IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

A Comprehensive Examination of Literature Exploring the Implementation of Machine Learning to Network Security's Intrusion Detection Systems

Anjali Pandathara
Msc. Computer Science
Mithibai College of Arts, Chauhan Institute of Science and
Amrutben Jivanlal College of Commerce and Economics, Mumbai, Maharashtra, India
anjaliradhakrishnant2001@gmail.com

Abstract: *The Internet and telecommunication technologies have developed quickly, the amount of data transferred has greatly increased. Attackers are continually devising new tactics to steal or modify these data because they are so highly desired. The threat these attacks pose to the security of our systems is growing. It is among the most tough issues to resolve for detection of intrusions. An idss is a programme that attempts to analyse network traffic in order to detect intrusions. Despite the fact that many researchers have examined and developed novel IDS systems, IDS even now must be enhanced in order to achieve satisfactory detection capability while reducing number of false alarms. Furthermore, numerous intrusion detection systems have difficulty detecting nil attacks. Machine learning techniques had also recently become popular among scholars as a quick and accurate method of detecting network infiltration. This article offers a taxonomy of machine learning approaches as well as an explanation of IDS. In addition to a list of current IDS that include machine learning and a discussion of the essential components for IDS analysis, this article also outlines the advantages and disadvantages of each machine learning approach. The veracity of the findings from the evaluated study is then discussed after specifics of the various datasets used in the studies are given. The preceding part looks at the results, study obstacles, and projected future trends.*

Keywords: Machine learning; data protection; systems for identifying and avoiding intrusions

I. INTRODUCTION

Computer Security Threat Monitoring and Surveillance, was released in 1980. which pioneered intrusion detection systems are indeed a theory.James worked for the US Air Force on the Defense Research Board Task Force on Internet Security. James explained Ways audit analysis could be utilised to find unauthorised or malicious behaviour is discussed in his work. For instance, examining a file record can reveal crucial details that help identify unusual usage. Premised on the outcomes of this research,was created HIDS.Shortly after, the very first intrusion detection system (IDS) has been established, which can identify threats from a list of known attacks.

Several systems administrators began utilising intrusion detection systems at the end of the 1980s. Although these continuously scanned the web, their are resource-intensive and were unable to identify zero-day attacks .

During the 1990s, in response to a rising number of attacks, a new detection technique called anomaly detection was developed. This approach involved identifying abnormal behavior or activity in a system and alerting administrators accordingly. However, the variability of networks during the 1990s and 2000s resulted in a significant number of false reports, causing administrators to lose confidence in intrusion detection systems (IDS) and ultimately abandon their use due to their lack of reliability.

Fortunately, as network technology has advanced, intrusion monitoring systems have been integrated into system security. and component processing capacity, as well as the emergence of machine learning. These improvements have enabled IDS to better detect and prevent unauthorized access to networks. Even though the potential of deep learning algorithms in this field has not yet been completely investigated, machine learning techniques have been thoroughly

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

assessed for use in IDS development. Deep learning may make it feasible to intensify the IDS's precision and reduction the frequency of fake warnings, providing new opportunities for boosting system security.

The key purpose of this article is to presentdepth and evident overview ofbasic agglomerates machine learning as well as systems for the detection of intrusions.The goal of is the article is to demonstrate most recent developments and findings regarding the utilization of machine learning through IDS. The article's objective is to evaluate the advantages and disadvantages of various machine learning paradigms by looking at various datasets, evaluation metrics, and machine learning techniques. The article also reviews the body of research in the area. The paper also finds applications for innovative machine learning approaches to improve IDS.In its final section, the paper outlines forthcoming advancements and future study needs in this field.

II. INTRUSION DETECTION SYSTEM

Any time an IDS detects harmful network activity, it will sound a warning. It can be either a hardware or software device. IDSs serve as the network's "watch-eye." It is very important in the authentication scheme of modern networks. By enabling the early detection of threats, it provides the chance to counteract them. Additionally, they make a variety of dangers, such as Man in the Middle (MitM) as well as Denial of Service, detectable (DoS). IDSs have the ability to track and record any and all network activity as needed.

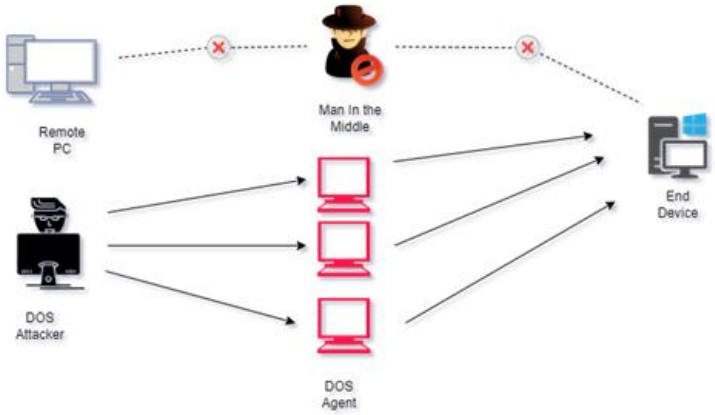

Figure 1. Network Attack.

Network administrators may better comprehend what transpired because IDSs may provide information on assaults as they happen.

- IDSs positioned strategically throughout the network are called network intrusion detection systems. To determine whether the network is being used for nefarious purposes, NIDS examines the network's overall traffic. NIDS is a vital part of the security of the majority of business networks and aids in the detection of assaults from your own hosts.
- On every client device connected to the network, IDSs, or host intrusion detection systems, are installed (hosts). HIDS, as opposed to NIDS, examines the traffic and activity of a particular host and notifies the user if it observes any unusual behaviour.

