Network Attack Detection System Based on Hybrid Deep Learning Technique in Cyber Security Application

Dr. Narendra Kandoi
Associate Professor
Department of Computer Sci & Engineering
S.S.G.M. College of Engineering
Shegaon
kandoin2011@gmail.com

Abstract- Technology, procedures, & controls used in cyber security are aimed at preventing cyber-attacks on systems, networks, programs, devices, & data. Its goal is to safeguard systems, networks, & technology against unauthorized use and cyber-attacks. Using a Network Attack Detection System, network security breaches may be detected & contained in companies. Creating a flexible & effective network intrusion detection system (NIDS) for unexpected & unanticipated assaults, on the other hand, poses numerous difficulties. A hybrid ID architecture built on DL-based prediction & classification of destructive network cyberattacks & protection of security is developed in this article using an AlexNet. The CNN pretrained Alexnet uses convolution to collect local features, while the recurrent neural network (RNN) catches temporal data to enhance the efficiency & predictions of the ID system. The hybrid method HARNN ensures that no intrusive packets transmit thru and enter our systems by using a combination of techniques. The hybrid convolution RNN intrusion detection system was tested using publicly accessible ID data, including the contemporary & realistic CSE-CIC-DS2018 data. Using 10-fold cross-validation, the simulation outcomes indicate that the suggested network attack detecting system surpasses existing ID methods in terms of malicious assault identification rate accuracy using CSE-CIC-IDS2018 data.

Keywords: Cyber Security; Recurrent neural network; deep learning; AlexNet; intrusion detection system; machine learning.

1. INTRODUCTION

Avoiding damaging attacks on computers, servers, smartphones, and other mobile devices are referred to as "cyber security." Security of electronic data or security of information technology is some terms that are used to describe it. Physical-layer wireless communication technology's fast advancement also introduces new security concerns. Wi-Fi eavesdropping, spoofing identities, & tampering with data are just a few of the security problems that plague those who use wireless communication networks. Many more physical devices are now linked to a network due to the advancement of Internet technologies. When devices are connected, a big no. of data is produced & stored. Over time, the 'big data' age will arrive [2]. Network assaults are common due to the system's complexity and the wide range of assault techniques available. There are two types of network attacks: active and passive. A malevolent party gains access to networks, monitors them, & steals sensitive data while creating no changes. This is called a passive network assault. Active network assaults alter, encrypt, or corrupt data. Since hackers' technology is improving all the time, network assault detecting systems must be more sophisticated & effective than ever before if they are to be effective. A robust, dependable, & precise intrusion detection model offers wide application possibilities for enhancing network security, given the constraints of conventional network security protection technologies.

Keeping an eye on one's cyber security condition is a new priority for information security professionals nowadays. Cyberspace security status identification & comprehension of possible incursion activities are two of its primary tasks. The prediction of specific network assaults has received a lot of attention to better comprehend possible incursion activity. The motives for network assaults, on the other hand, are much more significant [3]. The term can be used in a variety of contexts, from business to mobile computing, or it can be divided into a few broad groups. [4].

- **Network security** is a computer network against intruders, like targeted assailants or opportunist viruses, is the technique of network security.
- **Application security** is concerned with preventing malicious code from running on a computer's software or hardware. A compromised program may provide hackers access to information that was supposed to be secure. Before deploying software or device, successful security must be designed.
- **Information security** storage & transmission of data are safeguarded by information security, which keeps the data's integrity & privacy intact.
- **Operational security** process & decision making for managing & safeguarding digital assets are part of operational security. There are many rights consumers have when using a network, and procedures that regulate how and

where data may be stored and transferred are all part of it as well.

- **Disaster recovery and business continuity** in the case of a cyber-security accident or any other incident that outcomes in loss of procedures or data, disaster recovering & commercial continuity describe how an organization reacts. Policies for disaster recovery specify how the company will re-establish procedures & info to pre-event levels of performance. When a company is unable to use particular resources, business continuity is the strategy on which it falls backs.
- **End-user education** End-user education tackles the most unpredictability element in cyber-security: humans. In the absence of proper security procedures, anybody may unintentionally present a virus into an otherwise safe system. Ensuring that employees know how to remove suspicious email attachments, not put in unknown USB devices, & teaching them other key security skills is essential for every company's security [5].

There are two types of Network IDS. Signature-based detection systems (misuse detecting) Anomaly-based detecting systems are another. Signature-based detecting systems identify attacks by looking for certain patterns in the data. In other words, they've shown that it only works on well-known signature patterns. Consequently, since they can't identify patterns in prior data, the system fails to detect a new assault. On the other hand, anomaly-based detecting system performs their detection on basis of their observations of trend or behavior which has any deviation from the normal. So that this system can detect any new attacks without any previous knowledge based on the built-in model [6].

To prevent & defend against cyber security risks, it is necessary to identify them quickly. Thomas Friedman postulated that technological advancements had "flattened" the formerly uneven playing field. In other words, platforms, procedures, & individuals have all come together in a flattening effect. The complexity & dynamic properties of cyber security info are poorly understood, which makes it difficult to enhance total cyber security task efficiency. Since 1st computer virus made its way over the Internet, it's been clear that assaults may quickly propagate across national borders. [7].

