|| Volume 7 || Issue 9|| November 2023 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

# Intrusion Detection for real time Network Dataset using PCA and Random Forest Algorithms

**Mayuri Sanjay Narudkar, Prof. Anita Mahajan, Dr. Pankaj Agarkar,**

PG Student: Department of Computer Engineering, Ajeenkya DY Patil School of Engineering Pune

Prof.: Department of Computer Engineering, Ajeenkya DY Patil School of Engineering Pune

HOD: Department of Computer Engineering, Ajeenkya DY Patil School of Engineering Pune

**Abstract:** Ensuring robust network security is of utmost importance in the present era. To safeguard the integrity of in-network systems, numerous architectural solutions have been suggested to prevent unauthorized access by both internal and external users. Several techniques have been created to identify harmful activity on targeted machines. In some cases, an external user may engage in malicious behavior and gain illegal access to these devices. Such behavior is classified as malicious activity or intrusion. Several machine learning and soft computing algorithms have been developed to identify activities in real-time network log audit data. The data sets KKDDCUP99 and NLSKDD are commonly used to identify intruders in benchmark data sets. This study presents a method for detecting and identifying unauthorized individuals using machine learning methods. Two distinct methodologies have been suggested, namely signature-based detection and anomaly-based detection. The experimental investigation showcases the application of Principal Component investigation (PCA) and Random Forest (RF) algorithms on different data sets. It also evaluates the performance of the system in a real-time network context.

**Keywords:** Intrusion Detection System, Network security, Naïve Bayes, PCA, Artificial Neural Network, KDDCUP99.

## I. INTRODUCTION

The Intrusion Detection System (IDS) is designed specifically to identify and detect a single category of assault, such as a Sample or unknown attack, Denial of Service (DoS) attack, User to Root (U2R) attack, or Remote to Local (R2L) attack. Subsequently, it systematically activates a series of these subsystems, individually and in order. This serves a dual purpose: Initially, it is possible to train only a restricted set of features that can identify a specific sort of attack in each sub-phase. Furthermore, the sub-size device retains its small dimensions, thereby ensuring its functionality. Like our system, a typical drawback is that it increases the amount of communication between modules. By ensuring that each sub-phase is entirely autonomous from the other layers, we can effectively prevent this issue in our approach. Therefore, these traits may occur in multiple sub-phases. If a sub-phase is discovered without a central decision maker, it will effectively prevent an attack, based on the network's security policy. When established inside a certain layer, different subphases primarily function as filters that prevent abnormal associations. This allows for a quick response to intrusions and reduces the analysis required in later stages. It is important to acknowledge that in various sub-phases where attacks based on visual cues are used, distinct responses are frequently triggered. As assaults are discovered and thwarted, the amount of system-analyzed auditing information reduces at each tier and succeeding stage. If no attacks are detected before the final sub-phase, all staggered sub-phases in phase 2 will have an equal burden. Nevertheless, once attacks are identified and obstructed using any subsequent technique, the anticipated average workload is likely to be considerably reduced. Conversely, when the sub-phases are organized in parallel rather than in a sequential manner, within a subsystem's sequence configuration, the load is equivalent to the most unfavorable scenario. Repeating the initial step in a sequential setup can be done to achieve load

balancing and enhance performance. The study effort establishes a way for generating PCA rules by employing a role selection procedure that operates on both HIDS and NIDS systems. A genetic algorithm is a type of optimization technique that is employed to identify the optimal solution. The utilization of an ensemble technique, employing several classification algorithms, may yield the most effective detection of Network Intrusion Detection Systems (NIDS) for all categories of master class sub-attacks. The objective of our proposed study is to establish unambiguous guidelines and enhance the detection rates of Denial of Service (DOS), Probing (PROBE), User to Root (U2R), and Remote to Local (R2L) attacks for Network Intrusion Detection Systems (NIDS) and Host Intrusion Detection Systems (HIDS).

## II. LITERATURE SURVEY

The authors of the paper are Bhosale, Karuna S. et al. [1]. A deep neural network (DNN) Research is being conducted on developing deep learning systems that can efficiently and effectively detect and classify inadvertent and unexpected cyber-attacks, with a focus on scalability. The continuous evolution of network operations and the swift emergence of attacks necessitate the examination of recurring incidents that have been generated over the years using both dynamic and static methodologies. This study enables the identification of the appropriate algorithm that can effectively predict future cyber-attacks. An extensive assessment of the DNN experiments and other robust machine learning classifiers is demonstrated on multiple publicly available test malware databases. The system selects the optimized network parameters and modulation techniques for DNNs using the KDDCup 99 dataset. The number of hidden layers is determined by the system.

In their study, Chamou et al. [2], The science community has become accustomed to the complexity and improved efficiency of intrusion detection systems, which has led to numerous companies worldwide being targeted and threatened by the constant emergence of new threats. This application employs advanced deep learning models to assess suspicious activity in DDoS and malware cyber threats, making it a pioneering solution in the field. The fast proliferation of web apps and their widespread use has led to a consensus among internet users that cybersecurity effectiveness, data protection, and secure communication are of utmost importance. Concurrently, there has been a confirmed growth in the occurrence of increasingly advanced security risks across computer systems and internet networks in the academic and industrial digital realm, particularly affecting small and medium-sized firms (SMEs), with significant economic consequences.

