Deep Learning Techniques for Intrusion Detection Systems: A Survey and Comparative Study

1st Mohamed Ahmed Abdel-Rahman
Computer Engineering and Systems
Ain Shams University
Cairo, Egypt
2000694@eng.asu.edu.eg

2nd Mohamed Shalaby
Computer Department
ADRDC Egyptian Armed Forces
Cairo, Egypt
myousef73@hotmail.com

3rd Mohamed A. Sobh
Computer Engineering and Systems
Ain Shams University
Cairo, Egypt
mohamed.sobh@eng.asu.edu.eg

4th Ayman M. Bahaa-Eldin
Computer Engineering and Systems
Misr International University,
On leave from Ain Shams University
Cairo, Egypt
ayman.bahaa@eng.asu.edu.eg

2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) | 979-8-3503-0623-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/MIUCC58832.2023.10278326

***Abstract*— Nowadays cyber threats become increasingly sophisticated and prevalent. Intrusion Detection Systems (IDS) have been widely used, to achieve the necessary security requirements in computer networks because of their ability to detect network attacks. Recently, utilizing machine learning (ML) and deep learning (DL) models in IDS have demonstrated substantial improvements in identifying unknown attacks. This study conducts a comprehensive analysis of DL approaches for intrusion detection focusing on the recent research in the last five years, and explores the most used datasets in the field to highlight their characteristics and suitability for evaluating IDS performance. Finally, we present insights into the limitations, strengths, and future prospects of DL based IDS.**

Keywords— Intrusion Detection System, Deep Learning, DBN, CNN, RNN, LSTM, GAN, Autoencoders, Transformers

I. Introduction

Today, traditional methods are no longer effective enough to prevent cyber threats, because the number of attacks is increasing significantly and the variety of ways in which these attacks are carried out. Intrusion detection system is an efficient security tool that keeps track of network traffic and thwarts malicious requests. Conventional IDS approaches rely on rule-based or signature-based methods, which require a significant amount of human expertise and are limited in their ability to detect previously unseen attack. On the other hand, machine learning and deep learning approaches have been widely used to improve the IDS detection. Deep learning techniques have gained significant attention in the IDS research community because of their ability to handle large and complex datasets and learn representations that capture complex features and relationships. Compared to machine learning techniques, deep learning approaches have higher performance in detecting and classifying malicious network traffic. Several studies have demonstrated the effectiveness of deep learning approaches for IDS, these approaches have achieved remarkable levels of detection accuracy while simultaneously maintaining low false-positive rates. Despite the promising results, deep learning approaches for IDS are still an open research area with several challenges and research gaps.

In this paper, we shade the light on these challenges and research gaps, as well as presenting the related benchmark datasets. The remaining sections of this paper are structured as follows: Section II summarizes the related works in the field. Section III describes our methodology, including the search strategy and research questions. Section VI provides an analysis of IDS taxonomy, the most commonly used deep learning techniques, and the datasets used for training and evaluation. Section V discusses the challenges and future directions. Section VI concludes the remarks and findings.

II. Related Work

We review the existing literature on deep learning techniques for intrusion detection systems. We categorize and summarize the key contributions of related works in Table I.

TABLE I. RELATED WORKS SUMMARY

Related Work	Key Contributions	Strengths and Weaknesses	Challenges and Future Directions
Yang et al. [1]	Review of methods and datasets for anomaly-based network intrusion detection	Comprehensive coverage of literature and datasets	Lack of available labeled datasets for evaluating IDS.
Moustafa et al. [2]	Comprehensive survey of network anomaly detection systems	In-depth coverage of various network anomaly detection techniques. Lack of focusing on specific subareas or recent developments	Difficulty in handling high-dimensional and dynamic network data. Development of more accurate and efficient anomaly detection systems
Aleesa et al. [3]	Review of intrusion detection systems based on deep learning techniques.	Detailed taxonomy and analysis of deep learning-based IDS	Limited availability of labeled datasets for training and evaluating deep learning-based IDS
Lansky et al. [4]	Systematic review of deep learning-based intrusion detection systems	Analysis and evaluation of DL techniques for IDS. Provide insights into their applications in IDS	Insufficient interpretability and explainability of DL models. Exploration of novel architectures and algorithms
Liu et al. [5]	Conduct a comprehensive survey of machine learning and deep learning methods for IDS.	Cover various ML and DL models. Showcases the potential of ensemble learning and transfer learning techniques for IDS.	Imbalanced datasets that affect the performance of IDS. Development of hybrid models combining machine learning and deep learning techniques

979-8-3503-0623-1/23/$31.00 ©2023 IEEE

III. Methodology

To identify relevant papers for this survey, a Systematic Literature Review (SLR) was performed adhering to guidelines presented by Kitchenham et al. [6].

A. Research Questions

We used research questions that cover our survey objectives as following:

1) *What are the deep learning techniques used for IDS?*
2) *What are the most used benchmark datasets?*
3) *What are the evaluation metrics for these techniques?*
4) *What are the research gaps of DL based IDS?*

B. Search Strategy

To answer our research questions, we used two popular databases (Scopus and Web of Science) that cover the major publishers, and limited our search to publications between 2019 and 2023, as we aim at focusing on recent developments in the field. We used a combination of keywords to formulate our search query to be: ("deep learning" OR "neural network" OR "convolutional neural network" OR "recurrent neural network" OR "autoencoder" OR "transformers" OR "generative adversarial networks") AND ("intrusion detection" OR "network security" OR "cyber security").

C. Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for paper selection encompass English-language publications in peer-reviewed journals. Survey papers were excluded to maintain a narrowed focus on the contribution to deep learning techniques. The final selection comprised the top 10 most cited papers for each year, resulting in a total of 50 papers that specifically address deep learning techniques in IDS.