Monitoring for network intrusions has shown to be a useful tool for network security since it can identify unknown assaults & offer crucial assistance. National authorities are progressively playing the role of info brokers by establishing national cyber security centers & disseminating alerts about emerging attack vectors and also critical mitigation suggestions. These efforts have some success, but they are severely hampered by their shortcomings. For intrusion detection, it's essential to differentiate hostile assault activity from legitimate network traffic by looking for certain characteristics. As a result, intrusion detection may be seen as a challenge in network traffic categorization. It's possible to categorize network traffic into two types: normal and malicious, and additionally, there are 5 types of network traffic: normal, U2R (User to Root attacks), DoS (Denial of Service attacks), R2L (Root to Local attacks), & Probe (Probing attacks). Classifiers are used in intrusion detection to recognize malicious traffic more accurately, which is the primary goal. For the time being, conventional ML methods are used to classify network traffic. RNNs (that have been around for decades but whose full possibilities have only lately began to be broadly recognized, like CNNs) have currently created a substantial advancement in the field of DL as a result of increasing computational resources. [8].

We have suggested a deep learning method for an IDS utilizing ALEXNET, inspired by recurrent neural networks, which can extract superior descriptions from data to build significantly improved methods [9]. As computer technology advances & current attack detection techniques become more flawed, one of the most pressing issues in info security is how to identify network assaults. To correct the situation in this area, the current efforts of researchers are aimed at finding and applying new, including hybrid and adaptive, detection schemes. Computer network security requires the use of an IDS. Its main task is to monitor the network or system for malicious activity. Although the problem of detecting network attacks is quite old, it is still relevant. A few years ago, network intrusions were rare because they required extensive knowledge of operating systems, network protocols. Today, any user can carry out malicious actions on the network by downloading one of the exploit programs available on the Internet [10].

1.1 Cyber Security

Security of computer systems & networks, also known as information technology (IT) security, is the process of guarding against unauthorized access to, or theft or harm to, computer systems & networks or electronic data, and also interruptions to or rerouting of services these processes offer. [11].

Computer systems, the Internet [12], & wireless network protocols like Bluetooth & Wi-Fi are becoming more important because of growing dependence on them, as are "smart" gadgets like smartphones, TVs, or other items which are part of the "Internet of things." Because of its complexity, in both terms of political application & technological implementation, cybersecurity is a major problem in today's society. [13].

1) Security architecture

IT security construction is defined as "the building artifacts that clarify how the security mechanisms (security countermeasures) are positioned & how they relate to wider IT architecture" by this organization. To maintain a high degree of quality features like confidentially, integrity, accessibility & accountability, these controls have been put into place in the software system.[14]

According to Techopedia, security architecture is defined as "security architecture that takes into account both requirements & possible hazards of a given situation or environment. If and when security measures should be applied are also specified. In general, the designing procedure may be replicated." The following are some critical characteristics of a security architecture:[15]

the connections & interdependencies among several parts.

- Assessment of risks, good practices, economics, & legal issues to determine controls.
- the standardization of control measures.

Practicing security infrastructure lays the groundwork for addressing business, IT, & security issues in a systematic manner in a company.

2) Security measures

To achieve the conceptual ideal of computer "security," three procedures must be employed: threat prevention, detection, & response. The following policies & system parts provide the basis for these procedures: [16]:

- System files & data may be safeguarded via user account access controls & cryptography, accordingly.
- When it comes to network security, firewalls are the most popular preventive solution since they may protect access to inner network services while also blocking some types of assaults via packet filtering (if set correctly). Hardware & software firewalls are also available.
- Products like Intrusion Detection Systems (IDS) & audit trails and logs are used to identify network assaults while they are still in process and to aid in post-attack forensics.
- Responding to a security incident requires a thorough understanding of the system's security needs and could include anything from a basic security update to notifying legal authorities, mounting a counter-attack, etc. In certain instances, it's preferable to destroy the hacked system completely since not all of its affected resources have been discovered.

Computer security nowadays relies mostly on "preventive" measures such as with a firewall or an exit process. To filter network data between hosts or networks, such as the Internet, a firewall can be designed & implemented as computer software that connects to network stack (or, in most UNIX-based operating systems like Linux, is incorporated into the kernel) & allows real-time filtering & blocking of network traffic. Another option is to employ a "physical firewall," which utilizes a distinct computer to filter network data. Firewalls are common on computers connected to the Internet all the time.

Large data systems like Apache Hadoop are being used by some companies to increase data availability and use ML to identify more sophisticated risks.[17]

However, only a small number of businesses have computer systems equipped with reliable detection systems and even less have well-oiled reaction processes. This means, according to Reuters, Companies say for the first time that electronic theft of data causes them to lose more money than physical losses of assets. [18] The main difficulty in successfully combating cybercrime is an over-reliance on firewalls and other automatic "detection" schemes. However, it is the use of packet capture equipment to collect fundamental evidence that helps put offenders in prison. The CIA triad, which protects the confidentiality, integrity, & accessibility of a network, is the cornerstone of data security and must be maintained at all costs.

[19] Administrative, physical, & technological security measures are required to meet these goals. An asset's security value could only be calculated after its quantity is known.[20]

2. LITERATURE REVIEW

This section highlights some of the most significant recent developments in this field. Notably, our discussion is limited to studies that utilized the NSL-KDD dataset as an efficiency benchmark. As a result, from now on, every dataset shall be referred to as NSL-KDD. This method provides for a more precise comparing of research results with those from the literature. Another drawback is that most of the time, training data is used both for training & testing. As a final note, we describe a few deep learning-based methods which have been tested for comparison purposes using the CSE-CIC-DS2018 dataset.

C. Feng et al (2017) ICS networks make use of predictable & regular characteristics of interaction trends between so-called field devices. We build a generic package signature database by monitoring a system without abnormalities for an extended length of time. The signature database is stored in a Bloom filter, that is subsequently utilized to identify anomalies at the package content level. Our method of time-series anomaly detection also suggests a network-based stacking Long Short-Term Memory (LSTM) classifier that learns to anticipate the most probable package signatures depending on previously observed package traffic. At long last, we demonstrate that a gas pipeline SCADA anomaly identification method incorporating both methods may outperform different existing state-of-the-art algorithms by inspecting an actual dataset. [21].