As stated in reference [3], a system has been created to effectively identify possible assaults by employing decision-free techniques, random forest algorithms, and K-nearest neighbors (KNN) methods. A novel methodology is proposed to address the limitations of the prior method, which was unable to identify IPV6 assaults. The established framework for detecting IPV4-based assaults yields amazing and efficient results, taking into account the potential scope and the evaluated efficiency of various techniques. Measurements were taken to assess the consistency, accuracy, and recall % of detection.

According to the source [4], the successful utilization of clustering, KDD, and the identification of a novel phenomena called NEC can be achieved. An unsupervised anomaly detection method is employed to achieve high detection rates while minimizing false positive rates. It is a viable approach to address the problem and identify an irregularity without the need for gathering labeled data. Utilize the 2009 NSL-KDD dataset for the purpose of evaluating the system. The preprocessing model transforms all features into numerical values, and the standardized dataset will compare the predicted outcome with a precise result in the evaluation portion.

A survey on data mining and machine learning is undertaken to enhance cybersecurity in the field of intrusion detection. The purpose of the survey is to assure cybersecurity in relation to the CSID Data Mining and Machine Learning Survey[5]. The packet header and net flow packet header are utilized in networks and kernel-level data for the intrusion detection process. However, a possible challenge arises as data mining and artificial intelligence cannot be accomplished without real databases, resulting in significant time consumption. The discussion revolves around the existence of diverse statistics and

machine learning techniques. The study paper presents a set of criteria for comparing different approaches used in machine learning data mining for intrusion detection. This type of detection is focused on identifying, assessing, and acknowledging any unauthorized activities such as misuse, duplication, modification, or destruction of an information system.

The implementation of machine learning, ranking, and Voronoi techniques for detecting and monitoring intruders has resulted in enhanced security. The dataset's size and the detection's high precision are noteworthy. The ISOT dataset was utilized, taking into consideration the processing latency. The paper utilizes network flow patterns to forecast the invasion of botnets, leveraging the packet compliance feature that aids in packet encryption.

ADS-B IDS proposes the use of automated surveillance-broadcast systems based on ADS-B. The HMAC data set is utilized to improve the performance of air traffic control. The approaches function with minimal additional costs or burdens. The future scope indicates that its distance from the matching initial position must fall inside the safe zone in order for the ADS-B location to be considered genuine. GPS utilizes the cyber-physical environment, which is verified by attack detection, to establish radar requirements for aircraft, resulting in highly accurate location precision. ADS-B has emerged as a viable replacement to the existing radio system. A secure protocol is proposed for the secure exchange of keys used in the HMAC technique. The ATC Centre establishes secure connections with ATCs responsible for monitoring different zones along the flight route in order to exchange the private key via public key infrastructure (PKI) networks.

As per the reference [8], the paper indicates that a secure process mining technique is employed to develop a combined Intrusion Detection System (IDS) using data mining for a power device with data logging capabilities. The technique is a comprehensive way for designing the IDS prototype. One of the main benefits has been the high level of detection accuracy, reaching up to 73%. However, this strategy is insufficient for gathering significant problems like data logs, notwithstanding its complexity. The framework utilizes the capability of intrusion detection systems (IDS) based on characteristics and specifications. The data analysis process that consolidates audit logs to ascertain the normal trajectory taken by various machine units. With the automated method, there is no requirement to manually analyze and code the sequence.

Al states that Dr. Yogesh Kumar Sharma et al. have provided a definition of a 6G Network Access system and an Edge-Assisted Congestion Rule Mechanism using Software-Defined Networking [10]. In order to prevent congestion in the flow of traffic, the framework needs to handle high-frequency data. The evaluation results of this proposed approach suggest that it has the potential to optimize the effectiveness of the network. Furthermore, the network's security is a pressing concern, as a centralized security framework is crucial for ensuring the dependability of data throughout the network. Protection is essential because users transmit and link, increasing the risk of data corruption or piracy. Another method of enhancing network performance in wireless broadband is through network optimization, which can mitigate any potential slowdowns and provide high-speed connectivity for individuals. The software recommends the cutting-edge System Slicing Edge Admission System paradigm.

This article employs the Artificial Neural Network (ANN) of an Operating System Sensor to oversee malevolent actions in Android and iOS devices. It relies on the Flow Anomaly System [11], which is grounded on the Flow Anomaly Detection Platform for Android mobile devices. The detection rate and accuracy of this technique are 85% and 81%, respectively. Impersonation is evaluated based on CPU use, storage requirements, and improved visibility. This aids in developing a compact, adaptable, and efficient Intrusion Detection System (IDS) following an Integration node to counteract public attacks from different services. The data sources are analyzed using sophisticated data mining methods. Enhancing the precision and rate of classification necessitates consideration of future prospects.

PRADEEP and Dr. Yogesh Kumar [12] An efficient and secure method for implementing the Internet of Things using a combination of Fog Computing and Mobile Cloud Architecture. This paper assesses the

performance of cloud computing using the simulation model iFogSim. The model incorporates artifacts and cloud services to ensure a high level of consistency and precision.