IV. Taxonomy and Techniques

In this section, we present a taxonomy of IDS and discuss the various deep learning methodologies that have been employed in the field of intrusion detection. Additionally, we offer a comprehensive overview of the datasets commonly used to evaluate these approaches, as well as the evaluation metrics used to assess their performance. Mastering these approaches and their advantages and disadvantages, allows us to comprehend the most cutting-edge deep learning-based IDS currently available. and identify potential directions for future research.

A. IDS Taxonomy

Existing IDS can be classified according to three main categories: source of data, detection techniques, and deployment methods as illustrated in Figure 1.

According to data source, they can be divided into four levels: Packet-based, Flow-based, Session-based, and Log-based.

1) *Packet-based IDS:* It is one of the IDS types that works at the network packet level, it inspects the individual packets of network traffic in real-time and analyzes their content to identify security threats.

2) *Flow-based IDS:* It is one of the IDS types that works at the network flow level. We can define a flow as a sequence of packets that shared some common attributes, such as source port, source IP address, destination port, destination IP address, transport layer protocol, and time interval.

3) *Session-based IDS:* It is one of the IDS types that works at the network session level. A session is defined as a stream of data exchanged between two hosts over a period of time, typically using a specific protocol, such as TCP or UDP.

4) *Log-based IDS:* It is one of the IDS types that works at the system log level. It analyzes system log files to detect security threats, such as unauthorized access, malware infections, or system misconfigurations.

According to detection technique, they can be divided into two techniques: Anomaly detection and Signature-based.

1) *Anomaly Detection IDS:* It is one of the IDS techniques that functions by establishing patterns of normal network traffic or system activity and then monitoring for any deviations from these patterns. Any deviations that are detected are flagged as security threats and an alert is generated. ML and DL models are among the methods used in the anomaly detection, because they can detect anomalies that are difficult to detect using traditional methods.

2) *Signature-based IDS:* It is one of the IDS techniques that functions by comparing the content of network traffic or system activity against a pre-defined set of known attack signatures. Once a match is detected, the system generates an alert to notify the security administrator.

According to deployment method, they can be divided into two methods: Network-based IDS, and Host-based IDS.

3) *Network-based IDS*: It is one of the IDS methods that functions by analyzing network traffic to detect security threats. It can be implemented using two different approaches: inline and passive.

1) *Host-based IDS:* It is one of the IDS methods that works on individual hosts, such as servers, workstations, or endpoints. It analyzes system activity, such as user activity, or system configuration changes, to detect security threats.

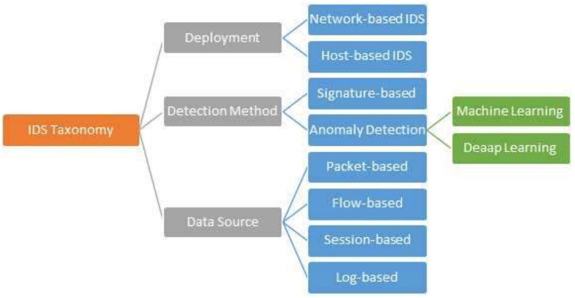

Fig 1. Intrusion Detection Systems Taxonomy

B. Deep Learning Techniques

DL techniques are preferred over ML techniques for detecting unknown attacks. According to Figure 2, DL models can be divided into two main techniques: supervised, and unsupervised. We will concentrate on deep learning approaches and provide up-to-date explanations of and comparison between the latest techniques used for IDS.

1) *Deep Neural Networks (DNN):* DNNs are consist of multiple layers of interconnected nodes. They characterized by their depth, meaning they have several hidden units. This makes them excel at learning intricate patterns and

characteristics from large amount of data and suitable for intrusion detection tasks. Faker et al [7], used three classifiers to improve IDS performance: DNN, Random Forest, and Gradient Boosting Tree (GBT). DNN achieve high accuracy compared to other classifiers. Priya et al [8], used dimensionality reduction and feature selection techniques to enhance the accuracy of DNN architecture in detecting network attacks. Recent studies combined autoencoders with DNN to learn robust representations of the input features and enhance detection of imbalanced attack [9].

2) Convolutional Neural Networks (CNN): CNNs are composed of various layers, including pooling layers, convolutional layers, and fully connected layers. The pooling layers downsample the data, reducing its dimensionality while retaining important information. The convolutional layers use set of kernels on the input data, capturing local patterns and features. The fully connected layers classify the extracted features. CNNs have proven to be highly successful in the domain of computer vision. Interestingly they have also demonstrated capabilities in IDS for identifying network attacks. Several studies used CNNs because they offer benefits when applied to IDS. One notable advantage is their capability to autonomously learn features from the data, eliminating the need for crafted features. This proves advantageous in detecting unidentified attack types. Additionally, CNNs demonstrate proficiency in handling inputs of varying lengths, which's crucial in network traffic analysis since packet lengths often differ significantly. Techniques such, as 1D convolutions and pooling layers can be employed to address this variability effectively. Zhang et al. [10] combined CNN with a new method to reduce class imbalance called SGM to design an efficient flow-based intrusion detection model. Saba et al. [11], implemented CNN as a feature extractor to enhance the acquisition of more effective representations of the input data, and introduced new feature selection technique depend on Reptile Search Algorithm (RSA). The RSA selects the most important features to enhances the performance of IDS. Yu et al. [12], presented a multi-scale CNN based IDS. Comparing it with models based on AdaBoost and RNN, this model achieved higher accuracy and less average detection rate error. Li et al [13], converted network traffic to a sequence of images to reduce the computational cost. Each packet, within the traffic can be depicted as an image, The CNN can then undergo training to categorize these images as either normal or malicious traffic.

