Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

An innovative network intrusion detection system (NIDS): Hierarchical deep learning model based on Unsw-Nb15 dataset

Mohammad A. Alsharaiah[a], Ahmad Adel Abu-Shareha[a*], Mosleh Abualhaj[b], Laith H. Baniata[c], Adeeb Al-saaidah[b], Qasem M. Kharma[d] and Mahran M Al-Zyoud[b]

[a]*Department of Data Science and Artificial Intelligence, Al-Ahliyya Amman, Jordan*
[b]*Department of Networks and Cybersecurity, Al-Ahliyya Amman University, Jordan*
[c]*Department of computer Science, Al-Ahliyya Amman University, Jordan*
[d]*Department of Software Engineering, Al-Ahliyya Amman University, Jordan*

C H R O N I C L E

Article history:
Received: November 7, 2023
Received in revised format: November 30, 2023
Accepted: January 8, 2023
Available online: January 8, 2024

Keywords:
UNSW-NB15
Classification
Machine learning
Deep learning
LSTM attention

A B S T R A C T

With the increasing prevalence of network intrusions, the development of effective network intrusion detection systems (NIDS) has become crucial. In this study, we propose a novel NIDS approach that combines the power of long short-term memory (LSTM) and attention mechanisms to analyze the spatial and temporal features of network traffic data. We utilize the benchmark UNSW-NB15 dataset, which exhibits a diverse distribution of patterns, including a significant disparity in the size of the training and testing sets. Unlike traditional machine learning techniques like support vector machines (SVM) and k-nearest neighbors (KNN) that often struggle with limited feature sets and lower accuracy, our proposed model overcomes these limitations. Notably, existing models applied to this dataset typically require manual feature selection and extraction, which can be time-consuming and less precise. In contrast, our model achieves superior results in binary classification by leveraging the advantages of LSTM and attention mechanisms. Through extensive experiments and evaluations with state-of-the-art ML/DL models, we demonstrate the effectiveness and superiority of our proposed approach. Our findings highlight the potential of combining LSTM and attention mechanisms for enhanced network intrusion detection.

1. Introduction

The widespread adoption of internet technologies, particularly with the integration of cloud computing, has led to a significant increase in intrusion incidents. Major platforms like Google and Amazon host numerous servers and provide services to various enterprises, making them attractive targets for malicious activities. Consequently, organizations face escalating costs for implementing security methods such as firewalls to protect their data and ensure uninterrupted services. Failure to detect and mitigate intrusions can have severe consequences for an organization's reputation and data integrity (Vaswani, 2017). Network Intrusion Detection Systems (NIDS) has a critical role in safeguarding networks against malicious activities (Chen & Guestrin, 2016; Wang, 2019). These systems, which can be hardware or software-based, are designed to monitor and detect any unauthorized or malicious network traffic. NIDS operate by passively monitoring network traffic, analyzing it for suspicious patterns, and comparing it against known attack signatures (Cisco, 2003). By strategically deploying NIDS at key points within a network, such as at the network medium or on specific devices, comprehensive traffic monitoring can be achieved. NIDS perform thorough analysis of all passing network traffic within a subnet and compare it against a predefined library of attack signatures. In the event of detecting abnormal behavior or identifying an attack, NIDS generates alerts for network administrators to investigate and take appropriate action (Pandya, 2013). The advantage of NIDS lies in its ability to detect

* Corresponding author.
E-mail address: m.sharaiah@ammanu.edu.jo (A. A. Abu-Shareha)

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2024.1.007

and respond to patterns shared across multiple hosts, allowing for early identification and prevention of attacks before they reach their intended targets. Thus, NIDS serves as an essential defense mechanism by continuously monitoring network packets and promptly alerting administrators to potential security breaches. However, it is important to note that NIDS face certain challenges and limitations. These include the need for continuous updates to attack signature databases to stay ahead of evolving threats, the potential for false positives or false negatives in intrusion detection, and the scalability of NIDS to handle high-volume network traffic. Addressing these challenges requires the development of advanced detection techniques, harnessing the potential of machine learning and deep learning algorithms to increase the precision and efficiency of intrusion detection systems. Taking these factors into account, the goal of this investigation is to develop a promising NIDS model that combines long short-term memory (LSTM) with an attention layer. By leveraging the spatial and temporal features of the benchmark UNSW-NB15 dataset, the proposed model seeks to overcome the limitations of existing ML/DL models, such as SVM and KNN, which often work with limited features and lower accuracy. The effectiveness of the proposed model will be evaluated through extensive experimentation and comparison with other ML/DL models, providing valuable insights into its benefits and limitations in the field of network intrusion detection.