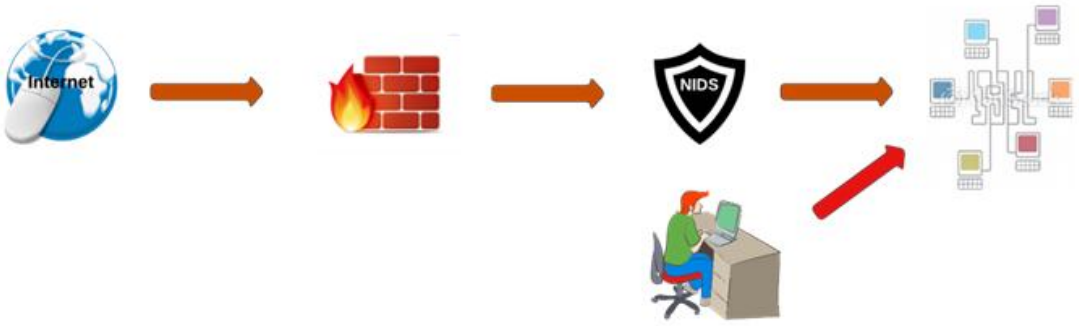

Figure 2. Network Intrusion Detection System

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

IDS has three distinct detection methods

- Misuse detection involves looking for recognisable intrusion patterns on a host or channel, often known as signature-based detection. Thus, every attack has an unique signature, which could be a header, originating IP address, or data of the protocol. The objective is to locate these signatures and utilise them to recognise and stop upcoming assaults. An alarm could be set off if the IDS detects an attack with a signature from its collection of known signatures. This tactic's advantage is its ability to spot common threats. Its effectiveness against unexpected or zero-day attack patterns, which are attack trends that have never existed before, is a drawback.
- Anomaly detection establishes a baseline for the usual condition of the host or network, and any divergence from this norm is regarded as an attack. For example, may create a baseline based on typical internet activity, such as the programmes provided and used by each host, as well as the activity level every day. Consequently, if such a hacker tries to enter an infrastructure at midnight, the IDS would then trigger an alarm, when there should be very little activity according to the baseline. The usefulness using anomaly detection is that it makes it simple to find intrusion risks that had previously gone undetected. Yet, the deceptive detection performance of these methods can be strong due to the fact that it is typically challenging to exactly define what a network's background is.
- The two previous observations are combined using hybrid detection. In theory, it could detect novel attacks and have a lower percentage of incorrect identification than anomaly methods.

III. PERFORMANCE COMPARISON OF IDS

We utilise efficiency metrics to assess IDSs that use machine learning methods. Usinga collection of test suite for which the actual values have been identified, a classification model (also referred to as a "classifier") performs is described in a table called a confusion matrix. The machine learning model's predictions are put up against the real target values in a matrix.

Following phrases employed to characterize a confusion matrix:

- True Positive: The proper classification of an attack outcome is attacking result.
- True Negative: The accurate identification of a typical sample as typical traffic was made.
- False Positive: A normal sample was mistakenly categorised as an assault.
- False Negative: A data from an incident was already incorrectly categorized as local traffic.

In order to prevent anomalies in the network, which cause network disruptions, IDS ought to have a high detection accuracy. Additionally, a lower false negative incidence is required to avoid undiscovered threats from accessing the network. We can review the numerous measures that are typically used to gauge an IDS's effectiveness using the abovementioned terminology and matrices from Table 1.

Prospective class

		Common	Offensive
True class	Common	TN	FP
	Offensive	FN	TP

Table 1: IDS performance evaluation metrics.

IV. MACHINE LEARNING

ML is a technology that is strongly associated with Artificial Intelligence (AI). It teaches an algorithm that searches a dataset for regular patterns. This learning produces a model which may be utilized to estimate or imbrute tasks. Machine learning may be employed by intrusion detection systems to detect unfamiliar or suspected threats if the model is properly trained. Figure 3 depicts three distinct classifications of techniques for machine learning: supervised, unsupervised, and semi-supervised.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

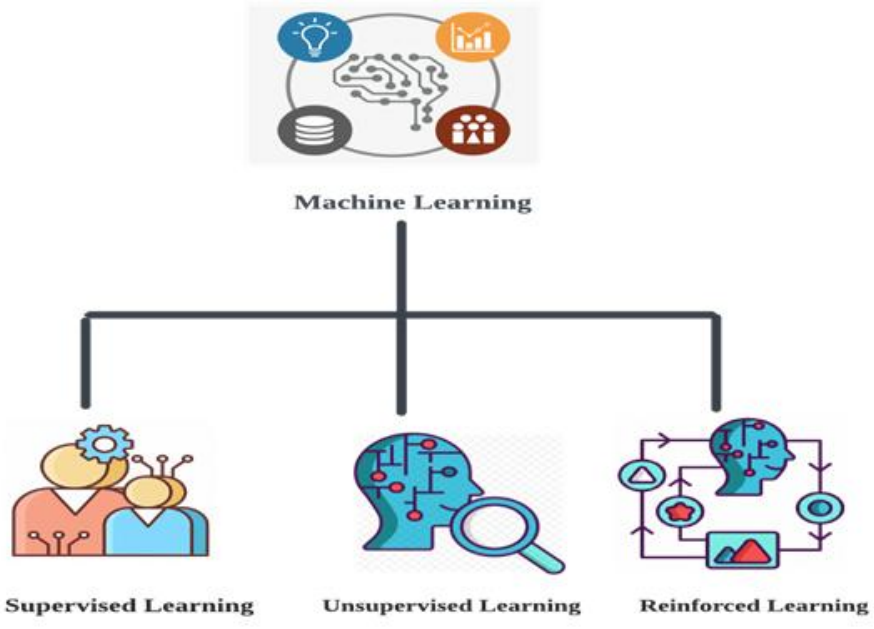

Figure 3. The various kinds of techniques for machine learning.