U. Sabeel et al (2019) have shown that DL is successful in identifying assaults with well-known profiles, i.e. assault patterns with which DL-based techniques have been trained. Furthermore, the effectiveness of these defenses against unknown or constantly altering assaults

has not been well studied. The growing complexity of assaults on network-based resources necessitates a better understanding of how DL-based techniques will fare in real-world situations or how well they could manage departure from their training prototypes. For binary predictions of unknown DoS & DDoS assaults, they compare the effectiveness of two widely suggested DL-based techniques: DNN & LSTM. Models are built using benchmarking CICIDS2017 dataset, & their efficiency is then evaluated using a new test dataset generated in a simulated setting. The True Positive Rate (TPR) for DNN & LSTM is improved by 99.8% & 99.9%, respectively, by retraining the frameworks on a dataset containing new unknown assaults [24].

M. Tan, et al (2019) Their system, which utilizes time slot-based characteristics, may be used for real-time assault detection, unlike several other approaches that rely on the attention mechanism already in place. The transformers framework has been developed & utilized in the language translation area as a modified version of the suggested solution. They do tests on a dataset derived from network traffic in a current repository that contains a variety of network attacks. They compare the suggested approach to two baselines, the "bidirectional LSTM" and the "conditional random fields," & find that the suggested solution exceeds both in terms of accuracy, recall, & FPR. They further demonstrate that by removing recurrent layers, their approach is more computationally efficient than the bidirectional LSTM prototype [23].

S. Nayyar, et al (2020) detected abnormalities in network traffic & sent all malicious requests to a honeypot-based black hole server using a machine learning framework depending on latent semantic analysis (LSTM). CICIDS2017 data was used to train and test the method, which included attacks like Patator, Web Brute Force, DDoS Hulk, DoS GoldenEye & DDoS LOIT, and other types of category assaults such as DoS slower & slow httptest. Detection accuracy for these assaults is excellent, and the framework may be used in distributed systems that already exist. [24].

A. Andalib and V. T. Vakili, (2020) suggested an IDS that minimizes the number of human communication & required expert knowledge while still delivering adequate efficiency under zero-day assaults. Instead of using one deep learning method, we're using three deep learning approaches simultaneously: gated recurrent unit (GRU), CNN, & random forest (RF). To aggregate the outcomes of these systems, two logics are used: majority vote & "OR" logic. To ensure that their suggested system is capable, they test it against the NSL-KDD dataset. Based on simulation findings, the system is capable of operating with relatively little involvement from technicians in the event of a zero-day assault. In the "KDDTest+" dataset from NSL-KDD, they got an accuracy of 87.28 percent, whereas, on the more difficult "KDDTest-21" dataset, we got an accuracy of 76.61 percent. [25].

S. N. Pakanzad and H. Monkaresi, (2020) enhanced the efficiency of IDS by utilizing a hybrid method combining a CNN with an LSTM network. While distinguishing among normal & abnormal traffic with some precision has been accomplished in earlier research, multi-class categorization precision has not been perfect. There is a desire to develop a technique for properly categorizing malware-infected communications in terms of attacks. Data from NSL-KDD & CICIDS2017 databases were used to verify findings. The NSL-KDD & CICIDS2017 datasets had multiple categorization accuracy of 98.1 & 96.7, correspondingly [26].

B. Wang, et al (2020) the development of a deep hierarchical network capable of learning about features of traffic from raw packet data to identify malicious activity at the packet level was presented. The spatial characteristics of raw packets were extracted using a one-dimensional convolutional layer, while the temporal characteristics were extracted using a GRU structure. Experiments were carried out on the suggested deep hierarchical network's effectiveness using the ISCX2012 dataset, the USTC-TFC2016 dataset, and the CICIDS2017 dataset, accordingly, to assess its efficiency. There are three measures for evaluating the product: Accuracy (ACC), detection rate (DR), & false alarm rate (FAR). In the ISCX2012 dataset, they obtained ACC, DR, & FAR accuracy rates of 99.42%, 99.74%, & 1.77%, correspondingly. There were 99.94%, 99.99%, & 0.99 percent in the USTC-TFC2016. The percentages were 100 percent in CICIDS2017, and the percentages were zero. Their discussion delved into how data balance impacts categorization effectiveness & time efficiency in comparison to GRU & LSTM models. As shown by the results of the experiments, their technique is capable of detecting malicious traffic with more accuracy & greater efficiency than several other state-of-the-art used ACC/DR approaches. [27].

C. Yue, et al (2021) in specific IP Scan, Port Scan, Denial of Service (DoS), & Man in Middle (MITM) assaults on the train ECN are suggested as a new ensemble intrusion detection technique for defensive network detection. The raw data produced by our ECN testbed is used to create a particular dataset, which contains 34 characteristics from various protocol components. The dataset will be improved using methods like data imaging & creating a temporal sequence. Six basic categories are constructed using various common convolutional & recurrent neural networks: SimpleRNN, LeNet-5, AlexNet, VGGNet, LSTM & GRU. To combine all of the existing classifiers, a dynamic weight matrix voting scheme is suggested. Their dataset is used to assess the suggested technique. These tests demonstrate their technique's exceptional ability to combine the best features of all previous classifiers, resulting in a better detection rate of 0.975. [28].

3. METHODOLOGY

In this section, we go over the details of our suggested prototype, problem statement, research methodology, proposed algorithm, and dataset.