Javier A. and colleagues suggested in their study [13] that information security can be enhanced by incorporating virus detection in a network setting. The platform proposed would be an efficient technique for malware detection in Ghana Limited's application security, thanks to its wide framework. In the concluding study, the participants have already developed a sense of assurance and satisfaction regarding the dependability and effectiveness of the subject matter. The study demonstrated that this technology successfully achieved the objective of the experiment. "High Quality" examined the procedures and resolution of the suggested approach. The successful implementation of the malware detection system by Asia Technology Security ensured the company's continued market leadership.

In their study, Bholanath Mukhopadhyay et al. [14] developed a novel method for task scheduling and protection in cloud-based Infrastructure as a Service (IaaS) applications. Their technique involved the implementation of both SSL-based protection and authorization access restrictions. We have also incorporated the feature of a search option for selecting Endpoint Protection. Our design allows for the creation of many profiles, each with its own distinct access policy, tailored for diverse network applications. To illustrate, a policy can be defined for authentication of internet connectivity for dynamic access point connections. By employing our unconventional methodology, it is feasible to rapidly categorize the user, determine the location of the customer, assess the prevailing network conditions during the connection, and evaluate the state of the server.

Self-Taught Learning (STL) with an inter PCA has been employed, as stated in reference [15], to maintain the best level of accuracy in Intrusion Detection Systems (IDS), except in unfamiliar scenarios. The Neural Learning method, which is based on deep learning, models vehicle routing to enhance the resilience of homes. It employs a self-healing technique throughout the IDS recovery phase. The simulation findings demonstrate the effectiveness of the suggested Intrusion Detection System (IDS) in safeguarding Unmanned Aerial Vehicles (UAVs) against cyber security assaults, specifically in terms of precision and accuracy. This technology has the capability to proactively mitigate potential cyber-attacks, such as GPS hacking, disruption of a drone's communication patterns, and unauthorized access or control of the drone.

A robust framework is implemented to counter cyber security threats, even in situations where signals are absent or unforeseen attacks occur. This framework utilizes a learning algorithm to safeguard every unmanned aerial vehicle. To ensure a secure journey back home or to the closest safe area, it is necessary to develop software that incorporates a deep learning model with a multi-class Support Vector Machine. This software should be capable of efficiently detecting both large-scale and small-scale attacks on surveillance drones (UAVs), while still being lightweight and responsive.

[16] provides a concise overview of the methodology, datasets, recent advancements in Re-ID, obstacles faced, as well as the approaches and strategies employed in computer vision systems. Common applications of re-identification (re-Id) technology include video surveillance, pedestrian enumeration, multi-camera tracking, and analysis of activities captured by several cameras. The objective of it is to categorize persons.

The photo depicts a solitary image, devoid of any physical examination or treatment-related symptoms. The purpose of this is to offer temporal, spatial, dynamic, and demographic information about individuals in a densely populated setting. The primary obstacle in this situation, in contrast to detection, is the constraint of time. The system identified a specific person within a group of individuals captured by a camera network located at different places with different perspectives. The most advanced results are achieved by utilizing CNN-based approaches for the individual re-identification process. The essential procedure of identifying individuals based on their physical characteristics is referred to as individual re-identification (re-ID). This analysis focuses on the recent advancements in deep learning algorithms and their use in detection. This has also tackled the ongoing investigations, assignments, and challenges of

individual re-identification. It becomes a complex undertaking as applications begin to proliferate in a diverse range of forms. Individual re-Identification is gaining attention not only in academia but also due to its wide range of applications in industries.

The article [17] proposes a detailed experimental investigation using various binaries to enhance the detection rate and minimize errors. Additionally, several experiments were conducted on intrusion detection systems using the Kddcup'99 dataset. The majority of them failed to effectively identify incursion with the new intrusion due to the inadequacy of the old dataset in safeguarding against contemporary threats. The primary goal of this system is to identify zero day attacks by utilizing a combination of Deep Learning (DL) and Binary algorithm (BA) for Anomaly Classification of IDS. It also showcases the precision of detection on various synthetic datasets such as KDDCUP99 and NSLKKD. In addition, this study utilizes deep neural network (DNN) to classify anomalies in intrusion detection systems (IDS). The deep learning (DL) platform and binary algorithms (BA) are employed, specifically the conditional bat algorithm, binary evolutionary algorithms (BGA), and binary magnetic simulated annealing, to optimize the detection speeds. The achieved outcomes for both DNN and the hybrid variant encompassed accuracy, retrieval, correctness, classification error, tolerance, specificity, with the exception of cost error. The study researcher will employ visualization techniques to present the results. Additionally, this might be employed for doing laboratory tests and executing operations with the aid of MATLAB. In addition, researchers generate graphs and diagrams to serve as a computational tool and illustrate the outcomes of the studies.

