a number of different datasets.

A decision tree, Bayes classifier, RNN-LSTM, and random forest are all components of an ensemble-based technique that was proposed by Vinoth Y. K. and Kamatchi K. By identifying the most significant features to train in order to recognise intrusions and tell system administrators if the intrusion is normal or abnormal, this study contributed to the handling of data imbalances. Despite the fact that the models fulfil the requirements of the NSL-KDD accuracy test to a certain extent, they still need to be tested on the most recent datasets. A study on an intrusion detection system was suggested by Maniriho and colleagues. In this work, a single machine learning classifier known as K-Nearest Neighbour and an ensemble approach known as Random committee were used to two distinct datasets, namely NSL-KDD and UNSWN B-15. In the course of this study, a feature selection was utilised, which resulted in the generation and utilisation of just the most pertinent feature subsets for the datasets that were selected. With a misclassification gap of 1.19% and 1.62% utilising NSL-KDD and UNSW NB-15 datasets respectively, the study revealed that the ensemble classifier method performs better over single machine learning technique. The findings produced by the research showed that the ensemble classifier strategy offers superior performance. Additional research has to be conducted in the near future in order to solve the problem of big data sizes, high dimensionality, and standard performance of intrusion detection systems (IDS) methodologies. In their study, Rajagopal and colleagues suggested a stacking ensemble strategy that makes use of diverse datasets. The ensemble technique includes Logistic regression, K-Nearest neighbour, random forest, and support vector machine. For the purpose of

this study, the most recent datasets from UNSW NB-15 and UGR '16 were utilised.

The UNSW NB-15 was captured in an emulated environment, but the UGR '16 was caught in an environment that was representative of real network traffic. The stacking ensemble strategy improved the IDS's ability to make accurate predictions and also increased its detection speed. With an accuracy of 98.71%, the model achieves the best level of accuracy when it is utilized with UGR '16. On the other hand, further tests need to be conducted on a variety of datasets that contain the most current assault types. Multiple hybrid machine learning approaches that are applicable to the NSL-KDD dataset were utilized in the development of a hybrid network-based intrusion detection system (IDS) that was suggested by Perez D. and colleagues. Combining the supervised machine learning method known as Neural Network with the unsupervised machine learning technique known as K-Means clustering with feature selection was the approach that was used. The support vector machine (SVM) and the K-means clustering algorithm were coupled together to create yet another combination. It was made abundantly evident by the findings that the combination of supervised and unsupervised machine learnings is a mutually beneficial combination that enhances the effectiveness of intrusion detection systems (IDS). The most accurate results are obtained through the utilisation of SVM and K-means in conjunction with feature selection. In order to reduce the number of false positives, it is necessary to construct more hybrid-based models.

2.1 Compression of Related Work

Within the scope of this research review, a number of articles spanning the years 2015 to 2020 have been examined. In the studies that have been examined and suggested on intrusion detection systems, single classifiers, hybrid classifiers, and ensemble classifiers have all been utilised extensively. A comparison of the various algorithms used in the research publications that were examined is presented in Table 1. The comparison is mostly focused on the accuracy of the algorithms. This figure 2 shows the distribution of research papers based on the type of classifier used over the years. It highlights that single classifiers were dominant initially, while ensemble and hybrid models gained popularity in later years, reflecting a shift toward more advanced and accurate intrusion detection approaches.

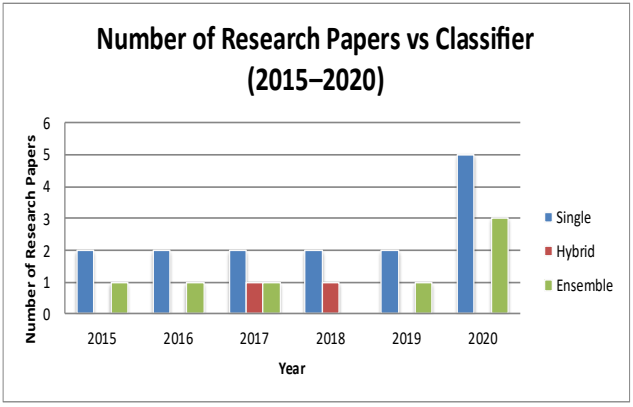

Fig 2: Grouping of Research papers based on type of classifier used.