3.1 Problem Statement

Currently, the most common issues with machine learning models are:

- With a wider spectrum of hostile incursions, these methods have a high false-positive rate (FPR) [29];
- These frameworks are not generalizable owing to obsolete ID datasets, which prevent most current ID systems from detecting new assaults
- High-speed network traffic is increasing rapidly nowadays, & maintaining it in a diverse environment requires state-of-the-art solutions.

Due to these difficulties, a hybrid convolutional RNN-based ID system is being developed, which will use a real-world dataset to test the effectiveness of machine learning & deep learning classifiers in the ID domain. The limits of ID techniques have been discussed above; thus, in their suggested ID system, they combine the two methods to overcome their drawbacks & offer a novel classical technique that combines the benefits of two methods which have improved efficiency over conventional ways. To enhance the ID system's learning capacity & performance, they present a new IDS which combines modern DL approaches like CNN with traditional ML techniques like RNN.

3.2 Dataset

Unverified & unreproducible incursions may be found in some of the publicly accessible ID databases. The Amazon Web Services (AWS) infrastructure created the well-known CSE-CIC-DS2018 [30] dataset to overcome these shortcomings & generate current traffic patterns. Several datasets are included for evaluating anomaly-based methods. The CSE-CIC-DS2018 intrusion dataset includes a variety of intrusion states & shows real-time network behavior. As an added bonus, it's spread throughout the entire network, encapsulating all of the internal network traces used to compute data packet payloads. These features of the CSE-CIC-DS2018 dataset encourage us to use it in their study for the development of a new intrusion detection system. A more logical & beneficial approach to network security is anticipated as a consequence of the suggested RNN- & CNN-based ID system. Using the concept of profiles, the Canadian Institute for Cybersecurity (CIC) and Communications Security Establishment (CSE) have worked together to create a systematic cybersecurity dataset. Details about intrusions are included, as are abstract distribution frameworks for apps, protocols, & lower-level network elements. There are 7 distinct assault situations included in the dataset, like

- Heartbleed,
- Botnet,
- Brute-force,
- DoS,
- Web attacks,
- DDoS,
- and internal network penetration.

Attackers use 50 machines, whereas the target company uses 420 PCs & 30 servers across 5 departments. This dataset contains the network traffic & log files of every victim computer, and also 80 network traffic features gleaned from traffic collected by CICFlowMeter-V3.

3.3 Proposed Methodology

The construction of the ID scheme is shown in Figure 1. There are two phases to the learning process, as shown here. In this method, we construct an IDS using AlexNet and RNN based deep learning approach. Our existing hybrid convolutional recurrent neural network is economical in terms of computational complexity while utilizing datasets with full features & offers enhanced precision with the lowest chance of FAR.

With a large data processing architecture, AlexNet & RNN-based deep learning target addressing actual ID issues. Solving this issue is difficult because of the limited amount of time & space available. Big data now have large quantities and is growing, but it requires great power, specialized resources, and a computing device to help in learning procedure which can manage the data effectively.

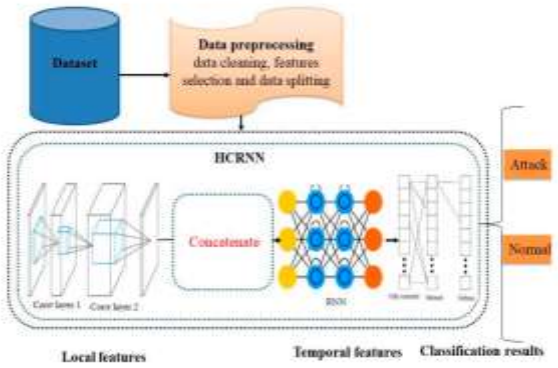

Fig. 1. Overview of the Model

A CNN AlexNet DL framework combined with an RNN reduces these difficulties. The experiment is based on the RNN's key structure, which may be found right here. Figure 1 illustrates the hybrid convolutional recurrent neural network in detail. As can be seen from the prototype overview, an AlexNet is made up of two main parts: a feature extractor or a classifier. Convolution & pooling layers make up the feature extractor. The feature map created from collected data is used as the input for the part of the second classification. The CNN-pretrained AlexNet picks up on the local details quickly in this manner. Furthermore, one flaw is that it neglects to take into account the interrelationship between several essential characteristics. After the Alexnet layers, we added recurrent layers to better capture spatial and also temporal features. In this manner, the disappearing & inflating gradient issues were successfully solved, improving the capacity to record

spatial & temporal relationships & quickly learn from variable extent sequences.

The NumPy, Pandas, & Scikit-learn libraries for Python programming language were used to perform data pre-processing procedures. The scientific community has focused a great deal of emphasis on the problem of class disparity. An unbalanced distribution of samples leads to a class imbalance; one class has the majority of samples while other classes include relatively few. As data dimensionality rises owing to limitless data values & imbalanced classes, the issue of the classification gets increasingly difficult.