3) Recurrent Neural Networks (RNN): RNNs used in IDS to analyze network traffic as a series of events by considering the time-based relationships, between these events they can effectively identify attacks. However, "Vanishing Gradient" problem limits the ability of traditional RNNs to identify long-term dependencies in the data. As a consequence, more advanced types of RNNs, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), were designed. Kasongo et al. [15], presented IDS framework depends on LSTM, GRU and Simple RNN and implemented feature selection algorithm using XGBoost-based to enhance the performance. XGBoost-LSTM was better in binary classification and XGBoost-Simple-RNN was better in multiclass classification. Sahu et al. [16], introduced IDS using LSTM and Fully Connected Network (FCN). this model provides an acceptable result in binary classification and multi classification problems. Almiani et al. [17], introduced multi-layered RNN to improve the detection rate of IDS. Imrana et al [18], used Bi-LSTM to enhance the detection of different attacks. Because BiLSTM can capture more complex patterns in the data, which can improve performance. Ashfaq et al [19], used a hybrid model of CNN and LSTM to detect the temporal features and spatial features from network traffic data and to achieve better performance. Henry et al [20], used a hybrid model of GRU and CNN instead of LSTM to minimize the training time.

4) Deep Belief Networks (DBNs): DBNs are composed of multiple layers of Restricted Boltzmann Machines (RBMs). One advantage of using DBNs in IDS is their ability to automatically learn complex representations of the input data, which can be especially useful in detecting new types of attacks that may not have been previously identified. Zhang et al. [21], combined DBN with an improved Genetic Algorithm (GA). The GA optimizes the DBN's architecture for various attack types, resulting in a compact structure with high detection rates. Yang et al. [22], used DBNs and the Modified Density Peak Clustering Algorithm (MDPCA) for intrusion detection achieving high accuracy and detection rates on the UNSW-NB15 and NSL-KDD datasets.

5) Autoencoders: they are a type of deep learning architecture that are composed of two parts, namely an encoder and a decoder. The encoder compresses the input into a lower-dimensional representation, while the decoder reconstructs the original input using this representation. Autoencoders can also handle high-dimensional data effectively, making them suitable for IDS applications. Khan et al. [23], used a stacked auto-encoder and soft-max classifier to create a two-stage deep learning model for IDS and achieve high recognition rate. Binbusayyis et al. [24], proposed an IDS framework using one-dimensional convolutional autoencoder (1D CAE) to achieve a lower feature representation. Then this representation is utilized in conjunction with a One-Class Support Vector Machine (OCSVM) for the detection of anomalies. Cui et al. [25], proposed an IDS consists of three parts: Stacked Autoencoder (SAE) for feature extraction, Gaussian Mixture Model (GMM) and the Wasserstein Generative Adversarial (GMM-WGAN) for processing class imbalance processing, and CNN-LSTM for classification.

6) Transformers: They are a type of deep learning models that has gained attention in natural language processing (NLP) tasks. It relies on a self-attention mechanism to capture dependencies between different words in a sentence, enabling it to process input sequences in parallel rather than sequentially. It can handle variable-length inputs and offer robustness. In their work, Wu et al. [26] introduced a transformer-based IDS. They used positional embedding to capture the sequential relationships among features, and employ a stacked autoencoder model to extract a compact feature representation from the original data. Farhan et al.

[27], developed IDS using transformer model for deep feature extraction, Synthetic Minority Oversampling Technique (SMOTE) for processing class imbalance, and CNN-LSTM for classifying network attacks. Yao, et al. [28] combined transformer and CNN to detect network attacks. They used XGBoost for feature selection and Adaptive Synthetic (ADASYN) as sampling technique.

7) Generative Adversarial Networks (GANs): They are a one of the deep learning techniques consists of two parts, namely a generator and a discriminator. The generator part is utilized for generating virtual data samples that mimic the characteristics of real network traffic. The discriminator part utilized to differentiate between the real and virtual samples. The two parts are trained simultaneously. Lee et al [29], employed GAN to produce synthetic data that closely resembles the existing data. The proposed model incorporates a random forest algorithm for classification, and its performance is evaluated both with and without addressing data imbalance using GAN. Huang et al [30], presented a new approach called IGAN, which incorporates an imbalanced data filter and convolutional layers into the traditional GAN framework in order to generate novel and representative instances for minority classes.

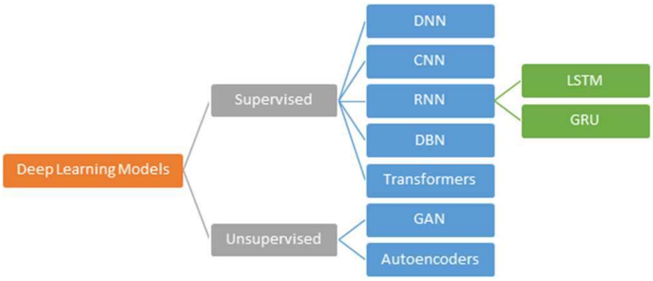

Fig 2. Deep Learning Models Taxonomy

C. Benchmark Datasets

The IDS performance is highly reliant on the quality and relevance of the datasets used for training. Numerous datasets have been used in IDS research. We will explore the latest and most used datasets.

1) KDDCup99: It is driven from the DARPA dataset and consists of 41-dimensional features. These features include duration, protocol type, service, flag, and others statistical features. However, the KDDCup99 dataset has several limitations requiring researchers to carefully filter it before using. It suffers from severe class imbalance, leading to biased classification results favoring the majority classes. It also has redundant and duplicate records.

2) NSL-KDD: It is introduced as an alternative to KDDCup99 dataset to solve its limitations. It addresses the issue of classification bias by carefully selecting instances from the KDD99 dataset and balancing the instances of different classes. Additionally, duplicate and redundant instances were removed, resulting in a dataset with a moderate number of instances. This allows for consistent and comparable results across different papers. While the NSL-KDD dataset improves data redundancy to some extent, it remains outdated as it doesn't incorporate new data and still lacks sufficient samples from the minority class.

3) ISCX-IDS 2012: It is created by the University of New Brunswick, it contains a labeled network traffic and packet payloads, allowing for detailed analysis. The dataset includes network activity for seven days, comprising both malicious and normal traffic. The dataset size is specified for each day, providing a substantial amount of data for researchers. This dataset is particularly useful for simulating user behavior and testing the IDS performance.