ML Approach	Algorithm utilised	Benefits
Supervised ML	Linear, Logistic, Random Forest, and SVM	It makes use of a learned or labelled dataset, which enables it to manage massive quantities of data. It excels in predictive scenarios and situations where accurate predicition are required
Unsupervised ML	K-means, Apriori, PCA	It equipped to work with unlabeled data and performs optimally in analytical environments
Reinforced Learning	Q-Learning, DL	This aids inside the resolution of complicated real-world issues that can be challenging to fix using overall methods.

Table 2. Tabulates the benefits and drawbacks of ML methods.

V. LITERATURE REVIEW

Throughout their study, A neural network built on convolution (CNN) with such a multi-layer perceptron was described by Lirim et al. as a potential approach to intruder detection. A totallylinked network, in every other neuron represents a layer in addition to therefore be interconnected into every other neuron throughout the following layers, is how the multi-layer perceptron can indeed be conceptualised. Pooling layer is used to represent one of the hidden levels of CNNs as opposed to a multiplication matrix in a traditional neural network. A tensor with various parameters, including input height, width, channels, and amount of inputs, serves as the input to a CNN. The information is concatenated by the convolutional layer, which further sends the output towards the following layer. Moreover, but after many convolutions, the tensor length may grow to an insurmountable size. Lirim et al. utilised space to condense the tensor's dimensionality in order to resolve this. Hyperparameter optimization was employed in order to train the model till the efficiency lowered. Ten classes comprised there own final model, which include nine for assaults as well as one for regular traffic. Numerous dual convolutional layers made the model stronger, which was then followed by pooling and dropout layers to preclude oversizing. The bootstrapping technique was employed in order to address the model's class imbalance seen between higher and lower classes. A consumer dataset that taken into account for 30% of the total dataset and the pre-partitioned UNSWNB15 dataset had been utilized by the authors to evaluate their model. With both databases, they attained accuracy rates of 94.4% and 95.6%, accordingly.

A CNN-based IDS with offline training and online recognition was proposed by Lin et al. in there own study (see Lin et al., 2018). Though the neural network as well as a maxpooling layer, those who was using a Cnn architecture in the unsupervised pre - training method to decrease this same input layer length from 9x9 to 1x1. However during internet

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

identification process, those who post the packets using the open-source IDS Suricata, subsequently employing the trained model to identify threats.The researchers employed the feature dataset as well as the unfiltered traffic dataset from the CICIDS2017 dataset, individuals evaluated there own model. Their model's correctness for both the feature dataset was 96.55%, as well as the exactness for the raw traffic dataset was 99.56%, showing that perhaps the algorithm performed better with the raw traffic compared to the collected characteristics.

Rohit et al. postulated an group-based methodology used intrusion detection throughout their investigation. Their approach must have been broken down into three phases. They began besides trying to normalize the KDD Cup99 dataset, after which features extracted using just a correlation method to information gain as that of the intends to take. Eventually, they consolidated optimization techniques: Nave Bayes, PART, and Adaptive Boost to generate an optimization technique. The conclusions of the each automated system were especially in comparison, as well as the typical or even the most popular findings was employed to determine the final choice. They utilized the bagging technique to decrease the variance error. Just on KDD Cup99 dataset, their strategy produced a precision of 99.9732%, proving the successfulness of their potential methodology.

Al-Yaseen et al. merge SVM as well as ELM in their innovative architecture for an intrusion detection system. So every part of there own five-level approach has been aimed at recognizing numerous categories of network communication. Whereas the second level isolates Probe traffic from Some of the other, the very first level differentiates between DoS as Well as other traffic. User to Root (U2R) assaults are differentiated from many other threats now at third level, as well as Remote to Local (R2L) assaults are distinguished from all other breaches somewhere at fourth level.The fifth stage ultimately differentiates among standard and prohibited activity. At tiers 1, 3, 4, and 5, Classification models are being used, whereas at level 2, an ELM classifier is being used since it has exhibited superior performance to SVM for distinguishing Probe activity. To establish the five categories which an one‘s approach could indeed distinguish, this same training set again from KDD dataset has been pre-processed, as well as an updated K-means methodology was employed to perform extraction of features. One‘s model's correctness was 95.75 %, just topping out the multi-level SVM approach's accuracy of 95.57 percent. Likewise, contrasted towards the multi-level SVM approach's fake threat rate of 2.17%, there own hybrid version used to have a reduced rate of false alarms of 1.87%. Such outcomes demonstrate the efficacy of one‘s suggested strategy for intrusion prevention is.

Yiping et al. established a technique for detecting intruders on wireless networks depending on the random forest algorithm. There own strategy entails establishing an algorithm for identifying hostile nonlinear encrypting incursion transmissions after establishing an algorithm for identifying channel estimation to catch significant signal characteristics. They collected a spectral characteristics of something like the fraudulent transmission to use an enhanced random forest algorithm, as well as subsequently employed a reinforcement learning technique as well as static characteristic synthesis to perform the highest suitable identification of fraudulent communication in a wirelessly. There own program's average precision of 96.93% showed how successful one‘s technique has been at detecting intrusions into wireless networks.