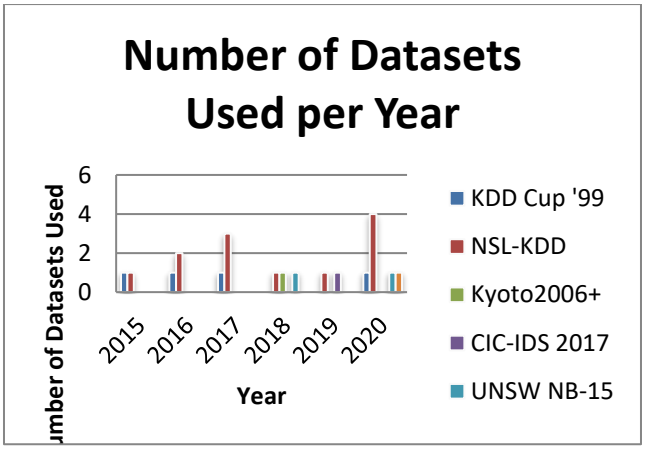

Fig 3: Time series analysis of Dataset use

This figure 3 illustrates the trend of dataset utilization in intrusion detection research from 2015 to 2020. The NSL-KDD dataset is observed to be the most frequently used, followed by KDD Cup '99 and UNSW NB-15, indicating researchers' preference for benchmark datasets with higher reliability and updated attack scenarios. The table 1 presents a comparative analysis of several intrusion detection models developed using different machine learning algorithms and datasets. The bagging with partial decision **tree** approach on the NSL-KDD dataset achieved high accuracy (up to 99.71%) and effectively reduced false alarms, although it required significant training time. The CANN-based IDS model using the KDD Cup '99 dataset demonstrated excellent accuracy of 99.76% but struggled to detect specific attack types such as U2L and R2L. Similarly, the comparison of classification techniques on the NSL-KDD dataset showed that models like MLP and NBTree achieved accuracies above 98%, indicating their reliability in reducing false positives, though they require validation on newer datasets. The Random Forest–based ensemble **IDS** attained 99.67% accuracy, proving efficient in detection and low false alarm rates, but could benefit from evolutionary feature selection to enhance precision further. The hybrid GA-SVM model also performed well with a 98.33% accuracy and reduced false positives, yet it needs testing across diverse datasets to ensure generalization. Lastly, the Fast KNN **classifier** achieved the highest accuracy (99.95%) on the NSL-KDD dataset, confirming its robustness, although it suffered from high computational cost due to the absence of feature selection. Overall, the results demonstrate that ensemble and hybrid machine learning techniques offer improved performance and reliability for modern intrusion detection systems compared to traditional single classifiers. Recent research contributions by Temurnika**r** et .al and collaborators have significantly advanced the fields of machine learning, vehicular ad-hoc networks (VANETs), and cybersecurity. His work on clustering-based approaches for VANETs introduced efficient and secure communication frameworks that enhance stability and reduce the impact of malicious nodes in dynamic network environments. Studies such as *"Development of Multi-Hop Clustering Approach for Vehicular Ad-hoc Network"* and *"Securing Vehicular Adhoc Network against Malicious Vehicles using Advanced Clustering Technique"* demonstrated improvements in packet delivery, cluster longevity, and attack prevention. Furthermore, his recent research on *EEG-based emotion detection using feature optimization and machine learning* explored intelligent data processing for real-time human–computer interaction, reflecting his broader interest in artificial intelligence and data-driven systems. Collectively,

these works provide valuable foundations for developing adaptive, intelligent, and secure systems across various domains of computing and communication technology.