An intrusion detection strategy is employed in [18] utilizing a deep learning model capable of discerning various types of attacks without the need for manually created rules or signature mapping. This utilizes RNN, Stacked RNN, and CNN supervised deep learning methodologies to detect five prevalent forms of threats by employing Keras on the pinnacle of TensorFlow. This technique solely necessitates understanding of the packet header and does not necessitate any payload from the user. The performance of Snort IDS is evaluated and compared with the usage of MAWI datasets, which are Winpcap files. The findings indicate that Snort was unable to identify the network scan attack over ICMP and UDP protocols due to the absence of user payloads. This system demonstrates that RNN and CNN can effectively detect port scan assaults, ICMP network scans, UDP network scans, and TCP network scans. Additionally, it achieves a high level of accuracy in detecting DoS and DDoS attacks. By employing deep learning techniques with TensorFlow, it is capable of identifying five prevalent attack categories, namely DoS, port scan, network scan via UDP, network scan via TCP, and network scan via ICMP. This approach, combined with RNN, achieves the utmost accuracy in detecting both network and host-based attacks. The proposed system will be analyzed and compared based on the output matrix.

In [19], a comprehensive and comparative analysis was conducted on the NSL-KDD and CIDDS-001 benchmark network audit datasets using machine learning classification techniques. Prior to implementing self-learning (Machine or Deep Learning) classification algorithms, this strategy utilized hybrid feature selection and ranking strategies to attain best outcomes. Some examples of machine learning algorithms are Principal Component Analysis (PCA), Naïve Bayes, k-Nearest Neighbors (k-NN), Neural Networks, Deep Neural Networks (DNN), and Denoising Autoencoders (DAE). The performance of IDS has been assessed using significant performance indicator criteria, such as accuracy.

As stated by [20], they employ deep learning algorithms, a novel method for detecting attacks, to analyze smart meter traffic and identify any attempts. The suggested approach incorporates intricate multi-layer techniques, arranged in a hierarchical framework to effectively detect adversaries. The analysis focuses on comparing the performance of the integrative solution learning algorithm and deep relational features with classifiers using the standard synthetic dataset in the proposed system. An exploit of network traffic analysis was implemented as a crucial security mechanism to detect intrusions. To inhibit the utilization of the feature selection technique in the training database. The simulation framework findings indicate that the DL-based IDS, which was specifically created, effectively and efficiently detect and execute assaults at a rapid pace. The equations are arranged in a chronological

manner according to the significance of the attacks. Every classifier is designed to be compositional and capable of identifying certain attack patterns within the network's activity data. The IDS employs a classification framework that is determined by the cost associated with each sort of assault. Once the categorization model detects the presence of malicious packets in a connection, the data is transmitted to the user. The data is available for additional examination.

Study [21] discusses the utilization of computer and machine learning technologies in Wireless Sensor Environments for IDS systems. The Deep Boltzmann Computer Distributed DBCD-IDS is introduced. Implementing numerous Intrusion Detection System (IDS) techniques is necessary for Wireless Sensor Networks (WSNs) in order to monitor and secure vulnerable systems effectively. The performance of RBC-IDS is analyzed and contrasted with the previously proposed efficient machine teaching IDS, namely the Adaptively Monitored and Grouped Hybrid IDS. The results of this study demonstrate that both RBC-IDS and ASCH-IDS achieve high levels of accuracy in duplicate detection. However, RBC-IDS outperforms ASCH-IDS in terms of detection capability. Security risks, including as breaches in network systems and compromised sink nodes, can arise in both cyber and digital portfolios. Remote Monitoring is a crucial computer security solution that has been established to address external attacks on communication systems and promptly detect any unauthorized infiltration. The Restricted Boltzmann Machine (RBM) is a neural network consisting of two layers: a visible layer (V) and a hidden layer (H).

The paper [22] explores the use of a classifier called Deep Naive Bayes to develop an IDS (Intrusion Detection System) that is both scalable and efficient in detecting and identifying unforeseen cyber-attacks. It is crucial to assess various datasets generated by both basic and complex strategies over the years due to the continuous evolution of network architecture and rapid advancement in attack techniques. This type of research facilitates the exploration of the optimal algorithm capable of efficiently detecting future cyber-attacks. A comprehensive evaluation of observations on datasets and other robust machine learning PCA classifiers was conducted on various publicly accessible benchmark minicomputers. The KDDCup 99 dataset was utilized to employ advanced configuration selection methods, resulting in the identification of optimal system setups and networking protocols for DNNs. The DNN tests were conducted for a maximum of 1,000 iterations, using a learning rate that ranged from 0.01 to 0.5. The DNN template that demonstrated effectiveness in the KDDCup 99 also enhanced the performance of the test on specific datasets such as NSL-KDD, UNSW-RF15, Kyoto, WSN-DS, and CICIDS 2017. By traversing multiple concealed layers, this deep neural network (DNN) model may acquire knowledge about the intricate yet very detailed picture classification of the Intrusion Detection System (IDS) data. Rigorous experimental research has confirmed that DNNs outperform traditional classifiers. An Artificial Neural Network (ANN) category is described as a directed graph that facilitates the transfer of various display adapters via edges, passing through a node without creating a cycle. Recurrent neural network

As per the findings of [23], the IDS-DLA framework utilizes a formulation and construction method to identify malicious activities by analyzing the structural steel parts. There is a substantial quantity of forms that have been deemed beneficial, particularly those that utilize composite materials. IDS-DLA utilizes Mathematical and CNN triplet filters to achieve accurate intrusion recognition and high precision from the Point cloud database. The IDS-DLA utilizes the Hu moments ranking method to select its top 5 ranking forecasts as the final results achieved. Undoubtedly, the experiments revealed that the current standards have acquired a significantly higher level of precision compared to prior ones. The utilization of a hybrid feature extraction approach and a multi-filtering approach leads to the attainment of superior efficiency. The primary determinants will be the synchronization of individuals, resources, and the respective machinery with accurate formulas to enhance or augment the total efficacy of the manufacturing lines.