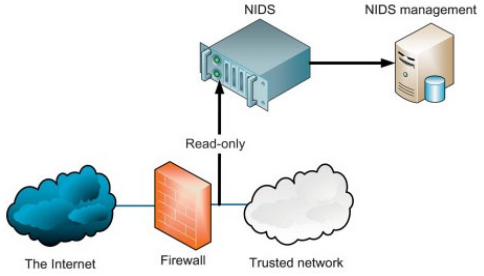

Fig. 1. Typical NIDS architecture

However, Machine learning and deep learning become a promising selection to develop a secure tool to protect data and develop trusted NIDs. Herein, we proposed a productive novel NIDS by merging different deep learning techniques in one model, it involves (LSTM) with an attention layer to detect the abnormal data packet network. Besides, this research concerned binary classification and the result is compared with other benchmark Machine learning models. Also, this research employs a very large dataset that has a high dimension feature space. In this research, the feature engineering process has not been used, this means no drop for any feature form the data set has been performed. The training and testing for the proposed model used all the features set even though it was time consuming. This research work uses the most popular public dataset UNSW-NB15 (The UNSW-NB15 Dataset, n.d.). Mainly, the University of South Wales released this dataset in 2015. It identifies the limitation of KDD98 and KDD99 datasets since they cannot include modern low footprint attacks (Moustafa & Slay, 2015). The improved form of KDD cup dataset was used, and it assisted us to analyze and distinguish the features for the untacked network or the attacked network. Essentially, the UNSW-NB15 dataset comprises several classes, and is termed as follows: Normal, Exploits, DOS, Back doors, analysis, Fuzzers, generic, shell code, Reconnaissance and worms. But, herein, we measured the binary version of the data set, the 0 for the regular class while 1 for the attacked class. Furthermore, in comparison with the models in literature (Sultana et al., 2019), the proposed model has been developed without feature engineering since our proposed model detention the features automatically while other models performed the feature engineering before developing the model to intensify the accuracy performance level in prediction.

Additionally, the proposed NIDS model aims to address the common practice of manual feature selection and extraction in existing models applied to the UNSW-NB15 dataset. Many previous models require a pre-manual stage to identify relevant features, which can be a time-consuming and error-prone process. In contrast, the proposed model leverages coupling the capabilities of LSTM, a recurrent neural network variety renowned for its proficiency in capturing extended contextual relationships in sequential information and integrating an attention layer to adaptively acquire and highlight critical features within the dataset. The benchmark UNSW-NB15 dataset, which contains a diverse range of real and simulated network traffic, will be used to train and test the proposed NIDS model. The dataset comprises diverse attack categories, such as DoS, DDoS, probe, and generic, making it well-suited for evaluating the effectiveness of intrusion detection models. The dataset consists of 49 features categorized under different types, including flow, content, time, basic, general purpose, and connection. To ensure the model's performance and generalizability, several data preprocessing steps will be applied. These steps include data normalization to scale the values of features within a relative range, conversion of symbolic data to numerical values, and extraction of normal data samples for training. Clustering will be applied to the training data to decrease the quantity of training records and enhance the pace of training. The evaluation of the proposed model will primarily focus on binary classification, determining whether an attack is occurring or not, as well as accurately classifying the category of the attack. The model effectiveness should be evaluated by using performance measures like accuracy, which are consequent from true positive (TP), true negative (TN), also, the false positive (FP), and false negative (FN) standards. Furthermore, we will employ the confusion matrix, offering a comprehensive breakdown of the model's predictions, to assess its performance in various

classes. Through extensive experiments and comparisons with existing ML/DL models, the proposed NIDS model aims to demonstrate superior performance in terms of accuracy and overall intrusion detection capability. The results and discussions will shed light on the advantages and limitations of the proposed model, providing valuable insights for improving network intrusion detection systems and enhancing cybersecurity measures in the face of evolving threats.