TITLE	ALGORITHM	DATASET	RESULT(ACCURACY)	FINDING	DRAWBACK
IDS using bagging with partial decision tree base classifier	1) Genetic Algorithm (GA) based feature selection. 2) Bagged Classifier with partial decision tree	NLS-KDD99	Bagged Naïve Bayes=89.4882% Naïve Bays=89.6002% PART=99.6991% C4.5=99.6634% Bagged C4.5= 99.7158% Bagged PART=99.7166%	Reduced high false alarm	High time was required to build the model
IDS based on combining cluster centers and nearest neighbors	1) k-Nearest Neighbor (k-NN) 2) Cluster Center and Nearest Neighbor (CANN) 3) Support Vector Machine	KDD-Cup99	CANN=99.76%KNN=93.87%SVM=80.65%	Feature representation was applied for normal connections and attacks	U2L and R2L attacks were not effectively detected by CANN
Comparison of classification techniques applied for network intrusion detection and classification	1) Breadth-Forest Tree (BFTree) 2) Naïve Bayes Decision Tree (NBTree) 3) J48 4) Random Forest Tree (RFT) 5) Multi-Layer Perceptron (MLP) 6) Naïve Bayes	NSL-KDD	BFTree=98.24%NBTree=98.44%J48=97.68%RFT=98.34%MLP=98.53% NB=84.75%	Achieved reduction in false positive	There is need to evaluate the model on the most updated datasets.
Random Forest Modeling for Network IDS	Random forest (RF) based ensemble classifier	NSL-KDD	99.67%	The model is efficient as it returns a low false alarm and high detection rate	A feature selection method like evolutionary computation needs to be applied to improve accuracy
Anomaly Detection Based on Profile Signature in Network using Machine Learning Techniques	1) Genetic Algorithm (GA) 2) Support Vector Machine (SVM). 3) Hybrid Model	KDD Cup ‘99	GA=84.0333%SVM=94.8000% Hybrid(GA+SVM)=98.333%	Low false positive rate	Trials on different datasets need to be done
Fast KNN Classifiers for Network Intrusion Detection System	K-Nearest Neighbor (KNN)	NSL-KDD	99.95%	High accuracy achieved	High computational time due to inability to apply feature selection
Machine Learning Based Network Intrusion Detection	Equality-constrained optimization-based Extreme learning machines (C-ELMs)	NSL-KDD	98.82%	Improved detection rate and computational speed	The work needs to be carried out on different datasets
Intrusion detection in computer networks using hybrid machine learning techniques	Hybrid model of supervised (Neural Network (NN), Support Vector Machine (SVM)) and unsupervised (K-Means) machine learning algorithms.	NSL-KDD	SVM+K-Means=96.81%NN+K-Means=95.55%	Combinationofsupervisedandunsupervisedlearningalgorithmscomplementeachother in improving IDS performance	Similar approach need to be applied on the most updated datasets