An outlier detection-based technique was developed by Jabez et al. to acknowledge unexplained assaults. They was using the neighborhood anomaly component to find pieces of data which are segregated from clustered observations. Also on KDDcup99 datasets, they examined one‘s algorithm, as well as they established it to be susceptible of identifying unexplained assaults. Their method offers a variety of benefits, including its rapid efficient implementation, something that surpasses alternatives like the backpropagation neural network, where it requires a considerable amount of computing power.

Gu et al.'s enhanced IDS proposal, predicated upon this SVM classifier, incorporated feature extraction by using Naive Bayes technique. They examined one‘s technique on two distinct sets of data, UNSW-NB15 and CICIDS2017, and obtained improved result than what they could possess using only a single SVM classifier. Their approach had accuracy scores of 98.92% on CICIDS2017 as well as 93.75% on UNSW-NB15. Their strategy, nevertheless, has been constrained to detecting malicious and therefore can distinguish between various kinds of assaults.

As a intrusion detection through wireless networks, Pan et al. presented a cloud-based approach. They utilized sink accordingly based inside the cloud to enhance system performance. To reduce the data density as well as computational burden, they was using a mixture of Polymorphic Mutation (PM) as well as Compact SCA (CSCA). For the optimal configuration, they adjusted a KNN method using PMCSCAThey generally used the UNSW-NB15 but also NSL-KDD

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

datasets to test their technique, and they obtained values of 99.327% and 98.27%, accordingly. Their tactic is successful and portable, making it appropriate for mobile networks with minimal resources.

Xiao et al. generated a device for detecting intrusions solely on CNN. After obtaining the characteristics utilising Principal Component Analysis as well as Auto-Encoder, they converted the dataset dimension into a two-dimensional matrix then use back propagation techniques to retrain the Cnn architecture. Also on KDDcup99 dataset, its model was assessed so it surpassed DNN as well as RNN models with such a precision of 94%. U2R as well as R2L attacks, that were underrepresented inside the dataset, have been inadequately detected by that of the algorithm.

A innovative multi-layer algorithm for intrusion detection which incorporates CNN as well as GcForest has been proposed by Zhang et al. To determine various types of assaults and legitimate traffic from the input information, they was using an enhanced CNN optimization technique GoogLeNetNP during the initial layer. To add the most assault subtypes and increase the solution's precision, their model's second level employs GcForest, a random forest approach that creates a flow configuration of decision trees.As in GcForest layer, they created trees sequentially utilising XGBoost rather than creating a random forest till the goal feature was streamlined. The correctness of their approach, which has been assessed utilising UNSW-NB15 as well as CICIDS2017 datasets around each other, seems to have been 99.24% total, which in itself is higher than that of the correctness of the techniques used independently.

A Few-Shot Learning (FSL)-based IDS concept with CNN as well as DNN integration algorithms for extracting features has been presented by Yu et al. Their approach intended to retain this same important data whereas understanding from either a modest amount of information. They assessed their program here on UNSW-NB15 as well as NSL-KDD datasets, and indeed the results indicated estimation with 92.34% and 92% correctness, accordingly.

Gao et al.introduced a combination-based detection system for intrusions which extrapolates utilising Principal Component Analysis. Here on NSL-KDD dataset, they conducted a trial count, as well as they merged the to create the same composite technique, , Random Forest, Decision Tree, K - nearest neighbors, DNN, but also MultiTree techniques were used. Throughout a bid to improve precision, distinct weight values have been given to each program when generating a ensemble model's outcome. Their ensemble method outperformed using such a single algorithm by itself with an effectiveness of 85.2%. There own technique could prove to be less successful in spotting assaults which appear occasionally, though.

Deep Belief Networks (DBN) and a group of Support Vector Machines have been merged in Marir et al.'s innovative intrusion detection system (IDS) plan. Unsupervised networks like Restricted Boltzmann Machines help compensate the unsupervised networks together in neural network is known a DBN . To understand its important aspects for DBN, They firstly used an unsupervised post methodology based on a layout that could be described as greedy. Following it which, a group of SVMs classifies its features extracted, as well as a selection technique yields an outcomes. On four datasets— CICIDS2017,NSL-KDD, KDDcup99, and UNSW-NB15—he evaluated each person's technique, and they found that accuracy and reliability were, respectively, 94.76%, 97.27%, 90.47%, and 90.40%. They did acknowledge, though, moreover adding so much DBN layers makes their approach so much time-consuming.

Wei et al. posited an optimal solution strategy to improve DBN effectiveness in IDS. They utilised Genetic Algorithm, Artificial Fish Swam Algorithm, and Particle Swarm Optimization in combination. AFSA was used to improve PSO to identify the most effective component search. GA was then employed to determine the ideal response globally for the early search for particles, which was then used to enhance the precision of the DBN prototype. On NSL-KDD dataset, the suggested solution was tested, as well as a correctness of 82.36% was obtained.

A adaptable IDS modeled here on Deep Neural Network (DNN) model, that also includes of such an input nodes, five hidden nodes, as well as an output neurons, has been developed by Vinayakumar et al. In accordance with the requirements, the amount of concealed layers could be modified from one to five. Their solution is adaptable to all these HIDS as well as NIDS, but instead they initiated utilize the Apache Spark computing environment. Contingent on the amount of DNN layers were being used, on a variety of datasets, including KDDcup99, NSL-KDD, CICIDS2017, UNSW-NB15, and Kyotohe examined his strategy for NIDS, as well as procured a best average of 93%, 79.42%, 87.78%, 76.48%, and 94.5%, successively.