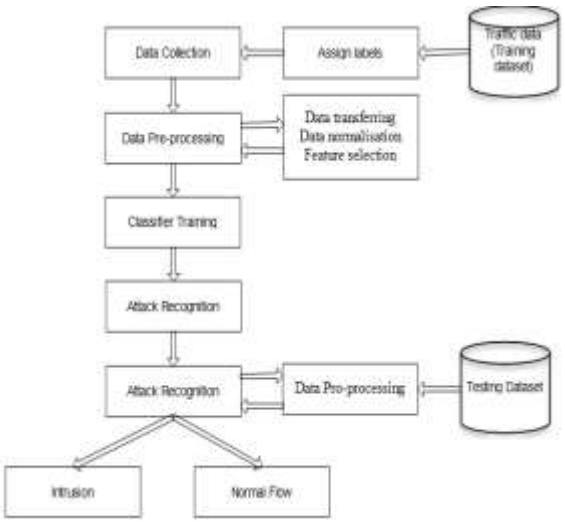

Fig. 2. Proposed System Architecture

3.4 ALGORITHMS USED

1) Recurrent Neural Network

Neural networks that use the result of earlier steps as inputs are referred to as recurrent neural networks (RNNs). Traditional neural networks have inputs & outputs that are completely independent of one another, but in certain instances, including if trying to predict the next word in a phrase, it is necessary to keep track of words that have come before. As a result, RNN was created, which used a Hidden Layer to address this problem. RNN's most critical feature is its Hidden state memory, which saves certain sequence info.

Steps:

Allow us to speculate about the existence of a more complex network, one that has three hidden levels and a single output layer. As with previous neural networks, each hidden layer will have its weights & biases, such as (w1, b1) for hidden layer 1, (w2, b2) for hidden layer 2, and (w3, b3) for hidden layer 3. To put it another way, because every layer is independent of the preceding one, past outputs are not remembered.

- The network receives a single time step of input.
- Then, given the existing input and the prior state, determine the system's current state.
- For the following time step, the present ht becomes ht-1.
- As numerous time steps as necessary to solve the issue may be taken, and the data from all prior stages can be combined.
- The last present state is utilized to compute the output after all the time steps have been finished.
- Afterward, the output is contrasted to the intended output to determine the error.
- Afterward, the mistake is sent backpropagated into the network, which modifies the weights. As a result, the RNN becomes trained.

MATHEMATICAL MODEL:

As a sampling from training distributions for RNN ∈, the stochastic gradient descending in Contrastive Divergence utilizes a learning rate of x1. The input b is an RNN offset vector unit are represented by the weight matrix W, which has a size of (number of hidden units, number of inputs). c is RNN offset vector for the hidden unit

Notation: Q (h2i = 1|x2) is the vector with elements Q (h2i = 1|x2)

Step 1: for all hidden units i do

Step 2: compute Q ($h_{1i} = 1|x_1$) (for binomial units, sigm($c_i + \sum_j Wijx1j$))

Step 3: sample $h_{1i} \in \{0, 1\}$ from Q ($h_{1i}|x_1$)

Step 4: end for

Step 5: for all visible units j do

Step 6: compute P ($x_{2j} = 1|h_1$) (for binomial units, sigm($b_j + \sum_i Wijh1i$))

Step 7: sample $x_{2j} \in \{0, 1\}$ from P ($x_{2j} = 1|h_1$)

Step 8: end for

Step 9: for all hidden units j do

Step 10: compute Q ($h_{2i} = 1|x_2$) (for binomial units, sigm ($c_i + \sum j W_{ij}x_{2j}$))

Step 11: end for

Step12: $W \leftarrow W+ \in (h1x'_1 - Q (h_{2i} = 1|x_2)x'_2)$

Step 13: $b \leftarrow b+ \in (x_1 - x_2)$

Step 14: $c \leftarrow - c+ \in (h_1 - Q (h_{2i} = 1|x_2))$

2) Proposed Alexnet Model

AlexNet is convolutional neural network design is called AlexNet. For cyber security, we utilized an AlexNet convolutional neural network. The fundamental building elements are convolutions, max pooling, & dense layers. To fit the prototype over two GPUs, we utilize grouped convolutions. The ImageNet Large Scale Visual Recognition Challenge 2012 was won by Alex-Net, which used an 8-layer CNN. This network demonstrated for an initial time that learning-based characteristics may go beyond those created by hand, upending the established order in computer vision.

LeNet & AlexNet have extremely similar architectural designs, as seen in Figure 3. Certain architectural oddities were required in 2012 to fit the prototype on

two tiny GPUs, which we have removed from our version of AlexNet.

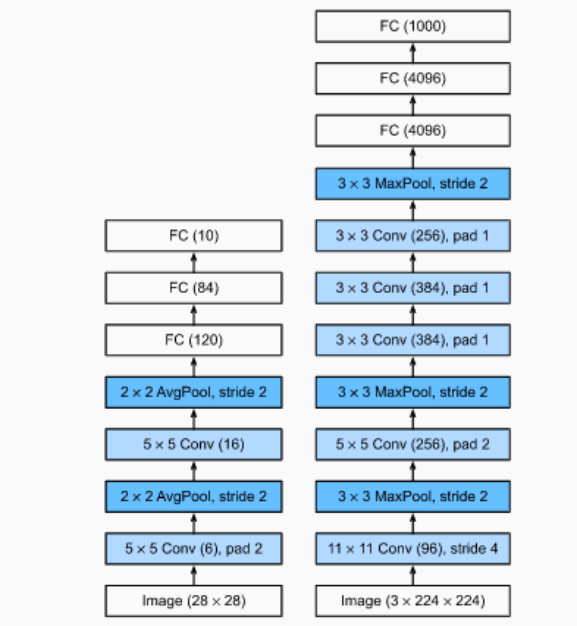

Fig. 3. From LeNet (left) to AlexNet (right).

While AlexNet & LeNet's design principles are strikingly similar, there are some major distinctions as well. To begin, Alex-Net is much larger than LeNet5, which is considerably smaller in comparison. Eight layers make up AlexNet, five of which are convolutional, two of which are fully connected hidden layers, and one of which is a fully connected output layer. Secondly, AlexNet's activation function was the ReLU rather than the sigmoid. Let's take a closer look at the specifics below:

Architecture of AlexNet:

There are 5 convolutional layers & three fully connected layers in the system's eight-layer design. These aren't the only characteristics that are novel methods to CNNs, but they are what set AlexNet apart from other networks.