When utilized together with [24], IDS's resilience in the context of computer vision is enhanced against intrusion detection. The term "min-max method" in the UNSW-RF 15 dataset is used to train intrusion detection systems against crown prosecution samples. Alternatively, this approach employs the current minimum technique as a security measure to enhance the detection mechanism that minimizes the loss incurred by the training data of the integrated adversarial samples. This study examines and quantifies the

effectiveness of malicious attack strategies and the resilience of training pictures against these attacks. This system utilizes remote attack methods that are specifically designed for binary systems. It is recommended to deploy this system in a secure and reliable environment to create network attacks. In conclusion, eliminating the major component analysis function will enhance the resilience of the sensor network by employing a learning algorithm. The dataset consists of 49 attributes for each column, including the class name. You select the optimal 28 features, which are commonly referred to as feature ranking or score. Regularization is employed to choose many features that have a wider range of characteristics in order to improve the classifier's efficiency and dependability.

Flow-dependent anomaly. The user's text is "[25]". This study presents a flow-anomaly intrusion detection framework for Android Mobile Devices that utilizes Artificial Neural Networks (ANN) within the Android operating system to identify abnormal behavior in Android devices. The detection rate for this approach is 85% and 81%, respectively. The emulation of the central processing unit (CPU), memory, and battery capacity is taken into account. The objective of this study is to categorize efficient, adaptable, and effective Intrusion Detection Systems (IDS) that may be used in an Android context to detect and mitigate various types of public attack services. The data streams are analyzed using sophisticated machine learning algorithms. In the future, there will be an enhancement in the rate of detection and precision.

A clandestine Markov Intrusion Detection System (IDS) is established, as proposed[26], utilizing Software-Defined Networking (SDN). Through the analysis of the interconnected network structures within a given system and the implementation of Artificial Neural Network Intrusion Detection Systems (ANN IDS), the Software-Defined Networking (SDN) network is capable of monitoring and ensuring the overall security of the system. The paper offers advantages such as enhanced conduct and heightened security. In order to enhance the HMM vector's capabilities in assessing the harmful nature of a dataset, it was utilized to indicate the likelihood of an application being a threat to network security.

This study presents a defined PS-Poll DOS assault intrusion detection system for 802.11 networks. The system relies on a single real-time event system, as described in reference [29]. This approach employs Real-Time Detection and Response System (RTDES) on a discrete event system to promptly identify Denial of Service (DOS) attacks as they occur. An important advantage is the elevated rate of detection and accuracy, whereas the primary drawback is the absence of frames. PS-DOS attack detection involves monitoring for alterations in protocols or the installation of specialized hardware.

The user's text is "[30]". Cybersecurity has been acknowledged as a significant issue in the digital realm. The research showcases a cognitive neuromorphic computer technology for the cyber safety network identification method using a deep learning approach. This method employs the process of differentiating vectors into factors. The NSL-KDD dataset enhances precision and score by a significant margin of 90.12% and 81.31% respectively. Deep learning is a very effective learning approach that integrates classification features and excels in recognition tasks. The future challenge is to determine the appropriate interpretation to be employed in the spike format of the existing northern data structure.

## III. PROPOSED SYSTEM

The research methodology employed machine learning techniques to conduct intrusion detection and prevention. The training for anomalous and remote monitoring will be conducted by the packet environments characterized block. Subsequently, it will propose a compilation of functions for a certain packet action. If everything is in good condition, proceed collectively. In order to identify individual attacks, samples of misbehavior will undergo feature selection to assess different characteristics. The suggested system consists of two phases: training and testing. We have utilized a network dataset for these purposes. The framework comprises many components. Figure 1 illustrates the complete implementation of the system through the utilization of specified algorithms. Different machine learning techniques have been employed to create training and testing modules, respectively.