The aim of this study is to advance the precision of network intrusion detection systems (NIDS) within the domain of cybersecurity. The specific goal is to propose a novel model that combines (LSTM) with an attention layer for binary classification of network traffic data. The research strives to tackle the constraints found in current machine learning and deep learning models applied in NIDS, including SVM and KNN, which frequently operate with restricted features and demonstrate reduced accuracy. Moreover, numerous prevalent models employed with the standard UNSW-NB15 dataset demand the manual selection and extraction of features, a process that can be time-intensive and might not produce highly accurate outcomes. To attain the research goal, the study makes use of the UNSW-NB15 dataset, housing a wide variety of authentic and simulated network traffic behaviors. The dataset consists of various categories of traffic, including normal behavior and different types of attacks. The model in question places its emphasis on acquiring insights from the spatial and temporal attributes within the dataset, achieved by integrating LSTM, renowned for its ability to capture enduring correlations in sequential data. The attention layer is introduced to dynamically select important features and assign higher weights to relevant information. By leveraging these techniques, the model aims to improve the accuracy of binary classification, distinguishing between normal network traffic and malicious activities. The research methodology includes data exploration, preprocessing, model development, training, and evaluation. The UNSW-NB15 dataset is thoroughly examined to understand its characteristics and categories. Preprocessing steps involve data normalization, conversion of symbolic data to numeric values, and extraction of normal data samples. It consists of several steps to achieve the objective of proposing an enhanced network intrusion detection system (NIDS) model. The methodology encompasses data exploration, preprocessing, model development, training, and evaluation. The comprehensive research methodology is outlined as follows:

Data Investigation: The investigation commences with an in-depth examination of the UNSW-NB15 dataset, a recognized popular dataset commonly employed in the field of network intrusion detection. This involves analyzing the dataset's characteristics, understanding its categories, and gaining insights into the nature of the network traffic data.

Data Preprocessing: The dataset undergoes preprocessing steps to prepare it for model development. This includes removing redundant and duplicate records from the selected training sets. Data normalization is applied to scale the values of each feature within a relative range, avoiding biases towards features with higher values. Symbolic data is converted into numeric representations. Furthermore, the dataset is segregated into training and testing subsets, utilizing a designated division ratio. Model Development: The proposed model architecture is developed, which combines (LSTM) with an attention layer. LSTM is employed to capture temporal relationships and patterns within the sequential network traffic data. The attention layer dynamically selects important features and assigns higher weights to relevant information, enhancing the model's ability to learn from spatial and temporal features.

Training: The developed model is trained using the training dataset from the UNSW-NB15 dataset. The training process involves optimizing the model's parameters using an optimization procedure, such as the Adam optimizer. The learning rate and momentum are set accordingly. The batch size, number of LSTM units, and attention dimensions are specified for the training process.

Evaluation: The trained model is evaluated using the testing dataset from the UNSW-NB15 dataset. The model's performance is assessed by metrics like accuracy, which measures the proportion of correctly classified instances. Additionally, other evaluation metrics like the confusion matrix are employed to provide further insights Assessing the model's efficacy, encompassing elements such as correct positives, correct negatives, erroneous positives, and erroneous negatives.

Contrasting with Established Models: The performance of the suggested model is evaluated in relation to other machine learning (ML) and deep learning (DL) models that have previously been utilized with the UNSW-NB15 dataset. This evaluation is rooted in accuracy and possibly other pertinent criteria, highlighting the advancements brought about by the proposed model.

Experimental Results: The experimental results acquired from the proposed model are analyzed and presented. The accuracy scores, performance metrics, and comparisons with Alternative models are considered to confirm the efficacy of the suggested method in augmenting the precision of NIDS.

The model's efficiency is assessed through extensive experiments, comparing it with other ML and DL models applied to the same dataset. The main metric used for evaluation is accuracy, which measures the proportion of correctly classified instances. Additional evaluation metrics, such as confusion matrix analysis, provide insights into the model's prediction performance.

The paper adheres to an organized layout, consisting of multiple distinct sections. Section 2 provides the related work, while section three discusses the main work details the methodology employed, including the architectural design and preprocessing steps applied to the benchmark UNSW-NB15 dataset. The dataset description presents pertinent information about the UNSW-NB15 dataset, such as its categories and distribution. Data set preprocessing elucidates the techniques used to

normalize and convert the data, ensuring its suitability for analysis. The proposed model basics and architecture section delves into the intricate components of LSTM and attention models, highlighting their synergistic integration. Outcomes and deliberations present the empirical results, elucidating the model's performance through comparisons with established ML/DL methods. Additionally, this section offers an inclusive analysis of the proposed model's benefits and limitations, facilitating a deeper understanding of its efficacy. Finally, the conclusion section encapsulates the key findings, reiterates the research contributions, and outlines future research prospects in the domain of NIDS.