In their innovative approach to IDS, Shone et al. combined Random Forest as well as Non-symmetric Deep Auto-Encoder.Unlike a traditional auto-encoder, which uses a symmetric encoder-decoder scheme, one‘s approach also uses the encoding process, reducing calculation time whereas keeping IDS precision. Individuals use 2 NDAE, that each

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

have three hidden layers in order to execute extraction of features before adding Random Forest as one's classifier more to improve accuracy rate. This allows them to handle complicated datasets. Participants put one's approach towards the evaluation against a DBN solution through using KDDcup99 as well as NSL-KDD databases. Even though there own proposed alternative does have trouble picking up smaller class sizes like R2L and U2R, it still succeeded in achieving accuracy level of 97.85% but also 85.42%, correspondingly.

In order to increase IDS efficiency, Yan et al.showed how Stacked Sparse Auto-Encoder can be used for extracting features. A sparsity penalty is used by SSAE, a type of auto-encoder, to lower the dimensionality and amount of hidden nodes in the training dataset. Those that tried one's SSAE just on NSL-KDD dataset utilising different classifiers and optimised it using error backpropagation. Their findings demonstrated that the highest overall performance of 99.35% has been attained while SSAE as well as SVM classifier were mixed. Their solution's considerably shorter training and testing periods than competing products are one of its main benefits. Their method, though, has trouble detecting R2L and U2R classes.

A two-stage deep learning model (TSDL) was implied from Khan et al. to optimise IDS. The structure consists of two levels, the first of which assigns a significance level to the traffic as well as categorises this as either normal or abnormal. The said probability value serves as an optional accessory to teach the classifier with in second phase. The Deep Stacked Auto-Encoder (DSAE) was used by the authors as a feature extractor and Soft-max as a classifier in a Deep Neural Network (DNN) approach with both phases. For multi-class categorization issues, neural networks frequently use soft-max. They was using the KDDcup99 as well as UNSW-NB15 datasets to test one's algorithm, and also the results showed that it was 99.996% and 89.134% effective altogether, correspondingly.

Have used two auto-encoders along with a one-dimensional CNN, Andresini et al. put forward an approach which combines two distinct approaches. The auto-encoders have been trained and evaluated using both typical and attack packets, as well as the recreated test results have been introduced to the training sample. Just the one CNN is employed to distinguish between typical as well as attack traffic, and just a Soft-max classification model is employed to determine the database's category. Here on KDDcup99, CICIDS2017 & UNSW-NB15 datasets, the solution make relevant levels of accuracy of 92.49%, 93.40%, but also 97.90%, respectively. Their solution, even so, fails to offer details on the kinds of threats identified.

Ali et al. proposed a novel method called "PSO optimized FLN " to boost the detection precision of intrusions Systems. They was using PSO to optimise the neural network strength training, which may be ineffectual in FLN. The planned algorithm was evaluated against the other FLN alternatives somewhat on KDDcup99 dataset. They outperformed some other alternatives in detecting this same various classes, accomplishing an average precision of 89.23%. Nevertheless, their highest accuracy was hampered by their incapability to recognise one of the few attack classifications (R2L).

To improve intrusion detection, Dong et al. posited a combination of hybrid cluster analysis as well as SVM. They was using K-means tend to cluster to split the information into subgroups, followed by SVM to categorise each subset. Mostly on NSL-KDD dataset, one's technique produced a total accurateness of 99.45%, which really is greater than some other methods. Their method also lowered computation time in comparison to SVM algorithms with various parameters. The authors, even so, did not provide information on the exactness of detecting so every kind of assault.

Table 3: Overview of the Literature Study with Pros and Cons of the IDS Model.

Researcher	Pros	Cons
Jabez et al.	The proposed approach for IDS is an outlier detection method that seeks to find isolated data points using the neighbourhood outlier factor. This approach outperforms the backpropagation neural network in terms of execution time.	The study utilized the outdated KDDcup99 dataset, and the authors did not provide any information regarding the accuracy of detecting different attack types.
Al-Yaseen et al.	The model has five tiers, each of which is in charge of identifying a specific type of attack. A modified K-means technique accustomed to identify features. Comparing this proposed model to multi-level SVM as well as ELM, a superior overall accuracy was attained.	The methodology was trained on an old dataset. Furthermore, the R2L as well as U2R attack precision of the model was very inadequate.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