4) UNSW-NB15: It is created by the University of South Wales and contains the network traffic that was captured from three virtual servers. It has 49-dimensional features that was extracted the traffic using tool called Bro. This dataset offers more attack types and a greater variety of features compared to the KDD99 dataset. It includes normal traffic and malicious traffic contains nine types of attacks.

5) CIC-IDS2017: It is established by the Canadian Institute for Cybersecurity (CIC) at the University of New Brunswick, it is a widely used benchmark dataset for training and testing IDS. It contains a comprehensive collection of real-world network traffic data, captured in a controlled environment. it has a total of 77 features, which include both packet-level and flow-level features.

6) CSE-CIC-IDS2018: It is established by the Communications Security Establishment (CSE) and Canadian Institute for Cybersecurity (CIC). This dataset contains a comprehensive collection of network traffic data, making it an invaluable resource for training and testing IDS algorithms. One key feature of the CSE-CICIDS2018 dataset is its large number of features. It includes a total of 80 features, which provide detailed information about network traffic flows.

7) NF-UQ-NIDS: Sarhan et al [31], combined four widely used IDS datasets namely, UNSW-NB15, BoT-IoT, ToN-IoT, and CSE-CIC-IDS2018 into novel dataset called NF-UQ-NIDS. They extract 43 features from the raw data using the Cisco NetFlow data collection protocol. The proposed dataset is now accessible, to the research community enabling a fair comparison of deep learning-based IDS across different datasets.

Table II shows the number of features, instances, and attacks in each dataset, as well its issuing year.

TABLE II. BENCHMARK DATASETS

Dataset	Features	Instances	Attacks	Year
KDDCup99	41	5,209,458	4	1999
NSL-KDD	41	148,517	4	2009
ISCX-IDS2012	41	2,450,324	4	2012
UNSW-NB15	49	2,540,038	9	2015
CIC-IDS2017	78	2,830,743	7	2017
CSE-CIC-IDS2018	80	4,525,399	7	2018
NF-UQ-NIDS	43	75,987,976	20	2022

D. Evaluation Metrics

Several metrics are employed to measure the performance of deep learning techniques, and these are used to select the most suitable models. It is common to utilize multiple metrics simultaneously to evaluate the detection effectiveness. First, we define some parameters that are used for the calculation of these metrics. True Positive (TP) means that the intrusion is detected correctly. True Negative (TN) means that the non-intrusive traffic is detected correctly. False Positive (FP) means that the model detects an intrusion when there is none. False Negative (FN) means that the model fails to detect an intrusion

1) Accuracy: It is the number of correctly classified instances divided by the number of all instances.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

2) Precision: It is the number of correctly predicted attacks divided by the number of instances predicted as attacks.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

3) Recall (Detection Rate): It is the number of correctly predicted attacks divided by the number of instances that are actually attacks.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

4) F1-Score: It is defined as the harmonic mean of the Precision and Recall

$$F1 - Score = \frac{2 \times Precsion \times Recall}{Precsion + Recall} \tag{4}$$

5) Missed Alarm Rate: It is the number of undetected attacks divided by the number of instances that are actually attacks. It is also called the False Negative Rate (FNR).

$$FNR = \frac{FN}{TP+FN} \tag{5}$$

6) False Alarm Rate: It is the number of normal instances that are predicted as attacks divided by the number of instances predicted as attacks. It is also called the False Positive Rate (FPR).

$$FPR = \frac{FP}{TP+FP} \tag{6}$$

7) Time Complexity: It is defined as the amount of time taken by the model to process and analyze input data. Time complexity is influenced by factors like the size and complexity of the input data, the architecture of the deep learning model, the number of layers and nodes, and the efficiency of the training and inference processes.

V. Discussion

This section aims at discussing the different methods and techniques utilized in IDS as described in the above section. The outcomes of this section can be valuable in emphasizing the challenges and future directions.

A. Methods

Through an extensive review of 50 deep learning approaches discussed in the previous section, we identified two primary categories of techniques that researchers have utilized: single methods and hybrid methods. As illustrated in Figure 3. CNN are the most used in IDS research, because of their ability to learn features automatically from the data, rather than relying on handcrafted features. This can be especially useful in detecting new types of attacks that may not have been previously identified. Another advantage is the ability of CNNs to handle variable length inputs. In network traffic, the length of packets can vary significantly, and CNNs can be designed to handle this variability by using techniques such as 1D convolutions and pooling layers. Autoencoders are utilized for extracting features and dimensionality reduction as an alternative of traditional methods such as Principal Component Analysis (PCA). One of the advantages of using autoencoders in IDS is their ability to learn representations of the input data that are robust to noise and variability. In network traffic, packets may be dropped or corrupted, leading to missing or noisy data. Autoencoders can be designed to handle this variability by using techniques such as denoising autoencoders. LSTM are well-suited for capturing sequential dependencies in network traffic. They can effectively simulate the temporal nature of network activities and detect anomalies based on historical patterns. Methods such as DNN, DBN are also used for detecting attacks in network traffic, because of the ability to learn complex representations of the input data.

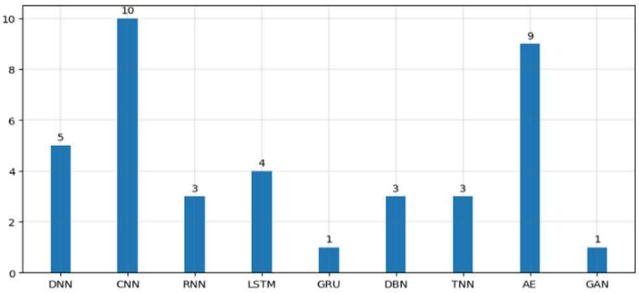

Fig 3. Deep Learning Models

In the recent studies GANs have been used in IDS to create new samples that is similar to those existing in dataset instead of methods such as ADASYN and SMOTE. On the other hand, Transformer Networks have been used in classifying network attacks. As they analyze event sequences, considering temporal dependencies and learning complex relationships. This property makes them well-suited for IDS, where network traffic data often exhibits sequential and temporal characteristics.