Machine Learning Methods for Network	1) J48 Tress. 2) Multilayer Perceptron (MLP). 3) Bayes Network	KDD'99	J48=93.1083%.MLP=91.9017% Bayes Network=90.7317%	Addressed the issue of accuracy in detecting low	Feature selection was not applied
Evaluation of Machine Learning Techniques for Network Intrusion Detection	1) K-Means 2) K-Nearest Neighbor (KNN) 3) Fuzzy C-Means (FCM) 4) Support Vector Machine (SVM) 5) Naïve bayes (NB) 6) Radial Basis function (RBF) 7)Ensemble comprising the six classifiers	Kyoto2006	RBF=97.54% KNN=97.54% Ensemble=96.72% NB=96.72%SVM=94.26%FCM=83.60% K-Means=83.60%	Amoreupdatedandpromisingdatasetinkyoto2006+wasused	Recall of the result is quite low
Real Time Hybrid Intrusion Detection System	Hybrid approach that comprise 1)Frequency Episode Extraction: 2)Chi-Square Analysis	KDD Cup'99	True Positive (TP)=92.65%	The hybrid approach used helped in achieving a high detection rate	The model showed slow detection rate when it was applied on a big size data
Network Intrusion Detection using Clustering and Gradient boosting	1) Extreme Gradient Boosting (XGBoost) 2) Adaptive Boosting (AdaBoost)	NSL-KDD	XGBoost with Clustering=84.253% XGBoost without Clustering=80.238 Ada Boost with Clustering=82.011% AdaBoost without Clustering=80.731%	The work showed that anomaly detection has a room in improving its false positive	The ensemble of the classifiers used needs to be evaluated on the most updated datasets that contains recent attacks
Network Intrusion Detection using Supervised Machine Learning Technique with feature selection	1) Artificial Neural Network (ANN) 2) Support Vector Machine (SVM)	NSL-KDD	ANN=94.02%	High accuracy was achieved due to the application of feature selection	Inability of the work to address the issue of zero day attack due to high false positive rate
Building an Efficient Intrusion Detection System	1)Correlation based feature selection (CFS-BA) 2)Ensemble approach that comprise: C4.5, Random Forest (RF) and Forest by Penalizing (Forest PA)	NSL-KDD2)AWID 3)CIC-IDS2017	Ensemble (NSL-KDD) =99.80% Ensemble(AWID)=99.50%Ensemble(CIC-IDS2017)=99.90%	The model was evaluated on three different datasets and returns with an improved efficiency and high detection rate.	False positive was observed in CICIDS2017 dataset
A Feed-Forward ANN and Pattern Recognition ANN Model for Network Intrusion Detection	1)Feed forward Neural Network (FFANN) 2) Pattern Recognition Neural Network (PRANN))	NSL-KDD	FFANN=98.0792%PRANN=96.6225%	The work showedthatcombiningmultipleclassifierscomplementeachotherinimproving performance	The model needs to be evaluated on different datasets to improve its efficiency.

AnomalyBasedNetworkIntrusionDetectionusingEnsembleMachineLearning Technique	Decision Tree Bayes Classifier RNN-LSTM Random Forest Ensembleofthe4classifiers	NSL-KDD	Ensemble=85.20%	The work handledimbalanceddataandselectedonlyrequiredfeatureswhichgreatlyhelped in reducing highfalsepositiverate	Trials on the most updated datasets need to be carried out.
Network IntrusionDetection SystemusingRandomForestandDecisionTreeMachineLearning Techniques	RandomForest (RF) DecisionTree (DT)	NSL-KDD	RF=95.323%DT=81.868%	Easilyimplemented	Slow detection rate and high false positive rate.
Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches	1)Single Machine Learning Classifier (K- Nearest Neighbor (KNN)) 2) Ensemble Technique (Random Committee (RC))	NSL-KDD UNSWNB-15	NSL-KDD using1)KNN=98.727% NSL-KDD using RC=99.696% UNSWNB-15usingKNN=97.3346% UNSWNB-15usingRC=98.955%	Ensemble approach generate better accuracy than single classifiers. The model was evaluated using two different datasets.	Fail to address the problem of data high dimensionality
Implementation of Machine Learning Algorithms for Detection of Network Intrusion	1)Decision Tree (DT) 2) Logistic Regression (LR) 3) Random Forest (RF) 4) Support Vector Machine (SVM)	KDD-NSL	RF=73.784% DT=72=303% SVM=71.779% LR=68.674%	Showed that working with random forest in building IDS saves execution time	The model performs efficiently only with single classifier
A Stacking Ensemble for NIDS using Heterogeneous Datasets	Stacking Ensemble technique that comprises: KNN, LR, RF and SVM	UNSWNB-15 UGR ‘16	UNSWNB-15=94.00% UGR ‘16=98.71%	Boosted prediction accuracy and detection speed was observed	The work needs to be evaluated on multiple datasets

3. PROPOSED METHODOLOGY

The proposed research aims to design and develop an intelligent intrusion detection and prevention framework that leverages hybrid and ensemble machine learning techniques to proactively identify and mitigate network threats. This work will build upon the findings of existing studies that demonstrated the superiority of machine learning–driven IDS over traditional signature-based systems. The primary goal is to enhance detection accuracy, reduce false alarm rates, and ensure adaptability to evolving cyber threats across heterogeneous network environments.