Rohit et al.	The solution that is being presented combines three methodologies Bayes, PART, and Adaptive Boost in a composite technique. The final decision is made by averaging the results obtained from each algorithm. Feature extraction methods are also employed in the solution.	The dataset utilised, KDDcup99, is out-of-date, as well as the research did not have any details on the model's effectiveness for various types of attacks.
Marir et al.	A multi-layer group SVM method is employed in the definitive technique to identify unusual activity. The group SVM received the characteristics produced by the extraction of features step of the DBN technique.A voting algorithm was used to decide the final outcome. The solution was tested on multiple datasets, including old and recent onessuch as CICIDS2017 NSL-KDD, KDDcup99 &UNSW-NB15.	Increasing the quantity of layers in the model leads to increased computational time. The study did not provide sufficient information about the accuracy of identifying different attacks.
Shone et al.	A random forest algo was combined with a Non-symmetric Deep Encoder to detect unusual behaviour. For extracting features, the writers was using two NDAE algorithms that those who combined. Their program had a high accuracy rating when contrasted to a DBN approach.	The authors' suggested method has trouble correctly identifying small attack classes like R2L with U2R an old databases.
Yan et al.	The designed IDS uses SVM as the classifier as well as SSAE as the feature extractor. Compared to other solutions, their method significantly reduces the length of time required for evaluation and teaching.	The suggested NIDS has a poor recognition accuracy for U2R but also R2L attacks because it uses the outdated dataset NSL-KDD. Moreover, the model's that use SSAE but also SVM significantly reduces instruction and training time.
Ali et al.	To identify abnormal traffic in their analysis, the scholars used the Fast Learning Network as well as particle swarm optimization.	When recognizing among the small classes, the model's accuracy is poor, so it was examined on the outdated KDDcup99 dataset.
Xiao et al.	For IDS, a Cnn architecture is used PCA as well as Auto-Encoder are used to retrieve features.	The model's performance was evaluated on the old KDDcup99 dataset and revealed a lower than U2R with R2L threat detection.
Zhang et al.	The random forest method as well as CNN were combined either by experts to create a multi-layer model. Their approach, once evaluated on such a mixture of the two most current databases, UNSW-NB15 as well as CICIDS2017, obtained an outstanding correctness of 99.24%. They also provided detailed accuracy scores for each attack class, highlighting the effectiveness of their approach.	The method for extracting features utilized by the algorithm was not discussed in the article.
Gao et al.	Decision Tree, Random Forest, DNN, & KNN as well as MultiTree are all constituents about suggested IDS, an integrated machine learning algorithm. Weights are used by the algorithm to increase voting precision. PCA is indeed the method	Once studying assaults which aren't widespread, the model's effectiveness is restricted. The model's effectiveness is also built just on classic NSL-KDD dataset.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

	for extracting features employed by the algorithm.	
Wei et al.	The use of an optimising method enhances DBN for IDS efficiency. In the approach for improving performance, the Artificial Fish Swarm, Particle Swarm, and Genetic Algorithm are combined the DBN design.	The NSL-KDD dataset, that is quite old, was utilized to evaluate the algorithm. However, the method of feature extraction used was not specified.
Vinayakumar et al.	IDS hybrid which employs DNN. To evaluate their model, they used also recent and old datasets.	The model lacks a feature extraction method, and For weaker assault groups, its complication leads to a low prediction performance. It was however tried on both fresh and old datasets.
Khan et al.	The DNN method is used by the two-stage TSDL model to identify abnormal traffic. The model categorises the traffic in the initial step with a specific probability. The method makes use of this likelihood as an optional bonus in the second step to enhance the classification outcomes.	The TSDL model was analysed on both the old KDDcup99 and recent UNSW-NB15 datasets. While the model achieved an impressive 99.996% accuracy on the old dataset, there was a significant 10% accuracy gap when compared to the recent dataset. Therefore, further improvements are needed to reduce this gap and make the model more robust.
Dong et al.	The hybrid approach that has been recommended uses K-means clustering to SVM to detect anomalous traffic. The data is divided into many subgroups using K-means, and SVM is run across all of those groups. When contrasted to SVM methods utilising different parameters, the method performed faster as well as is quicker to process.	The method was evaluated on an outdated collection, and thus no data on the effectiveness of the each approach classification is provided.
Lin et al.	The solution proposes a NIDS using CNN for detecting abnormal traffic and achieves excellent results on the CICIDS2017 dataset.	The solution lacks information on the accuracy of each type of attack and the techniques used to identify features.
Yu et al.	The IDS employs DNN as well as CNN for extracting features and therefore is based on Few-Shot Acquisition. Even just a small part of the datasets were used by the algorithm, which still produced excellent precision findings. Both the old collection NSL-KDD as well as the more current dataset UNSW-NB15 were used to test the solution. However, no information was given regarding the precision of every assault classification, and it is not obvious if any extracted features or selection techniques were applied.	While the U2R and R2L detection rates are there remains a lot of space for development in these areas, though they are comparatively better than some other approaches.
Andresini et al.	Two auto-encoders are used by the IDS to features extracted, as well as soft-max classifier is used towards identify breach. This algorithm was assessed with both an old (KDDcup99) as well as new dataset (UNSW-NB15).	The effectiveness of identifying so every distinct type of assault was not disclosed in the approach.
Gu et al.	On the existing datasets CICIDS2017 and UNSW-NB15, an SVM model is employed in conjunction with Naive Bayes for selecting features, yielding	The sort of attack is frequently unknown due to their technique.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

	excellent accuracy findings.	
Pan et al.	The approach is indeed a KNN algorithm that maximizes with PM-CSCA and gets good precision on both the old NSL-KDD dataset , new UNSW-NB15 dataset. It really is made for wifi communication and makes use of cloud technology to cut down on the amount of effort as well as processing power needed by the technique.	The solution fails to offer data on the precision of various threat classes.
Yiping et al.	This method employs a random forest structure designed specifically to a wireless connection.	A precision discovered is 96.93% on average. Nothing used validated datasets.

VI. PERFORMANCE METRICS

The datasets used by the various articles analysed in Section 5 are comprehensively summarised in Table 4, along with the performance metrics applied in Section 5 to assess the effectiveness of the suggested solutions. The table provides pertinent information about each dataset's origin and size as well as the precise metrics employed to assess the algorithms and models described in the literature review's accuracy, precision and other performance indicators.