The convolution window shape of AlexNet's 1st layer is 11×1111×11. ImageNet pictures are 10 times wider & greater than MNIST pictures, thus items in ImageNet data often take up a larger percentage of the available pixels. Therefore, to collect the item, a wider convolution window is essential. The 2nd layer's convolution window shape is decreased to 5×55×5, followed by is 3×33×3. In addition, the network includes maximum pooling layers with a window shape of 3×33×3 and a stride of 2 after the 1st, 2nd, & 5th conv layers. The AlexNet convolution network contains 10 times the number of convolution channels as LeNet, for instance. There are two 4096-output layers after the final convolutional layer. The prototype parameters generated by these two massive, fully connected layers are close to 1 GB in size. Because early GPUs had restricted memory, AlexNet originally utilized a split data stream architecture, which allowed every one of its two GPUs to store & compute just half of the prototype. As a result of GPU memory becoming more plentiful, it is no longer necessary to split up models among GPUs (our version of AlexNet prototype deviates from a real article in this aspect). And AlexNet switched from sigmoid activation to an easier ReLU activation in its latest release. ReLU activation function calculation is easier on the one hand. The exponentiation process present in the sigmoid activation function, for instance, is absent from this function. The ReLU activation function, on the other hand, simplifies prototype training if various initialization techniques are used. To put it another way: since this area's gradient is so close to 0, backpropagation cannot continue to update certain prototype parameters if the sigmoid activation function's outcome is near 0 or 1. The ReLU activation function's gradient in the positive interval, on the other hand, is always 1. Consequently, incorrect initialization may cause the sigmoid function to have a gradient in the positive interval of almost 0, making it impossible to train a model successfully. We proposed this alexnet layer model to improve performance measures, such as recall, precision, f-1 score, DR, FAR, and precision for detecting network attacks.

4. Experimental Results

4.1 Evaluation Metrics

An effective way to distinguish between actual & anticipated categorization is to use a confusion matrix (CM). Two classifications are created as a consequence of classification: Normal & Anomaly. Four key states must be measured in the confusion matrix.

1) Accuracy

Accuracy is a measure of how many predictions your model has done for the whole test dataset. The following formulation measures it:

$$Accuracy = \frac{Tp+TN}{Tp+Fp+TN+FN} \quad (1)$$

2) Recall

Recall – or the true positive rate – is the measure of how many true positive values of all the positive ones in the data set are expected. Sensitivity is also occasionally termed. The following formula collects the measurement:

$$Recall = \frac{Tp}{Tp+FN} \quad (2)$$

3) Precision

Precision is a measure of the accuracy of a positive forecast. In other terms, it signifies how positive you can be if a result is projected as good. The following formula is used for calculation:

$$Precision = \frac{Tp}{Tp+FP} \quad (3)$$

4) F1-Score

The F1-score is the most frequently utilized. It is a combination of accuracy and memory, that is, its harmonic significance. The following formula allows you to compute the F1-score:

$$F1\ Score\ =\ \frac{Tp}{Tp+1\,2\,(Fp+FN)} \qquad (4)$$

$\boldsymbol{Fp}$ = no. of false-negative cases
$\boldsymbol{FN}$ = no. of false-positive cases
$\boldsymbol{TN}$ = no. of true negative cases
$\boldsymbol{Tp}$ = no. of true positive cases

Using confusion matrix requirements listed above, we can compute the system's output. DR & FAR are two critical and often used IDS parameters. FAR is the total number of misclassified regular occurrences, while DR describes the total no. of intrusion events detected by a prototype.

$$FAR\ =\frac{FP}{(TN+FP)} \qquad (5)$$

$$DR\ =\frac{TP}{(TP+FN)} \qquad (6)$$

Because the DR increases while the FAR decreases, we believe the RNN method is superior to conventional ways.

4.2 Evaluation of the Proposed Hybrid Deep Learning

The classifier's results using CSE-CIC-DS2018 are shown in Table 1. The outcomes were generated using a method called random search hyperparameter optimization. Ensemble classifier XGB improves assault classification effectiveness considerably, with an accuracy of 83%. Compared to ensemble-based classifiers, the tree-based classifier is more accurate.

Table 1. Classifier performance with CSE-CIC-DS2018.

Several other methods have been combined as well. Many conventional machine-learning classifiers were employed, but our primary goal was to use RNN to collect both spatial & temporal data more reliably, thus we added recurrent layers following AlexNet's convolution layers. In that manner, we tried to deal with disappearing & inflating gradients in a way that increases the capacity to grasp spatial & temporal relationships and learn quickly from variable extent sequences. Because of a huge number of variables, conventional machine-learning-based classifiers do not work well with high-dimensional data (imbalances). Because of its large dimensionality & scale invariance, it's an excellent fit for this dataset. However, sophisticated DL methods like the CNN AlexNet framework, which correctly identified abuse in up to 98.6% of cases, provided the most significant improvement. Supplementary Materials provide implementation details that explain how the efficiency was improved because of long-term dependence between nonlinear features.