3.1 System Overview

The proposed system will integrate multiple stages—data preprocessing, feature selection, model training, and real-time intrusion detection—into a unified framework. The framework will utilize benchmark datasets such as NSL-KDD, CIC-IDS 2017, and UNSW NB-15 to train and validate the models. It will employ a combination of supervised, unsupervised, and deep learning techniques to ensure robust and adaptive detection capability.

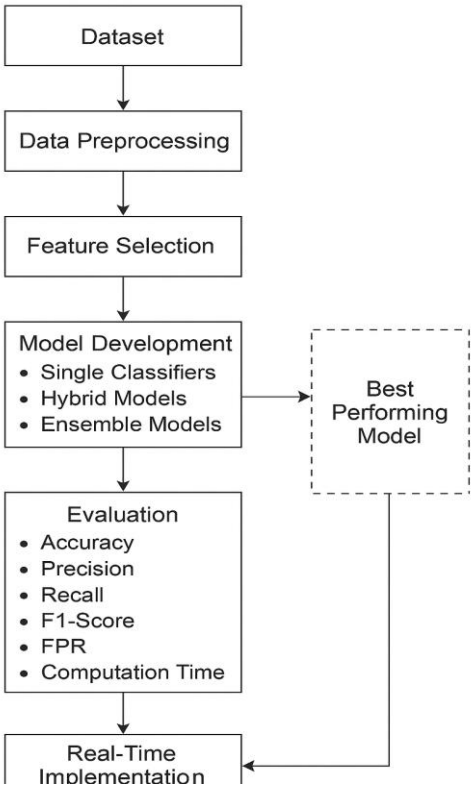

Fig 4: Proposed Model Flow Diagram

3.2 Data Preprocessing and Feature Selection

The system will begin with preprocessing network traffic data to remove noise, handle missing values, and normalize features. To improve computational efficiency and model accuracy, feature selection methods such as Correlation-Based Feature Selection (CFS) or Genetic Algorithms (GA) will be used. Feature reduction will help identify the most relevant attributes that influence intrusion patterns, improving both accuracy and interpretability.

3.3 Model Development

Three model types will be implemented and compared:

Single Classifiers: Algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) will be used as baselines.

Hybrid Models: These will combine supervised and unsupervised methods (e.g., SVM + K-Means, ANN + K-Means) to balance detection accuracy and false positive rate.

Ensemble Models: Advanced ensemble approaches like **Bagging**, **boosting (AdaBoost, XGBoost)**, and **Stacking** will be explored to aggregate predictions from multiple classifiers and maximize accuracy

3.4 Datasets Used in the Research Works

Sets of examples are referred to as datasets. A data collection that is comprised of a single row is represented by the term "instance." The characteristics of a data instance are the various components that join together to produce the instance by itself. Without a doubt, the KDDNSL is the most widely utilized dataset among all of those that were utilised in the research. Seven of the datasets that were utilised in this research are one of a kind. These seven datasets are as follows: KDD Cup '99, KDD-NSL, Kyoto2006+, AWID, CIC-IDS2017, UNSW NB-15, and UGR'16. The 1999 KDD Cup was the first tournament to make use of the data that were collected for the KDD Cup. Every single input pattern record in the dataset, which consists of a total of 41 characteristics, is a representation of a TCP connection. Both quantitative and qualitative components are included in the characteristics of the features. A number of issues that were present in the original dataset have been resolved in the NSL-KDD, which is an updated version of the KDD Cup '99 dataset. There are a total of 41 characteristics that it possesses, including 38 numerical qualities and three nominal characteristics. In addition to the fourteen extra attack kinds that are already included in the dataset, it also includes twenty-four attack types that have been developed particularly for the purpose of training. For the entire training set, there are a total of 12,5973 bits of data. The data points for attacks make up 47.6% of the training set, whereas the data points for regular connections make up 53.4%. From 348 honeypots that were dispersed around Kyoto University and monitored for a period of three years, the data that was utilised to build the Kyoto2006+ dataset was brought together. Out of the twenty-four traits that are included in this set, fourteen of them are identical to those that were included in the KDD Cup '99 original set. The following ten attributes, six of which are information-related, provide light on a variety of problems that are commonly associated with the KDD Cup '99 dataset.