Several studies have been employed hybrid architecture for IDS as shown in Figure 4. CNN-LSTM scheme is used due to its effectiveness in capturing both spatial and temporal dependencies in sequential data. CNN-GRU is used to enhance the performance, reduce the model complexity and provide faster convergence in training. In AE-LSTM scheme LSTM is used to process the compressed representations generated by the autoencoder. This makes the model learn temporal patterns and dependencies within the network traffic data, enhancing the detection of time-sensitive attacks that span across multiple packets or network flows. Architectures such as AE-CNN, AE-DNN, and DNN-GAN have been used to leverage the respective strengths of these techniques and attain outstanding performance.

B. Datasets and Evaluation

Benchmark datasets are crucial components utilized to measure the performance of the DL approaches. Figure 5 shows the most frequently used datasets in the discussed studies. It is illustrated that 70% of these studies used NSL-KDD, UNSW-NB15 and CIC-IDS2017 datasets for testing and validating processes. So, we will discuss these datasets with some details, focusing on the size of datasets, the number of features, attacks type, models accuracy, and the features importance using Random Forest Classifier.

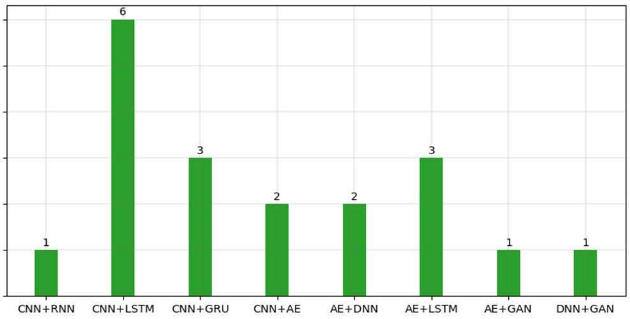

Fig 4. Hybrid Deep Learning Models

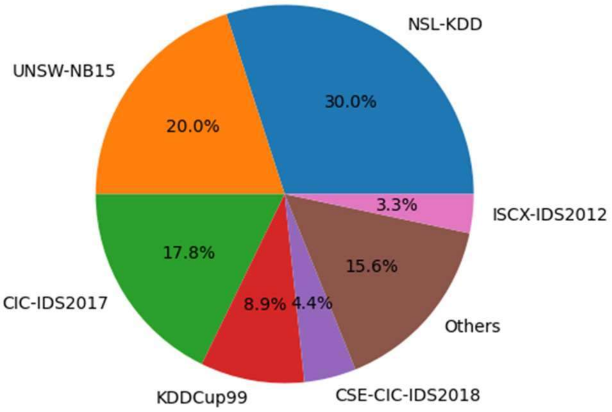

Fig 5. IDS Benchmark Datasets

Approximately 30% of the previous studies have utilized the NSL-KDD dataset. However, it is important to know that this dataset is considered outdated as it is based on the KDDCup99 dataset. NSL-KDD comprises a total of 148,517 instances, with 77,054 classified as normal and 71,463 classified as malicious. The dataset encompasses four types of attacks which are Probing, Denial of Service, User to Root, and Root to Local. Figure 6 illustrates that the dataset contains 41 features, with only 26 of these features demonstrating significant importance in classifying attacks. As shown in Figure 7, several studies such as [21], [33], [35] have employed the NSL-KDD dataset to achieve accuracy exceeding 99%. However, this dataset has a limited number of features and relatively fewer instances. Additionally, it suffers from class imbalance. As a result, any deep learning model can easily fit the NSL-KDD dataset and attain high accuracy.

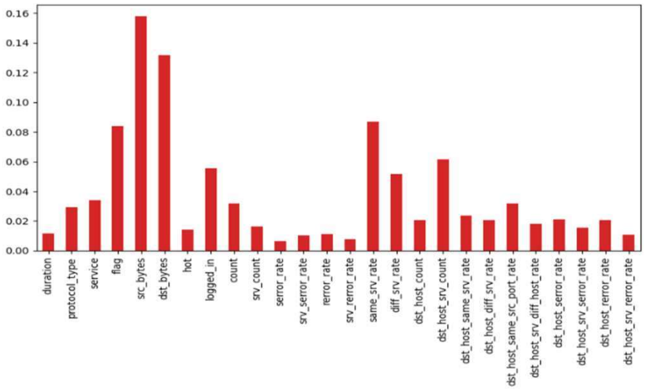

Fig 6. NSL-KDD Features Importance

UNSW-NB15 dataset has been utilized in approximately 20% of the previous studies. It comprises a large number of instances, specifically 2,540,038, with 2,218,755 classified as normal and 321,283 classified as malicious. This dataset encompasses nine types of attacks, namely Fuzzers, Backdoor, Analysis, Denial of Service, Exploits, Reconnaissance, Generic, Shellcode, and Worms. Figure 8 illustrates that the dataset contains a total of 49 features, of which 35 features exhibit high importance in classifying attacks. According to Figure 9, the accuracy of models used UNSW-NB15 dataset for training, varies from 84% to 99%. This wide range can be attributed to the dataset's large number of instances and large number of features, relative to the NSL-KDD dataset. Thus, it has proven to be a favorable choice for training and testing deep learning models.