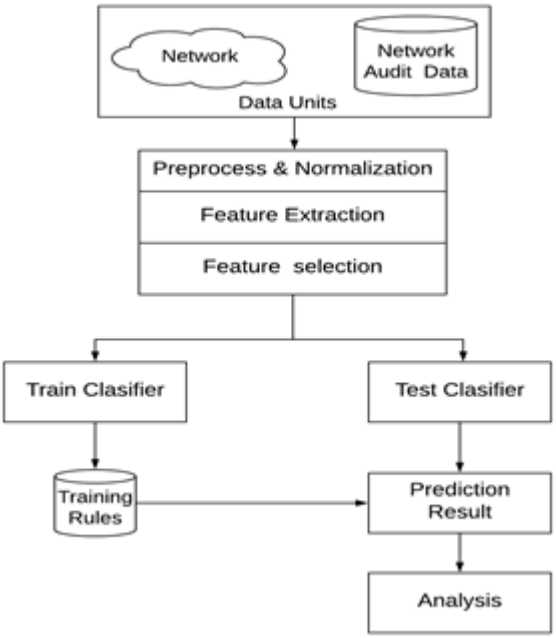


**Figure 1 : Proposed system architecture**

The NSL KDD CUP 1999 dataset is utilized for conducting the experiments, specifically the KDD data set from 1999. The KDD CUP 1999 dataset is a modified version of the MIT Lincoln Laboratory's original 1998 DARPA intrusion detection assessment software, designed to be prepared and regulated. Sampled awareness is divided into distinct layers. The database contains around 5 million association records pertaining to coaching expertise, as well as nearly 2 million association records pertaining to general knowledge. Furthermore, the dataset includes a compilation of forty-one alternatives derived from each relationship, together with a label indicating the status of association records as either a typical or specific form of attack. Various types of continuous, discrete, and symbolic variables possess these choices, encompassing a wide range of values that can be classified into four distinct categories: (1) The main class comprises the inherent choices of an association, which encompass the essential possibilities for connecting individual transmission control protocols. Various options encompass the duration of the association, the protocol type (TCP, UDP, etc.), and network access (HTTP, telnet, etc.). (2) The content options available in an association, as indicated by the Domain Data Area Unit, do not have any influence on the payload of the first transmission control protocol packets. (3) Persistent host options examine established connections that have a consistent target host within the past 2 seconds due to the existing association and collect data on protocol behavior, performance, etc. (4) Similar service options analyze connections that share the same service as the current link within the previous two seconds.

During the pre-processing stage, the second block of Figure 1 illustrates our utilization of a packet sniffer implemented with the winpcap library. This sniffer is employed to collect comprehensive network packet details from each packet, encompassing the IP header, TCP header, UDP header, and ICMP header. Subsequently, the packet data is organized and transformed into a record by consolidating the information, taking into account the connections between each pair of science addresses (source science and destination IP). Each record comprises carefully selected features that serve as indicators of network data and activities, as musical notation features accurately represent the key aspects. To distinguish

significant opportunities that delineate the characteristics of normal vs malicious network traffic, we opt to carry out comprehensive tests. We employ data collection to carefully determine thirty-five crucial selections for our IDS strategy, based on factors such as data kind, available possibilities, and the cost-to-benefit ratio. The price of any data benefit associated with a feature is determined by the relationship between the component and the output group. The knowledge parameters obtained are X and Y. X specifies individual choices such as the range of protocol packets for transmission control and the content of supply ports for transmission control protocol. Y defines category groups that measure conventional knowledge, probe attack, and DoS attack. However, each element of the present is crucial for both the DoS assault and the Probe attack. The data acquisition results indicate that all 35 network data features must be taken into account for intrusion detection and categorization.

## IV. RESULTS AND DISCUSSION

Upon the system's successful implementation, we proceed to compute the confusion matrix. Table 1 and Table 2 display the categorization achieved with the utilization of PCA techniques. Figure 2 illustrates the accuracy of data collection by KDDCUP utilizing the density-based technique of the machine learning algorithm program. Figure 3 is employed to categorize and predict the precision of the proposed system using various ways, including the RNN algorithm.

**Table 1 : Confusion matrix calculation using PCA for classification**

| **Class** | **Normal** | **Attack** |
|---|---|---|
| Normal | 1760 | 19 |
| Attack | 9 | 1640 |
| | 1769 | 1659 |

**Table 2 : Confusion matrix calculation using RF for classification**

| **Class** | Normal | Attack |
|---|---|---|
| Normal | 1830 | 227 |
| Attack | 169 | 1202 |
| | 1999 | 1429 |

**Table 3 : Performance evaluation with PCA and RF**

| | **PCA** | **RF** |
|---|---|---|
| **Accuracy** | 0.9892 | 0.9525 |
| **Precision** | 0.9867 | 0.9797 |
| **Recall** | 0.9933 | 0.9463 |
| **F-Score** | 0.9899 | 0.9529 |

Based on the analysis of both experiments, Principal Component Analysis (PCA) demonstrates superior classification accuracy compared to the Random Forest (RF) technique, as depicted in Figure 3. Based on the analysis of the experiment mentioned above, we can infer that the suggested system yields higher accuracy for trust calculation in the IoT in-service environment. The research incorporates simulation environmental elements and employs a fusion of machine learning methods. Machine learning algorithms have been applied to cluster differentiation and id.mi.com utilizing different computation settings.

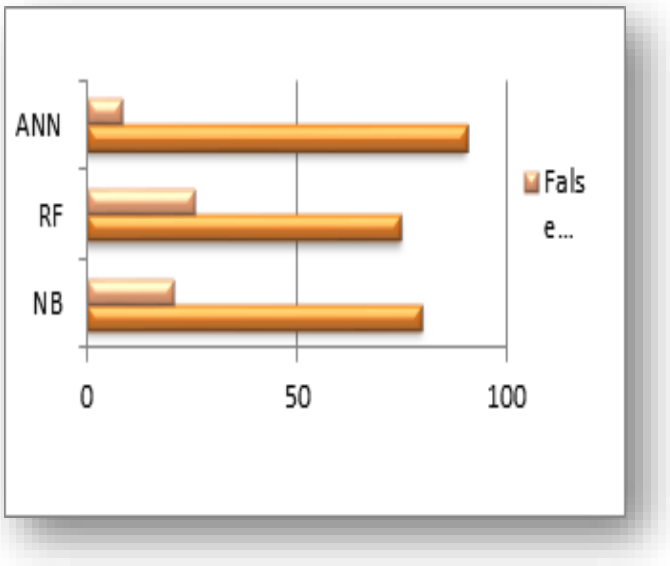


**Figure 2: Detection accuracy for KDD : CUP99 dataset using machine learning**

The diagram depicted in Figure 2 illustrates the accuracy of the classification findings for the kddCup 99 dataset, which consists of five distinct classes. The average software output for the machine learning algorithm is approximately 88.50% across all classes.