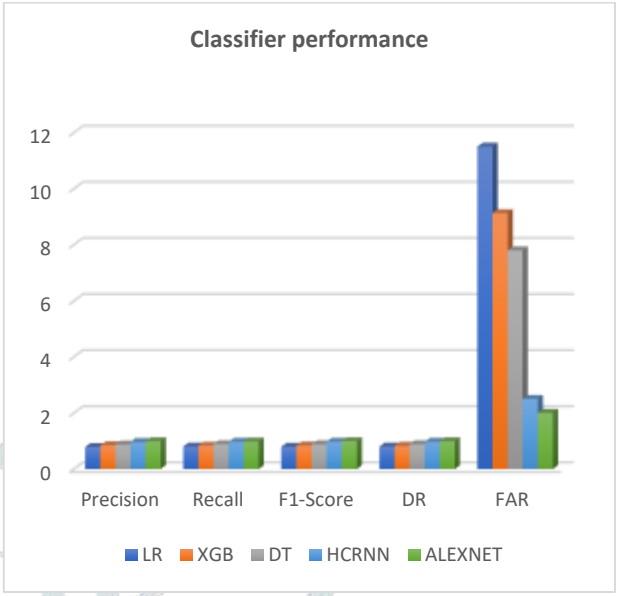

Fig. 4. Bar Graph of parameters of classifier performance comparison

The parameters like precision, recall, F1-Score, DR, and FAR of classifier performance comparison between existing methods and proposed method HARNN represent by bar graph as shown in figure 4.

4.3 Comparative Analysis between existing dataset and proposed dataset

We contrasted the CSE-CICIDS2018 dataset sample size with the CICIDS2017 dataset. Table 2 shows the findings. When contrasted to CICIDS 2017 ID dataset, the sample size of CSE-CICIDS2018 has grown substantially, especially in Botnet & Infiltration assaults, where it has grown by 143 & 4497 accordingly. However, there are only 928 Web Attacks accessible in CSE-CICIDS2018.

Table 2. The contrast of CSE-CIC 2018 ID dataset with CICIDS-2017.

Classifier	Precision	Recall	F1-Score	DR	FAR
LR	0.781	0.801	0.791	0.80	11.50
XGB	0.845	0.834	0.839	0.83	9.13
DT	0.8733	0.885	0.879	0.88	7.8
HCRNN	0.9633	0.9712	0.976	0.97	2.5
HARNN	0.9911	0.9822	0.994	0.99	2.0

Dataset	Normal	DDoS	Dos	Botnet	Brute Force	Infiltration	Web Attacks	Port Scan
ICIDS2017	1,743,179	128,027	252,661	1966	13,835	36	2180	158,930
CSE-CICIDS2018	6,112,151	687,742	654,301	286,191	380,949	161,934	928	-

5. Conclusions and Future Work

Our deep learning method for cyber security has been suggested. It is the process of preventing digital assaults on vital systems & sensitive data that is known as cybersecurity. Cybersecurity measures often referred to as IT security, are intended to fight threats to networked systems & apps, both from inside and outside a company. Deep learning architectures that are commonly used for intrusion detection are examined, as are a few specific applications. As a result, the AlexNet method to feature learning was suggested. After that, this was improved upon by presenting a brand-new classifications method based on an RNN classifications model. The outcomes indicate that the proposed method provides excellent levels of accuracy, precision, & recall while also requiring less training time. Recurrent neural networks enhance the accuracy of the suggested NIDS system. The NIDS was developed with the aid of a cybersecurity-friendly HARNN. We used the CSE-CIC-DS2018 dataset to train the ID system architecture. Several conventional classifications methods were used in the IDS, and also the HARNN methodology for the suggested ID system (LR, DT, XGB, HCRNN, and so on). After Alexnet's convolution layers, we inserted recurrent layers to better capture spatial & temporal information. As a consequence of the suggested ID system, intrusion detection accuracy & DR have both improved while also decreasing computing complexity. Reputable categorization criteria were used to evaluate conventional ML as well as deep learning (DR, Accuracy, Precision, Recall, & F1 score). HARNN can effectively calcify malicious assault events, according to simulated findings. According to the CSE-CIC-IDS2018 statistics, the total accuracy of normal and other kinds of assaults is approximately 98.6 percent.

We want to use the suggested model in a real SDN system in the future and evaluate its throughput & latency properties.
We'll also evaluate the model's performance on a variety of different datasets. Finally, the current work may be expanded to build a management module for taking preventative measures following intrusion detection using categorized outcomes. Our framework outperforms current techniques & points in the direction of a bright future for intrusion detection in big, unbalanced data sets, as shown. In the future, we'll look at other feature selection techniques to see if we can have better results with detection. There are numerous potential applications for this system in the future, including finding abnormalities & misuses in actual picture collections from the Internet of Things. We'll use a feature extraction approach to learn about additional ID problems in current, realistic datasets by examining several other deep learning techniques.

REFERENCES

[1] M. Liu, J. Yang, T. Song, J. Hu, and G. Gui, ‘‘Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks,’’ IEEE Trans. Veh. Technol., vol. 68, no. 1, pp. 641–653, Jan. 2019.

[2] Y. Li, X. Cheng, and G. Gui, ‘‘Co-robust-ADMM-net: Joint ADMM framework and DNN for robust sparse composite regularization,’’ IEEE Access, vol. 6, pp. 47943–47952, 2018.

[3] Q. Zhang, D. Man and W. Yang, "Using HMM for Intent Recognition in Cyber Security Situation Awareness," *2009 Second International Symposium on Knowledge Acquisition and Modeling*, 2009, pp. 166-169, doi: 10.1109/KAM.2009.315.

[4] https://www.kaspersky.co.in/resource-center/definitions/what-is-cyber-security

[5] Z. Xue, J. Wang, G. Ding, Q. Wu, Y. Lin, and T. A. Tsiftsis, ‘‘Deviceto-device communications underlying UAV-supported social networking,’’ IEEE Access, vol. 6, pp. 34488–34502, 2018.