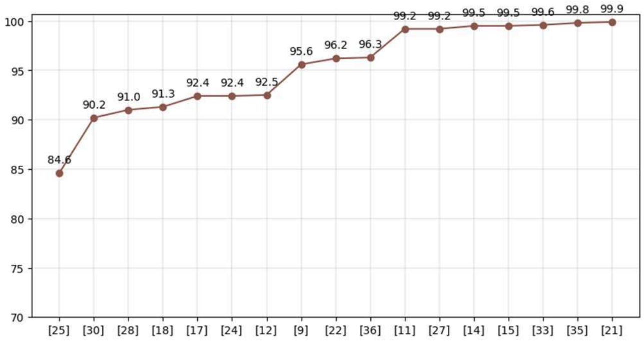

Fig 7. Accuracy of NSL-KDD based DL Models

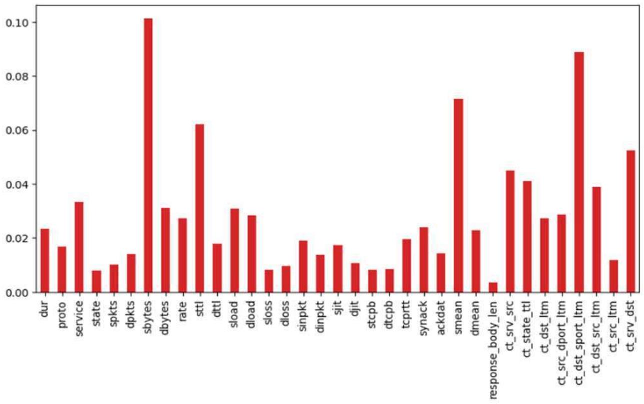

Fig 8. UNSW-NB15 Features Importance

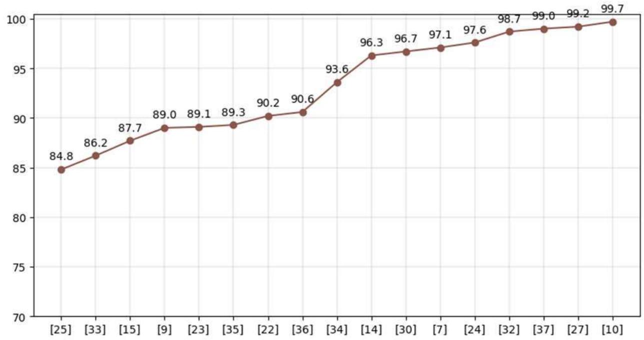

Fig 9. Accuracy of UNSW-NB15 based DL Models

CIC-IDS2017 dataset has been utilized in approximately 18% of the previous studies. It consists of a large number of instances, specifically 2,830,743, with 2,273,097 classified as normal and 557,646 classified as malicious. This dataset encompasses seven types of attacks which are Denial of

Service, Brute force, Web Attacks, Botnet, Port scan, and Infiltration. Figure 10 reveals that the dataset comprises a total of 78 features, of which 32 features demonstrate high importance in classifying attacks. As shown in Figure 11 several studies such as [11], [32], [35] have employed the CIC-IDS2017 dataset and achieved accuracy levels exceeding 99%. This remarkable accuracy can be attributed to the dataset's large number of instances with diverse attack types, making it suitable for training deep learning models.

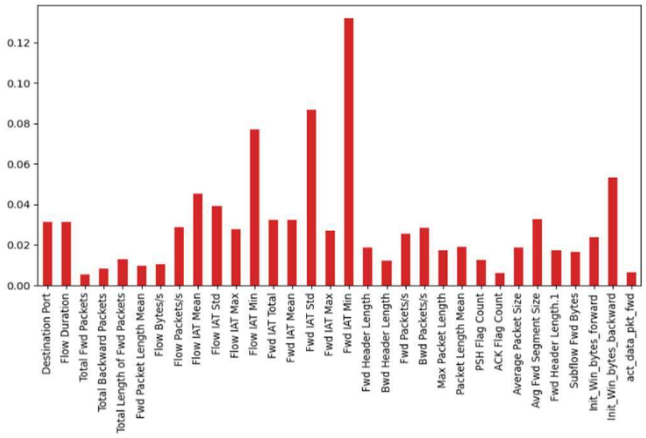

Fig 10. CIC-IDS2017 Features Importance

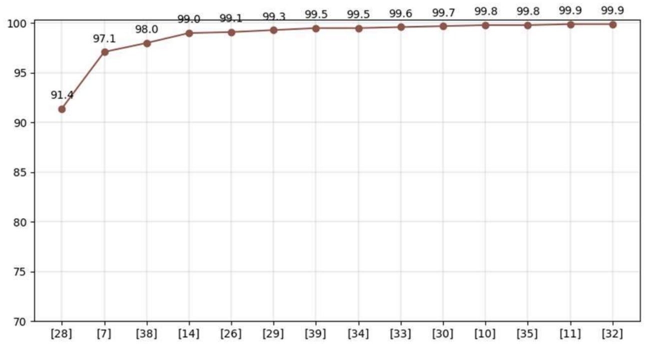

Fig 11. Accuracy of CIC-IDS2017 based DL Models

To improve the DL models generalization, studies such as [14], [30] used multiple datasets during training and testing. This allows researchers to capture a wide range of network activities and attack scenarios. Each dataset may have different characteristics, such as different types of attacks, network configurations, or traffic patterns by using diverse datasets, researchers can ensure that their DL models are trained on a variety of scenarios and are robust enough to handle different types of attacks. This also makes it possible to compare the performance of different DL models or algorithms.

Figure 12, shows the evaluation metrics used by the researchers. the Detection Accuracy is the most widely used. But the accuracy alone may not be a sufficient metric to judge the DL models performance, especially when dealing with imbalanced data. Accuracy can be misleading when the dataset is skewed towards one class. When the number of normal instances is much larger than the number of attack instances the model that simply labels everything as normal can achieve high accuracy, but it fails to detect attacks effectively. To overcome this limitation, researchers should use additional evaluation metrics like F1-score, detection rate (recall), and precision. By considering these metrics, researchers can get more comprehensive understandings of a DL model's performance in detecting attacks. It is also essential to consider the time complexity of DL models when evaluating their performance, as it can impact their practical usability in real-world scenarios.