### *A.* Proposed Result

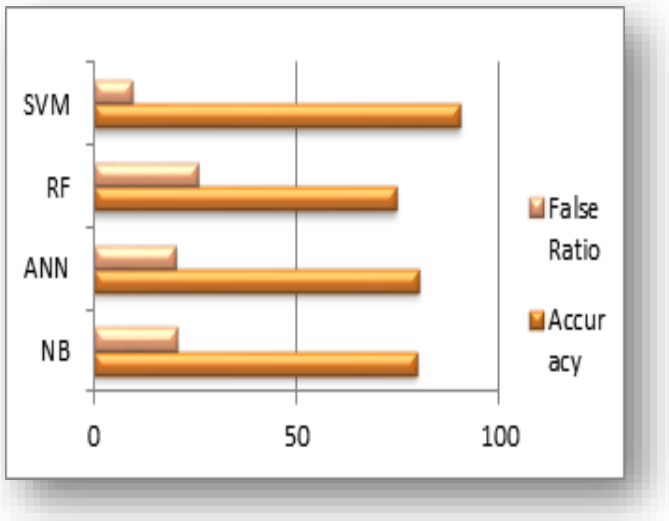


**Figure 3 : Detection accuracy various network dataset using PCA and RF**

The picture above, picture 3, displays the mean accuracy of identification across multiple databases for (n) distinct categories. The system achieves an average performance of approximately 95% across all (n) classes using the machine learning technique.

## 5. CONCLUSION

This work presents a novel intrusion detection system (IDS) approach that utilizes deep learning techniques to recommend a highly effective IDS solution. In order to evaluate the precision of anomaly detection, we employed the NSL-KDD dataset, which consists of synthetic-based intrusion data. Our future strategy involves integrating Intrusion Detection Systems (IDS) into the cloud environment using the deep learning technique. We additionally examine and contrast several techniques of deep learning, specifically. The application utilizes artificial intelligence and conditioning algorithms to identify intrusions in the RF network and perform PCA on the NSL-KDD dataset. Its primary function is to detect unknown cases during the data inspection. The streamlined rule framework results in enhanced categorization and sophisticated identification. Several trials employed experimental analysis to evaluate the algorithm's efficiency through multiple testing and determined that we are attaining suitable outcomes.

## REFERENCES

[1] Bhosale, Karuna S., Maria Nenova, and Georgi Iliev. "Modified Naive Bayes Intrusion Detection System (MRFIDS)." 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS). IEEE, 2018.
[2] Chamou, Dimitra, et al. "Intrusion Detection System Based on Network Traffic Using Deep Neural Networks." 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD). IEEE, 2019.
[3] Mohammed ARFar, Rosni Abdulah, Izan H. Hasbullah, Yung- Wey Chong; Omar E. Elejla, “Comparative Performance Analysis of classification algorithm for Internal Intrusion Detection “, 2016 14th Annual Conference on Privacy Security and Trust (PCT), Dec 12-14,2016, Penang, Malaysia.
[4] Weiwei Chen, Fangang Kong, Feng Mei, GuiginYuan, Bo Li, “a novel unsupervised Anamoly detection Approach for Intrusion Detection System”, 2017 IEEE 3rd International Conference on big data security on cloud, May 16-18,2017, Zhejiang, China.
[5] Anna L. Buczak, Erha n Guven, "A Survey of Data Mining and Machine Learning methods for cybersecurity intrusion detection", IEEE communication surveys and tutorials, vol. 18, Issue 2,2016.
[6] Manoj s. Koli, Manik K. Chavan, “An Advanced method for detection of botnet traffic using Interhnal Intrusion Detection”, 2017 International Conference on (ICICCT), March 10-11, 2017, Sangli, India.
[7] Thabet Kacem, Duminda Wijesekera, Paulo Costa, Alexander Barreto, “An ADS-B Intrusion Detection System”, 2016 IEEE on ISPA, 2016, Fairfax, Virginia.
[8] Shengyi Pan, Thomas Morris, Uttam Adhikari, “Developing a Hybrid Intrusion Detection System using Data Mining for power system”, IEEE Transactions on, vol. 6, issues. 6, Nov. 2015.
[9] Mehdi Ezzarii, Hamid Elghazi, Hassan El Ghazi, Tayeb Sadiki, "Epigenetic Algorithm for performing Intrusion Detection System", 2016 International Conference on ACOSIS, Oct17- 19,2016, Rabat, Morocco.
[10] BOROLE, Prajakta; SHARMA, Yogesh Kumar; NEMADE, Santosh. 6G Network Access and Edge-Assisted Congestion Rule Mechanism using Software-Defined Networking. International Journal of Future Generation Communication and Networking, 2020, 13.1s: 107-112.
[11] Panagiotis I. Radogloa-Grammatikis; Panagiotis G. Sarigannidis, "Flow anamoly based Intrusion Detection System for Android Mobile Devices", 2017 6th International Conference on MOCAST, May 4-6, 2017, Kazani, Greece.
[12] PRADEEP, S.; SHARMA, Dr Yogesh Kumar. Effectual Secured approach for Internet of Things with Fog Computing and Mobile Cloud Architecture Using IFogSim. WE C-2019-London, UK, DOI, 2019, 978-988.
[13] Jaevier A. Villanueva, Luisito L. Lacatan, Albert A. Vinluan, Information Technology Security Infrastructure Malware Detector System, International Journal of Advanced Trends in Computer Science and Engineering, pp. 1583-1587 ,Volume 9, No.2, 2020.