2. Related work

Many techniques have been engaged for (NIDS) to classify either the normal or the abnormal network data packets. For instance, the NIDS that is based on classical machine learning utilized a special schema such as kernel machines and ensemble for classification (Panda et al., 2011; Fu, 2017). Further, (SVM) is a common technique for classification, and SVM employs the kernel machine and kernel Gaussian (Ahmad et al., 2018). The kernel helps the SVM to deal with nonlinear datasets. Kernel projects the data into upper dimensional space and marks it separable. On the other hand, ensemble classifiers may construct a set of weak classifiers in one to avoid over-fitting through the training process. This method provides robust classification functionality such as random forest (Zhang et al., 2008) and adaptive boosting (Hu et al., 2013). However, the mentioned classifier methods encounter a limitation in handling extensive datasets due to their lack of scalability. Additionally, they seldom achieve a perfect balance between validation accuracy and the size of the training dataset. Classical machine learning approaches possess another constraint, which is their capacity to generalize features when utilizing raw data. They learn from the inputs solely based on the set of features provided, and their learning effectiveness relies on the specific features chosen. Machine learning techniques prioritize learning the significance of features, while feature availability methods also emphasize the use of dimensionality reduction techniques to identify the most optimal correlations among the dataset's features. Besides, it predicts the best result with finest time steps. Therefore, this leads to choosing the deep learning approaches to cross over the drawback. Recently, artificial neural networks (NN) have become an alternative in developing the NIDS. NN organizes the learning algorithm inside a group of hidden layers to learn and make artificial decisions. For illustration, an initial class of the feed forward deep neural network termed as multi-layer perceptron (MLP) has been employed on the NIDS fields (Pal & Mitra, 1992). This type reduces the error rate by involving the backpropagation algorithm over the training process. Another form of NN like convolutional neural network (CNN) becomes a successful method for system modeling due to the capability to mining the features from the raw data (LeCun et al., 1998; Szegedy et al., 2015; He et al., 2016). CNN introduced a feature map to manifest the spatial relations inside the dataset. However, in case the dataset has a long range of dependency, then CNN cannot deal with it. Recurrent neural networks (RNNs) are employed in the construction of NIDS to overcome the constraints observed in CNNs since they can extract temporal features from the input data. Moreover, the (LSTM) is a frequently used type of RNN (Hochreiter & Schmidhuber, 1997). LSTM saves the tendency for the long-term relationships in the sequential data and drops the noise that appeared in the short-lived temporal. Recently, hierarchical models have been utilized to learn the spatial and the temporal features that are available in network traffic data. For instance, prediction models that are related to network time series data are popular ones. This type of data set has a nonlinear specificity because several data points are changed over time, this causes an irregular fluctuation. ML such as SVM, k –nearest neighbor and naïve bays have engaged to develop a NIDS's (Biswas, 2018; Gao et al., 2019). Besides, these statistical methods do not involve mutual relations that occur between data and depend on feature selection. As result, for the real time practice they become ineffective even with smaller detection rates. Recently, deep learning methods are used in developing the NIDS's Such as RNN - Recurrent Neural Network, CNN- Convolutional neural network, etc. However, for practical practice these models are still under research because of their high false positive rate (Bengio et al., 2005). To be precise, the model must be trained over the data set with features that define the usual behavior for the network. In general, the label would be 0 for the normal and set as 1 for the attack. Although certain datasets classify the attack through several subdivisions, in this case the classification becomes time consuming with less accuracy. The literature provides several machine-learning models for classification. Bayesian model Michalas and Murray (2017) implement a special Bayes theorem to train the data to generate a classification model. Furthermore, neural networks NN were used.

Sarnovsky and Paralic (2020) present a study titled “Hierarchical Intrusion Detection Using Machine Learning and Knowledge Model”, in which they introduce a hierarchical intrusion detection system (IDS) that integrates machine learning techniques with knowledge-based methods to identify and categorize network intrusions. This IDS employs a multi-stage hierarchical prediction strategy to differentiate between regular network connections and attacks, as well as predict the specific classes and types of attacks. To assess its performance, the proposed IDS was tested on the KDD 99 dataset and benchmarked against similar methodologies. Han and Pak (2023) centered on elevating the efficiency of network intrusion detection systems (NIDSs) by leveraging comprehensive packet data. The authors introduce an innovative classifier for NIDSs that relies on hierarchical long short-term memory (LSTM). Their suggested NIDS attains superior detection accuracy in contrast to current approaches by proficiently managing the entirety of packet information. The evaluation of performance validates the enhanced detection capabilities. Elsayed et al. (2021) unveiled a secure automated two-tier intrusion detection system (SATIDS) tailored for IoT environments. The system they propose incorporates feature selection techniques and an improved variant of (LSTM) to elevate the performance of the (IDS). Via training and evaluation on real-world datasets, the suggested algorithm showcases enhanced accuracy and detection rates when compared to existing IDS solutions.