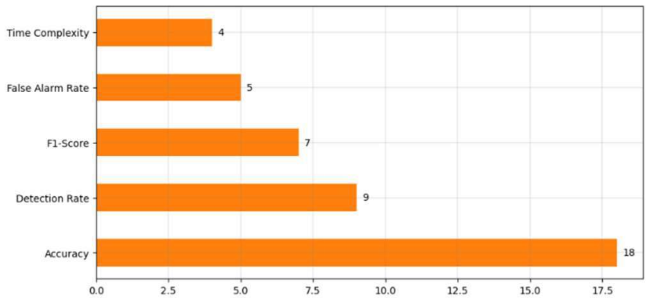

Fig 12. Deep Learning Models Evaluation Metrics

C. Challenges and Future Directions

- Data quality and availability: The process of training and testing DL models requires large amounts of high-quality labeled data. However, obtaining labeled intrusion data is challenging due to privacy concerns and the rarity of real-world attacks. Generating synthetic data or utilizing transfer learning techniques can help mitigate this challenge.

- Imbalanced datasets: Intrusion data is often imbalanced, with a majority of instances belonging to the normal class. DL models tend to struggle with unbalanced data as they can become skewed towards the majority class. Techniques such as oversampling, undersampling, or using cost-sensitive learning algorithms can address this challenge.

- Explainability and interpretability: DL models are commonly perceived as black boxes, making it difficult to comprehending their decision-making process. Research is focused on developing explainable DL models that can provide insights into how and why a model makes certain predictions.

- Adversarial attacks: DL models are susceptible to adversarial attacks, which involve manipulating input data in order to deceive the model and evade detection. Adversarial training and robustness techniques are being explored to enhance the resilience of DL-based IDS against such attacks.

- Real-time processing: DL models require significant computational resources for real-time intrusion detection. Optimizing model architectures, leveraging hardware accelerators, and developing efficient algorithms are crucial for achieving real-time performance.

- Hybrid models: Combining DL techniques with traditional rule-based or signature-based IDS can leverage the strengths of both approaches. Hybrid models can enhance detection accuracy, reduce false positives, and provide better coverage against known and unknown attacks.

- Transfer learning and domain adaptation: Transferring knowledge from pre-trained models in related domains or adapting models to different network environments can improve the performance of DL-based IDS. This approach can help address the challenge of limited labeled intrusion data.

• Online learning and incremental training: DL models that can adapt and learn continuously from streaming data are essential for dynamic network environments. Developing online learning algorithms that can update the model in real-time without the requirement of retraining the entire model from scratch.is an important future direction.

• Ensemble methods: Combining multiple DL models or different DL architectures can enhance the overall detection performance and robustness of the IDS. Ensemble methods can mitigate the influence of individual model biases and enhance the generalization capabilities of the system.

VI. CONCLUSION

This comprehensive study conducted a systematic literature review of publications over the past five years, delving into the utilization of DL techniques in IDS and examining the datasets commonly used in the field. The analysis revealed the widespread adoption of DL techniques like CNNs, LSTM, and Autoencoders, which have demonstrated promising results in enhancing IDS performance. Recent studies have also explored the use of Transformers for attack classification and GANs for generating synthetic data to address imbalanced datasets. Through our study, we notice that several studies relied on outdated datasets with a limited number of instances to train their models, resulting in artificially inflated accuracy measures. Additionally, a notable portion of the reviewed studies focused primarily on achieving high accuracy without considering the associated time complexity. We present the challenges posed by limited labeled datasets, imbalanced data, and the interpretability of DL models, so that, future research should prioritize the addressing of these issues. Additionally, exploring techniques such as transfer learning and anomaly detection can further enhance IDS performance.

REFERENCES

[1] Z. Yang, X. Liu, T. Li, D. Wu, J. Wang, Y. Zhao and H. Han, "A systematic literature review of methods and datasets for anomaly-based network intrusion detection," *Computers & Security,* vol. 116, p. 102675, 2022.

[2] N. Moustafa, J. Hu and J. Slay, "A holistic review of network anomaly detection systems: A comprehensive survey," *Journal of Network and Computer Applications,* vol. 128, p. 33–55, 2019.

[3] A. M. Aleesa, B. B. Zaidan, A. A. Zaidan and N. M. Sahar, "Review of intrusion detection systems based on deep learning techniques: coherent taxonomy, challenges, motivations, recommendations, substantial analysis and future directions," *Neural Computing and Applications,* vol. 32, p. 9827–9858, 2020.

[4] J. Lansky, S. Ali, M. Mohammadi, M. K. Majeed, S. H. T. Karim, S. Rashidi, M. Hosseinzadeh and A. M. Rahmani, "Deep learning-based intrusion detection systems: a systematic review," *IEEE Access,* vol. 9, p. 101574–101599, 2021.

[5] H. Liu and B. Lang, "Machine learning and deep learning methods for intrusion detection systems: A survey," *applied sciences,* vol. 9, p. 4396, 2019.

[6] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey and S. Linkman, "Systematic literature reviews in software engineering–a systematic literature review," *Information and software technology,* vol. 51, p. 7–15, 2009.

[7] O. Faker and E. Dogdu, "Intrusion detection using big data and deep learning techniques," in *Proceedings of the 2019 ACM Southeast conference*, 2019.

[8] R. M. Swarna Priya, P. K. R. Maddikunta, M. Parimala, S. Koppu, T. R. Gadekallu, C. L. Chowdhary and M. Alazab, "An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture," *Computer Communications,* vol. 160, p. 139–149, 2020.

[9] Y. Yang, K. Zheng, C. Wu and Y. Yang, "Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network," *Sensors,* vol. 19, p. 2528, 2019.

[10] H. Zhang, L. Huang, C. Q. Wu and Z. Li, "An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset," *Computer Networks,* vol. 177, p. 107315, 2020.

[11] A. Dahou, M. Abd Elaziz, S. A. Chelloug, M. A. Awadallah, M. A. Al-Betar, M. A. A. Al-Qaness and A. Forestiero, "Intrusion detection system for IoT based on deep learning and modified reptile search algorithm," *Computational Intelligence and Neuroscience,* vol. 2022, 2022.

[12] J. Yu, X. Ye and H. Li, "A high precision intrusion detection system for network security communication based on multi-scale convolutional neural network," *Future Generation Computer Systems,* vol. 129, p. 399–406, 2022.