[14] Bholanath Mukhopadhyay , Dr. Rajesh Bose, Dr. Sandip Roy, A Novel Approach to Load Balancing and Cloud Computing Security using SSL in IaaS Environment, International Journal of Advanced Trends in Computer Science and Engineering, pp. 2130-2137,Volume 9, No.2, 2020.
[15] Arthur, Menaka Pushpa. "Detecting Signal Spoofing and Jamming Attacks in UAV Networks using a Lightweight IDS." 2019 International Conference on Computer, Information, and Telecommunication Systems (CITS). IEEE, 2019.
[16] Jaiswal, Shradha, and Dinesh Kumar Vishwakarma. "State-of-the-Arts Person Re-Identification Using Deep Learning." 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2019.
[17] Atefi, Kayvan, Habibah Hashim, and Touraj Khodadadi. "A Hybrid Anomaly Classification with Deep Learning (DL) and Binary Algorithms (BA) as Optimizer in the Intrusion Detection System (IDS)." 2020 16th IEEE International Colloquium on Signal Processing & Its Applications (CSPA). IEEE, 2020.
[18] Chockwanich, Navaporn, and Vasaka Visoottiviseth. "Intrusion Detection by Deep Learning with TensorFlow." 2019 21st International Conference on Advanced Communication Technology (ICACT). IEEE, 2019.
[19] Rashid, Azam, Muhammad Jawaid Siddique, and Shahid Munir Ahmed. "Machine and Deep Learning Based Comparative Analysis Using Hybrid Approaches for Intrusion Detection System." 2020 3rd International Conference on Advancements in Computational Sciences (ICACS). IEEE, 2020.
[20] Vijayanand, Radhakrishnan, D. Devaraj, and B. Kannapiran. "A novel deep learning-based intrusion detection system for smart meter communication network." 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization, and Signal Processing (INCOS). IEEE, 2019.
[21] Otomo, Safa, Burak Kantarci, and Hussein T. Mouftah. "On the feasibility of deep learning in sensor network intrusion detection." IEEE Networking Letters 1.2 (2019): 68-71.
[22] Vinayakumar, R., et al. "Deep learning approach for the intelligent intrusion detection system." IEEE Access 7 (2019): 41525-41550.
[23] Sheu, Ruey-Kai, et al. "IDS-DLA: Sheet Metal Part Identification System for Process Automation Using Deep Learning Algorithms." IEEE Access 8 (2020): 127329-127342.
[24] Abou Khamis, Rana, M. Omair Shafiq, and Ashraf Matrawy. "Investigating Resistance of Deep Learning-based IDS against Adversaries using min-max Optimization." ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020.
[25] Panagiotis I. Radogloa-Grammatikis; Panagiotis G. Sarigannidis, "Flow anamoly based Intrusion Detection System for Android Mobile Devices", 2017 6th International Conference on MOCAST, May 4-6, 2017, Kazani, Greece.
[26] Trae Hurley, Jorge E. Perdomo, Alexander Perez-pons, "HMMBased Intrusion Detection System for software-defined networking", 2016 15th IEEE Conference on Machine Learning and Application, Dec 18-20, 2016, Miami, Florida.
[27] Mariem Belhor, Farah Jemili, "Intrusion Detection based on genetic fuzzy classification system", 2016 IEEE 13th International Conference on Computer Systems and Application (AICCSA), Nov 29 2016-Dec 2, 2016, Sousse, Tunisia.
[28] Sharad Awatade, Shweta Joshi. "Improved EAACK: Develop Secure Intrusion Detection System for MANETS using hybrid cryptography", 2016 International Conference on computing communication control and automation (ICCUBEA), Aug 12-13, 2016, Maharashtra, India.
[29] Mayank Agarwal, Sanketh Purwar, Santosh Biswas, Sukumar Nandi, "Internal Detection System for PS-Poll DOS attack in 802.11 networks using real-time discrete event system",IEEE,vol.4,issue4,2017.
[30] Md Zahangir Alom, Tarek m. Taha, "Network Intrusion Detection for cybersecurity on neuromorphic computing system", 2017 International Joint Conference on Neural Networks (IJCNN), May 14-15,2017, USA.