[6] Y. Y. Chung and N. Wahid, “A hybrid network intrusion detection system using simplified swarm optimization (sso),” Applied Soft Computing, vol. 12, no. 9, pp. 3014–3022, 2012

[7] J. J. Walker, T. Jones and R. Blount, "Visualization, modeling and predictive analysis of cyber security attacks against cyber infrastructure-oriented systems," *2011 IEEE International Conference on Technologies for Homeland Security (HST)*, 2011, pp. 81-85, doi: 10.1109/THS.2011.6107851.

[8] M. M.Sakr, M. A. Tawfeeq, and A. B. ElSisi, ‘‘Network intrusion detection system based PSO-SVM for cloud computing,’’ Int. J. Comput. Netw. Inf. Secur., vol. 11, no. 3, pp. 22–29, Mar. 2019

[9] Rieck K, Laskov P . Language models for detection of unknown attacks in network traffic[J]. Journal in Computer Virology, 2007, 2(4):243-256

[10] .A. Karpathy, The unreasonable effectiveness of recurrent neural networks, 2015, [online] Available: http://karpathy.github.io/2015/05/21/rnn-effectiveness/

[11] Jing, Y., Feng, W.: Simulation Modeling of Network Intrusion Detection Based on Artificial Immune System. In: Life System Modeling and Intelligent Computing. Communications in Computer and Information Science, v. 97. Springer, Berlin, Heidelberg, pp. 1-4 (2010)

[12] *Schatz, Daniel; Bashroush, Rabih; Wall, Julie (2017).* "Towards a More Representative Definition of Cyber Security". *Journal of Digital Forensics, Security and Law. 12 (2).* ISSN 1558-7215

[13] "Reliance spells end of road for ICT amateurs", 7 May 2013, The Australian

[14] *Stevens, Tim (11 June 2018).* "Global Cybersecurity: New Directions in Theory and Methods" (PDF). *Politics and Governance. 6 (2): 1–4.* doi:10.17645/pag.v6i2.1569.

[15] Definitions: IT Security Architecture Archived 15 March 2014 at the Wayback Machine. SecurityArchitecture.org, Jan 2006

[16] *Jannsen, Cory*. "Security Architecture". *Techopedia. Janalta Interactive Inc*. Archived *from the original on 3 October 2014*. Retrieved 9 October 2014.

[17] *Woodie, Alex (9 May 2016).* "Why ONI May Be Our Best Hope for Cyber Security Now". Archived *from the original on 20 August 2016*. Retrieved 13 July 2016.

[18] "Firms lose more to electronic than physical theft". *Reuters. 18 October 2010.* Archived *from the original on 25 September 2015*.

[19] *Walkowski, Debbie (9 July 2019).* "What Is The CIA Triad?". *F5 Labs*. Retrieved 25 February 2020.

[20] "Knowing Value of Data Assets is Crucial to Cybersecurity Risk Management | SecurityWeek.Com". *www.securityweek.com*. Retrieved 25 February 2020.

[21] C. Feng, T. Li and D. Chana, "Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks," *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 2017, pp. 261-272, doi: 10.1109/DSN.2017.34.

[22] U. Sabeel, S. S. Heydari, H. Mohanka, Y. Bendhaou, K. Elgazzar and K. El-Khatib, "Evaluation of Deep Learning in Detecting Unknown Network Attacks," 2019 International Conference on Smart Applications, Communications and Networking (SmartNets), 2019, pp. 1-6, doi: 10.1109/SmartNets48225.2019.9069788.

[23] M. Tan, A. Iacovazzi, N. M. Cheung and Y. Elovici, "A Neural Attention Model for Real-Time Network Intrusion Detection," *2019 IEEE 44th Conference on Local Computer Networks (LCN)*, 2019, pp. 291-299, doi: 10.1109/LCN44214.2019.8990890.

[24] S. Nayyar, S. Arora and M. Singh, "Recurrent Neural Network Based Intrusion Detection System," 2020 International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 0136-0140, doi: 10.1109/ICCSP48568.2020.9182099

[25] A. Andalib and V. T. Vakili, "An Autonomous Intrusion Detection System Using an Ensemble of Advanced Learners," *2020 28th Iranian Conference on Electrical Engineering (ICEE)*, 2020, pp. 1-5, doi: 10.1109/ICEE50131.2020.9260808.

[26] S. N. Pakanzad and H. Monkaresi, "Providing a Hybrid Approach for Detecting Malicious Traffic on the Computer Networks Using Convolutional Neural Networks," *2020 28th Iranian Conference on Electrical Engineering (ICEE)*, 2020, pp. 1-6, doi: 10.1109/ICEE50131.2020.9260686.

[27] B. Wang, Y. Su, M. Zhang and J. Nie, "A Deep Hierarchical Network for Packet-Level Malicious Traffic Detection," in IEEE Access, vol. 8, pp. 201728-201740, 2020, doi: 10.1109/ACCESS.2020.3035967.

[28] C. Yue, L. Wang, D. Wang, R. Duo and X. Nie, "An Ensemble Intrusion Detection Method for Train Ethernet Consist Network Based on CNN and RNN," in *IEEE Access*, vol. 9, pp. 59527-59539, 2021, doi: 10.1109/ACCESS.2021.3073413

[29] Kumar, K.P.M.; Saravanan, M.; Thenmozhi, M.; Vijayakumar, K. Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks. Concurr. Comput. Pr. Exp. 2021, 33, 5242. [CrossRef]

[30] A Collaborative Project between the Communications Security Establishment (CSE) & the Canadian Institute for Cybersecurity (CIC). Available online: https://www.unb.ca/cic/datasets/ids-2018.html (accessed on 31 March 2021)