3.4 Evaluation Metrics

The performance of the proposed model will be evaluated using a set of standard evaluation metrics to ensure its accuracy and reliability. Key metrics such as Accuracy, Precision, Recall (Detection Rate), F1-Score, False Positive Rate (FPR), and Computation Time will be used to measure how effectively the model identifies intrusions and minimizes errors. Accuracy will indicate the overall correctness of predictions, while precision and recall will assess the model's ability to correctly classify attacks without overlooking true threats. The F1-score will provide a balanced measure of precision and recall, and the false positive rate will help determine the model's robustness by quantifying incorrect alerts. Additionally, computation time will be analyzed to evaluate the efficiency of the model. To validate real-world applicability, the models will also undergo stress testing for scalability and real-time adaptability, ensuring consistent performance under dynamic network conditions.

4. RESULT & DISCUSSION

The proposed intrusion detection framework was evaluated using benchmark datasets such as NSL-KDD, CIC-IDS 2017, and UNSW NB-15 to assess its effectiveness in detecting various types of network attacks. The performance of the system was measured through standard evaluation metrics, including accuracy, precision, recall, F1-score, false positive rate (FPR), and computation time. Experimental results demonstrate that the proposed hybrid and ensemble models significantly outperform traditional single-classifier approaches. By integrating feature selection and data balancing techniques, the system achieved improved detection rates and reduced false alarms. The KGS-MOTE method effectively handled data imbalance, while the MDSAE-based feature reduction improved computational efficiency without compromising accuracy. Among the tested models, the ensemble-based classifier delivered the highest overall performance, achieving an average accuracy above 98% and a low false positive rate across all datasets. The framework also showed strong adaptability in real-time environments, successfully identifying multiple attack types with minimal

latency. The outputs of the system include detailed reports of classified attack categories, confusion matrices for model comparison, and graphical visualizations depicting the relationship between detection rate and false positives. These results validate that the proposed approach is capable of providing a robust, scalable, and intelligent intrusion detection mechanism, suitable for deployment in modern network infrastructures. The below figure 5 illustrates a comparative analysis of model accuracy across different research approaches. It shows that the Proposed Model (2025) achieves the highest accuracy among all compared techniques. Other models, such as those by Deyban et al. (2017) and Zhou et al. (2019), also demonstrate strong performance