Table 4: Performance metrics

Authors	**Performance metrics**						
	Authenticity	Distinctness	Discovery rate	F1 Score	False detection	ROC analysis	Other
Jabez et al.	-	-	-	-	-	-	Yes
Al-Yaseen et al.	Yes	-	Yes	-	Yes	Yes	-
Rohit et al.	Yes	Yes	Yes	-	-	-	-
Marir et al.	Yes	Yes	Yes	Yes		Yes	Yes
Shone et al.	Yes	Yes	Yes	Yes	Yes	-	Yes
Yan et al.	Yes	-	Yes	-	Yes	-	Yes
Ali et al.	Yes	-	Yes	-	Yes	-	
Xiao et al.	Yes	-	Yes	-	Yes	-	Yes
Zhang et al.	Yes	-	Yes	-	Yes	-	
Gao et al.	Yes	Yes	Yes	Yes	Yes	-	Yes
Wei et al.	Yes	-	Yes		Yes	Yes	-
Vinayakumar et al.	Yes	Yes	Yes	Yes		-	-
Khan et al.	Yes	Yes	Yes	Yes	Yes	-	-
Dong et al.	Yes	-	Yes	-	Yes	-	-
Lin et al.	Yes	-	-	-	-	-	-
Yu et al.	Yes	Yes	-	Yes	Yes	-	-
Andresini et al	Yes	-	-	Yes	-	-	-
Gu et al.	Yes	-	Yes	-	Yes	-	-
Pan et al.	Yes	-	Yes	-	Yes	Yes	-
Yiping et al.	Yes	-	-	-	-	Yes	-

VII. STUDY LIMITATIONS

7.1 Inadequacy of Recent Dataset

The research assessed a critical issue related to current datasets' inadequacy in representing new and emerging cyber-attacks. As a result, due to a lack of diverse attack samples for model training, the majority of the analysed solutions failed to detect zero-day attacks. A comprehensive and diverse dataset containing both old and new attacks is required for such an efficacious Intrusion Detection System (IDS), letting the IDS to detect intrusions, to learn and recognise the

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

critical features of each attack type. As a result, researchers must obtain and maintain an up-to-date dataset with a sufficient amount of evidence representing all kinds of computer attacks, and it must be regularly refreshed with fresh intrusion vials from different scenarios.

7.2 Lower Precision for Smaller Classes

The study also revealed that most proposed methods are ineffectual in detecting minor classes, despite having a high overall accuracy in detecting unusual behavior. The issue of imbalanced datasets in detecting minor classes of abnormal behaviour, resulting in lower accuracy for minor classes compared to major classes. Although using a current dataset with sufficient occurrences of small classes is a potential option, such a dataset is not yet available. To split the dataset into major and minor classes and improve accuracy in detecting minor classes, feature extraction methods and a multi-level framework using various methods of machine learning may be employed.

7.3 Inefficient Performance in Real-World Scenarios

Among the biggest challenges that IDS confronts is a lack of test results in real - life conditions. Whereas many alternatives have indeed been investigated, the majority of them have been evaluated using outdated datasets that don't really represent modern network traffic. Moreover, neither of these strategies has been tested with actual statistics, therefore it's unknown the way they might perform in practical uses. To address this issue, prospective remedies must be examined in practical environments to guarantee their potency.

7.4 Resource-Intensive Models

As demonstrated in Section 6 of this research, often these IDSs seem to be incredibly complicated and necessarily require a substantial portion of computational and storage time as well as resources. The said necessity could have a significant impact on the effectiveness of IDS in a real-world setting. utilizing multi-core GPUs can assist in decreasing the amount of time required, but this solution is expensive. As a result, the developed algorithms must use extraction of features to pick out the crucial characteristics to supervise in to expedite the computation. Future solutions should look for new ways to extract information in order to decrease a processing time technique.

VI. FUTURE TRENDS

8.1 Efficacious NIDS

Latest research has revealed that despite new methods being suggested to improve their accuracy, NIDS have a limited ability to detect zero-day attacks. Future research might focus on creating a current dataset that perfectly captures real-world scenarios and contains a sufficient amount of instances of every intrusion classification even though it has been determined that no approach is sufficient. A further option is to create a structure that combines the far more latest techniques for deep learning, with both the outcomes utilized to constructive play distinct traits so every time an altogether new instance of such an attack is encountered.

8.2 Advanced Model Solutions

Deep learning techniques are increasingly used in recent research, which is leading to more complex models. One remedy for this issue involves employing a cloud platform and perhaps a elevated GPU tool for one‟s computational resources. These methods, however, can be costly. Another option is to use new techniques for extracting features to optimise the dataset that's going to be utilized even during training process. This method can reduce the model's complexity while increasing the solution's accuracy.

8.3 Encrypted Traffic Identification

With the growing utilisation encryption across nearly everything, such as malicious hacker operations, there is an increasing need to identify unusual activity from encrypted traffic. Existing approaches for detecting encrypted traffic, however, are inefficient, and the could go used it to evaluate these options do not properly depict today's networks, where a lot of information is encrypted. To detect unusual occurrences in today's networks, new strategies that really can retrieve key aspects from computer networks are required.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

8.4 Usage of Extracting Features

Feature extraction is indeed a newer method in IDS strategies with demonstrated promising outcomes such as decreasing model complexity. Despite this, only 60% of the research literature in this field used ways to extract features, and in many instances, outdated techniques such as Principal Component Analysis were used. As a result, future research efforts must prioritise by mean of innovative DL approaches to extractfeatures, accompanied by simple ML way in order to lower the data processing adequate funding for implementation.

IX. CONCLUSION

This research paper offers insights into Intrusion Detection Systems as well as their potential for advancement by employing machine learning techniques. The paper initially examines the intrusion detection systems concept (IDS) and their different kinds, such as Network, Host, but also Hybrid IDS, in addition to how those who can detect attacks by using signatures as well as comparing network behaviour to a baseline. The paper then delves into the metrics used to evaluate IDS effectiveness, such as Accuracy, Detection Rate, as well as F-Measure.