[13] Y. Li, Y. Xu, Z. Liu, H. Hou, Y. Zheng, Y. Xin, Y. Zhao and L. Cui, "Robust detection for network intrusion of industrial IoT based on multi-CNN fusion," *Measurement,* vol. 154, p. 107450, 2020.

[14] V. Ravi, R. Chaganti and M. Alazab, "Recurrent deep learning-based feature fusion ensemble meta-classifier approach for intelligent network intrusion detection system," *Computers and Electrical Engineering,* vol. 102, p. 108156, 2022.

[15] S. M. Kasongo, "A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework," *Computer Communications,* vol. 199, p. 113–125, 2023.

[16] S. K. Sahu, D. P. Mohapatra, J. K. Rout, K. S. Sahoo, Q.-V. Pham and N.-N. Dao, "A LSTM-FCNN based multi-class intrusion detection using scalable framework," *Computers and Electrical Engineering,* vol. 99, p. 107720, 2022.

[17] M. Almiani, A. AbuGhazleh, A. Al-Rahayfeh, S. Atiewi and A. Razaque, "Deep recurrent neural network for IoT intrusion detection system," *Simulation Modelling Practice and Theory,* vol. 101, p. 102031, 2020.

[18] Y. Imrana, Y. Xiang, L. Ali and Z. Abdul-Rauf, "A bidirectional LSTM deep learning approach for intrusion detection," *Expert Systems with Applications,* vol. 185, p. 115524, 2021.

[19] M. A. Khan, M. R. Karim and Y. Kim, "A scalable and hybrid intrusion detection system based on the convolutional-LSTM network," *Symmetry,* vol. 11, p. 583, 2019.

[20] A. Henry, S. Gautam, S. Khanna, K. Rabie, T. Shongwe, P. Bhattacharya, B. Sharma and S. Chowdhury, "Composition of hybrid deep learning model and feature optimization for intrusion detection system," *Sensors,* vol. 23, p. 890, 2023.

[21] Y. Zhang, P. Li and X. Wang, "Intrusion detection for IoT based on improved genetic algorithm and deep belief network," *IEEE Access,* vol. 7, p. 31711–31722, 2019.

[22] N. Balakrishnan, A. Rajendran, D. Pelusi and V. Ponnusamy, "Deep Belief Network enhanced intrusion detection system to prevent security breach in the Internet of Things," *Internet of things,* vol. 14, p. 100112, 2021.

[23] F. A. Khan, A. Gumaei, A. Derhab and A. Hussain, "A novel two-stage deep learning model for efficient network intrusion detection," *IEEE Access,* vol. 7, p. 30373–30385, 2019.

[24] A. Binbusayyis and T. Vaiyapuri, "Unsupervised deep learning approach for network intrusion detection combining convolutional autoencoder and one-class SVM," *Applied Intelligence,* vol. 51, p. 7094–7108, 2021.

[25] J. Cui, L. Zong, J. Xie and M. Tang, "A novel multi-module integrated intrusion detection system for high-dimensional imbalanced data," *Applied Intelligence,* vol. 53, p. 272–288, 2023.

[26] Z. Wu, H. Zhang, P. Wang and Z. Sun, "RTIDS: A robust transformer-based approach for intrusion detection system," *IEEE Access,* vol. 10, p. 64375–64387, 2022.

[27] F. Ullah, S. Ullah, G. Srivastava and J. C.-W. Lin, "IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic," *Digital Communications and Networks,* 2023.

[28] R. Yao, N. Wang, P. Chen, D. Ma and X. Sheng, "A CNN-transformer hybrid approach for an intrusion detection system in advanced metering infrastructure," *Multimedia Tools and Applications,* vol. 82, p. 19463–19486, 2023.

[29] J. Lee and K. Park, "GAN-based imbalanced data intrusion detection system," *Personal and Ubiquitous Computing,* vol. 25, p. 121–128, 2021.

[30] S. Huang and K. Lei, "IGAN-IDS: An imbalanced generative adversarial network towards intrusion detection system in ad-hoc networks," *Ad Hoc Networks,* vol. 105, p. 102177, 2020.

[31] M. Sarhan, S. Layeghy and M. Portmann, "Towards a standard feature set for network intrusion detection system datasets," *Mobile networks and applications,* p. 1–14, 2022.

[32] A. Basati and M. M. Faghih, "APAE: an IoT intrusion detection system using asymmetric parallel auto-encoder," *Neural Computing and Applications,* vol. 35, p. 4813–4833, 2023.

[33] B. Cao, C. Li, Y. Song, Y. Qin and C. Chen, "Network intrusion detection model based on CNN and GRU," *Applied Sciences,* vol. 12, p. 4184, 2022.

[34] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi and R. Ahmad, "CNN-LSTM: hybrid deep neural network for network intrusion detection system," *IEEE Access,* vol. 10, p. 99837–99849, 2022.

[35] G. Andresini, A. Appice and D. Malerba, "Autoencoder-based deep metric learning for network intrusion detection," *Information Sciences,* vol. 569, p. 706–727, 2021.

[36] P. R. Kanna and P. Santhi, "Unified deep learning approach for efficient intrusion detection system using integrated spatial–temporal features," *Knowledge-Based Systems,* vol. 226, p. 107132, 2021.

[37] Y. Yang, S. Tu, R. H. Ali, H. Alasmary, M. Waqas and M. N. Amjad, "Intrusion detection based on bidirectional long short-term memory with attention mechanism," 2023.

[38] R. V. Mendonça, A. A. M. Teodoro, R. L. Rosa, M. Saadi, D. C. Melgarejo, P. H. J. Nardelli and D. Z. Rodríguez, "Intrusion detection system based on fast hierarchical deep convolutional neural network," *IEEE Access,* vol. 9, p. 61024–61034, 2021.

[39] C. Ieracitano, A. Adeel, F. C. Morabito and A. Hussain, "A novel statistical analysis and autoencoder driven intelligent intrusion detection approach," Neurocomputing, vol. 387, p. 51–62, 2020.