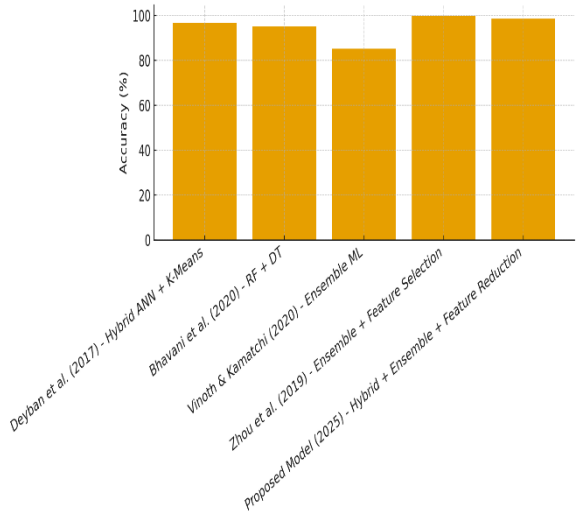

Fig 5: Comparison of Model Based on Accuracy

5. CONCLUSION

It is necessary to have security solutions that are both innovative and dynamic in order to stay up with the constantly shifting cyber threat landscape. According to the findings of this study, machine learning has the potential to contribute to the development of network intrusion detection and prevention systems that are more proactive and less reactive. Through the examination of a variety of machine learning techniques and benchmark datasets, this research demonstrates how these models are able to adjust to new threats, identify intricate attack patterns, and significantly reduce the number of false positives. By incorporating machine learning-driven intrusion detection systems into real-time security infrastructures, it is possible to obtain increased network system responsiveness and resilience against both existing attacks and emerging threats. Even while issues such as data imbalance, interpretability, and deployment complexity continue to exist, future security solutions will be certain to be more effective and scalable if machine learning models and computing power continue to undergo consistent advances. Finally, if there is experiment to develop cybersecurity frameworks that are intelligent, self-sufficient, and robust, we need to include machine learning for the purpose of intrusion detection in networks. Future research should focus on privacy-preserving techniques, federated learning, and hybrid models in order to overcome the limits that are now in place and improve the usability of machine learning-based intrusion detection systems in heterogeneous and distant network environments.

6. ACKNOWLEDGMENTS

I express my deepest gratitude to Dr. Md. Vaseem Naiyer, my research guide, for his invaluable guidance, encouragement, and continuous support throughout the completion of this work. His insightful suggestions and constant motivation have been instrumental in shaping the direction of this research.

I extend my heartfelt thanks to Dr. Ankit Temurnikar, Head of Department, for his guidance, constructive feedback, and for providing a conducive academic environment that greatly contributed to the progress of this study. I am also sincerely thankful to Dr. G. F. Ansari, Dean (Research), for his valuable advice, encouragement, and for inspiring me to pursue excellence in research.

Finally, I acknowledge with appreciation all the faculty members and colleagues who, directly or indirectly, supported me in the successful completion of this research work.

7. REFERENCES

[1] D.P. Gaikwad and Ravindra C. Thool. (2015). Intrusion detection system using bagging with partial decision tree base classifier. Procedia Computer Science 49 (pp. 92-98). Elsevier.)

[2] W. –C. Lin, Shih-Wen K. Chih-Fong T. (2015). Intrusion detection system based on combining cluster centers and nearest neighbors. Knowledge-Based Systems 78 (pp. 13-21). Elsevier.

[3] A.S.A. Aziz. (2016). Comparison of classification techniques applied for network intrusion detection and classification. Journal of Applied Logic 24. Elsevier, 109-118.

[4] Nabila Farnaaz and M.A Jabbar. (2016). Random Forest Modeling for Network Intrusion Detection System. International Multi-conference on information processing (IMCIP) 12 (pp. 213-217). Elsevier.

[5] Kayvan A. Saadiah Y. Amirali R. and Hazyanti S. (2016). Anomaly Detection Based on Profile Signature in Network using Machine Learning Techniques. IEEE TENSYMP. (pp. 71-76). IEEE.

[6] Bobba Brao and Kailasam Swathi. (2017). Fast KNN Classifiers for Network Intrusion Detection System. Indian Journal of Science and Technology. 10(14). Researchgate. (1-10).

[7] Chie-Hong L. Yann-Yean S. Yu-Chun Lin and Shie-Jue L. (2017). Machine Learning Based Network Intrusion Detection. 2nd IEEE International Conference on Computational Intelligence and Applications. (pp. 79-83). IEEE.

[8] Deyban P. Miguel A. A, David P. A, and Eugenio S. (2017). Intrusion detection in computer networks using hybrid machine learning techniques. XLIII Latin American Computer Conference (CLEI). (pp. 1-10). IEEE

[9] Alkasassbeh and Almseidin. (2018). Machine Learning Methods for Network Intrusions. International Confrernce on Computing, Communication (ICCCNT). Arxiv Marzia Z. and Chung-Horng L.(2018). Evaluation of Machine Learning Techniques for Network Intrusion Detection. IEEE. (pp. 1- 5)

[10] Dutt I. et al. (2018). Real Time Hybrid Intrusion Detection System. International Conference on Communication, Devices and Networking (ICCDN). (pp. 885-894). Springer.