Following that, the paper offers a summary of machine learning as well as its three major classifications: supervised, semi-supervised, as well as unsupervised learning. The paper after which review sites published recently papers that use algorithms for intrusion prevention, that utilizes machine learning and finds that deep learning techniques have grown increasingly popular, though at this same cost of increasing model complexity and computing resources. According to the paper, numerous IDS solutions are based on extracting features with Auto-Encoder. The study outlines several limitations in current IDS research, such as the use of out-of-date datasets and the requirement of a dataset that includes latest attack instances as well as minor classes to low detection rates. The paper also mentions that there are still significant challenges with IDS complexity but also low accuracy for minimal classes.

Finally, the paper suggests several possibilities for further study, including the formation of an NIDS framework which can be developed primarily through cloud computing and the development of a method for discussing network traffic that is encrypted.

REFERENCES

[1]. The History of Intrusion Detection Systems (IDS), 2022
[2]. What Is an Intrusion Detection System?Checkpoint., 2022
[3]. All Machine Learning Models Explained in 6 Minutes, 2022
[4]. IBM Cloud Education. Machine Learning, 2022
[5]. Baraa I. Farhan et al, Performance analysis of intrusion detection for deep learning model based on CSE-CIC-IDS2018 dataset, 2022
[6]. Chauhan, N. Naïve Bayes Algorithm: Everything You Need to Know, 2022
[7]. Danalakshmi Durairaj, Thiruppathy Kesavan Venkatasamy, Abolfazl Mehbodniya, Syed Umar, and Tanweer Alam. Intrusion detection and mitigation of attacks in microgrid using enhanced deep belief network, 2022
[8]. Vahid Majidnezhad,Avaz Naghipour, A new intelligent intrusion detector based on ensemble of decision trees, 2022
[9]. Chen, Y.; Yuan, F. Dynamic detection of malicious intrusion in wireless network based on improved random forest algorithm, 2022
[10]. Geeta Singh & Neelu Khare, A survey of intrusion detection from the perspective of intrusion datasets and machine learning techniques, 2021
[11]. Gu, J.; Lu, S. An effective intrusion detection approach using SVM with naïve Bayes feature embedding. Comput. Secur. 2021.
[12]. Pan, J.-S.; Fan, F.; Chu, S.C.; Zhao, H.; Liu, G. A Lightweight Intelligent Intrusion Detection Model for Wireless Sensor Networks. Secur. Commun. Networks 2021
[13]. Chen, L.; Kuang, X.; Xu, A.; Suo, S.; Yang, Y. A Novel Network Intrusion Detection System Based on CNN,2020.
[14]. Yu, Y.; Bian, N. An Intrusion Detection Method Using Few-Shot Learning. IEEE Access 2020,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, March 2023
Impact Factor: 7.301

[15]. Andresini, G.; Appice, A.; Mauro, N.D.; Loglisci, C.; Malerba, D. Multi-Channel Deep Feature Learning for Intrusion Detection,IEEE Access 2020

[16]. Xiao, Y.; Xing, C.; Zhang, T.; Zhao, Z. An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks. IEEE Access 2019

[17]. Zhang, X.; Chen, J.; Zhou, Y.; Han, L.; Lin, J. A Multiple-Layer Representation Learning Model for Network-Based Attack Detection. IEEE Access 2019

[18]. Gao, X.; Shan, C.; Hu, C.; Niu, Z.; Liu, Z. An Adaptive Ensemble Machine Learning Model for Intrusion Detection. IEEE Access 2019

[19]. Wei, P.; Li, Y.; Zhang, Z.; Hu, T.; Li, Z.; Liu, D. An Optimization Method for Intrusion Detection Classification Model Based on Deep Belief Network. IEEE Access 2019

[20]. Khan, F.A.; Gumaei, A.; Derhab, A.; Hussain, A. A Novel Two-Stage Deep Learning Model for Efficient Network Intrusion Detection. IEEE Access 2019

[21]. Liang, D.; Liu, Q.; Zhao, B.; Zhu, Z.; Liu, D. A Clustering-SVM Ensemble Method for Intrusion Detection System, 2019

[22]. Gautam, R.K.S.; Doegar, E.A. An Ensemble Approach for Intrusion Detection System Using Machine Learning Algorithms, 2018

[23]. Gautam, R.K.S.; Doegar, E.A. An Ensemble Approach for Intrusion Detection System Using Machine Learning Algorithms, 2018

[24]. Shone, N.; Ngoc, T.N.; Phai, V.D.; Shi, Q. A Deep Learning Approach to Network Intrusion Detection. IEEE 2018

[25]. Yan, B.; Han, G. Effective Feature Extraction via Stacked Sparse Autoencoder to Improve Intrusion Detection System. IEEE Access 2018

[26]. Ali, M.H.; Mohammed, B.A.D.A.; Ismail, A.; Zolkipli, M.F. A New Intrusion Detection System Based on Fast Learning Network and Particle Swarm Optimization. IEEE Access 2018

[27]. Al-Yaseen,W.L.; Othman, Z.A.; Nazri, M.Z.A. Multi-level hybrid support vector machine and extreme learning machine based on modified K-means for intrusion detection system. Expert Syst. Appl. 2017

[28]. Jabez, J.; Muthukumar, B. Intrusion Detection System (IDS): Anomaly Detection Using Outlier Detection Approach 2015

[29]. Harafaldin, I.; Lashkari, A.H.; Ghorbani, A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization, 2018

[30]. Elhefnawy, R.; Abounaser, H.; Badr, A. A Hybrid Nested Genetic-Fuzzy Algorithm Framework for Intrusion Detection and Attacks. IEEE Access 2020