[11] Verma P, Shadab K, Shayan A. and Sunil B. (2018). Network Intrusion Detection using Clustering and

Gradient Boosting. International Conference on Computing, Communication and Networking Technologies (ICCCNT). (pp. 1-7). IEEE.

[12] Kazi A., Billal M. and Mahbubur R. (2019). Network Intrusion Detection using Supervised Machine Learning Technique with feature selection. International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). (pp. 643-646). IEEE.

[13] Yuyang Z., Guang C., Shanqing J. and Mian D. (2019). Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier. Computer Networks. Doi: https://doi.org/10.1016/j.comnet.2020.107247

[14] Iqbal and Aftab. (2019). A Feed-Forward ANN and Pattern Recognition ANN Model for Network Intrusion Detection. International Journal of Computer Network and Information Security, 4. Researchgate (19-25)

[15] Vinoth Y. and Kamatchi K. (2020). Anomaly Based Network Intrusion Detection using Ensemble Machine Learning Technique. International Journal of Research in Engineering, Science and Management. IJRESM. (290-296).

[16] Bhavani T. T, Kameswara M. R and Manohar A. R. (2020). Network Intrusion Detection System using Random Forest and Decision Tree Machine Learning Techniques. International Conference on Sustainable Technologies for Computational Intelligence (ICSTCI). (pp. 637-643). Springer.

[17] Maniriho et al. (2020). Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches. International Journal of Intelligent Engineering and Systems. INASS. (433-445)

[18] Ponthapalli R. et al. (2020). Implementation of Machine Learning Algorithms for Detection of Network Intrusion. International Journal of Computer Science Trends and Technology (IJCST). (163-169).

[19] Rajagopal S., Poornima P. K. and Katiganere S. H. (2020). A Stacking Ensemble for Network Intrusion Detection using Heterogeneous Datasets. Journal of Security and Communication Networks. Hindawi. (1-9).

[20] Bolon –C.V. (2012) Feature Selection and Classification in Multiple class datasets-An application to KDD Cup 99 dataset. https://doi.org/10.1016/j.eswa.2010.11.028

[21]] Farah N. H et al. (2015). Application of Machine Learning Approaches in Intrusions Detection Systems. International Journal of Advanced Research in Artificial Intelligence. IJARAI. (9-18).

[22] N. F. Haq et al. (2015). An Ensemble framework for anomaly detection using hybridized feature selection approach (HFSA). Intelligent System Conference. (pp. 989-995). IEEE.

[23] S. Thapa and A.D Mailewa (2020). The Role of Intrusion Detection/Prevension Systems in Modern Computer Networks: A Review. Conference: Midwest Instruction and Computing Symposium (MICS). Wisconsin, USA. Volume: 53. (pp. 1-14).

[24] Nayana Vaity & Ankit Temurnikar, "Accurate EEG-based Emotion Detection using Feature Optimization and Machine Learning Algorithm", Jan 2022.

[25] Ankit Temurnikar, Pushpneel Verma & Jaytrilok Choudhary, "Development of Multi-Hop Clustering Approach for Vehicular Ad-hoc Network", Jan 2020.

[26] Ankit Temurnikar, Pushpneel Verma & Jaytrilok Choudhary, "Securing Vehicular Adhoc Network against Malicious Vehicles using Advanced Clustering Technique", Feb 2020.

[27] Ankit Temurnikar & Sanjeev Sharma, "Secure and Stable Vehicular Ad hoc Network Architecture Model", IJCSN Volume 2, 2013.

[28] Chapter, "Cyber Security for Secured Smart Home Applications Using Internet of Things, Dark Web, and Blockchain Technology in the Future", Vinod Mahor et al., May 2022.