3. The main work

Herein, the proposed prediction model consists of LSTM with an attention layer. The prediction model executes a binary classification to make sure that the attack is occurring or not and predicts the accurate category of the attack. The NB15 dataset (Verma & Ranga, 2017) utilized to train and test the proposed model. In addition, data exploration has been applied. The exploration and analysis for the data set is in the coming subsection, this exploration provides a better understanding for the dataset.

4. Dataset Description

This study is dedicated to the widely recognized benchmark dataset UNSW-NB15, which was curated primarily by Mustafa and Slay (2015). They employed specialized tools, including IXIA Storm, for the dataset's preparation. They meticulously monitored the typical behavior of network traffic and incorporated contemporary attack techniques spanning nine distinct categories. The UNSW-NB15 dataset encompasses an extensive array of both real and simulated network traffic intrusion activities. Furthermore, they generated the dataset's features using both conventional and innovative approaches. In total, the UNSW-NB15 dataset comprises 49 features, which would be categorized into various groups. To facilitate our analysis, we partitioned the dataset into training and testing subsets, as outlined in Table 1. Additionally, Table 2 provides details about the dataset's features, including their data types and traffic feature categories. It also includes other information related to flow, basic attributes, content, time, connection, and common characteristics.

Table 1
UNSW-NB15 classification within numerous categories with Training and testing splitting

NO	Category	UNSW-NB15 Train Set	UNSW-NB15 Test Set
1	Normal	37,000	56,000
2	**Dos**	**4089**	**12,264**
3	DDos	-	-
4	Probe	-	-
5	U2R	-	-
6	Generic	18,871	40,000
7	Analysis	677	2000
8	Fuzzerz	6062	18,184
9	Worms	44	130
10	Exploits	11,132	33,393
11	Backdoor	583	1746
12	Shellcode	378	1133
13	Reconnaissance	3496	10,491
14	Theft	-	-
15	R2L	-	-
Total		82,332	175,341

Table 2
The attributes- Features inside UNSW_NB15 in standings of data type types and categories.

Category	NO.	Name	Datatype	Category	No.	Name	Data type
	1	Dstip	Nominal		25	res_bdy_len	integer
	2	Sport	Integer	content	26	trans_depth	integer
Flow	3	Proto	nominal		27	Syncak	Float
	4	Dsport	Integer		28	Djit	Float
	5	scrip	nominal		29	ackdat	Float
	6	Service	Nominal		30	Ltime	Timestamp
	7	Dur	Float	Time	31	Sintpkt	Float
	8	dttl	Integer		32	Dintpkt	Float
	9	Dloss	Integer		33	Tcprit	Float
	10	sttl	Integer		34	Sjit	Float
Basic	11	sload	Float		35	stime	Timestamp
	12	sloss	Integer		36	Ct_ftp_cmd	integer
	14	state	Nominal		37	Is_ftp_login	Binary
	15	sbytes	Integer	General Purpose	38	Ct_flw_http_mthd	integer
	16	Diaod	Float		39	Ct_state_til	integer
	17	Spkts	Integer		40	Is_sm_ips_ports	integer
	18	Dpkts	Integer		41	Ct_dst_sport_ltm	integer
	19	Swin	Integer		42	Ct_srv_ltm	integer
	20	Dmeansz	Integer	connection	43	Class	integer
content	21	Stcpb	Integer		44	Ct_src_ltm	integer
	22	Dtcpb	Integer		45	Ct_src_doprt_ltm	integer
	23	Smeanz	Integer		46	Attack_cat	Nominal
	24	dwin	Integer		47	Ct_dst_src_ltm	integer
					48	Ct_dst_ltm	integer
					49	Ct_srv_src	integer

5. Dataset pre-processing

Indeed, most of the dataset required a pre-processing step. It is a procedure which contains several processes such as data normalization, cleaning, reduction, and transformation. The mentioned steps are significant since they affect the performance for the classifier model (da Costa et al., 2021). This research utilizes perfect inputs from the UNSW-NB15 dataset for the proposed model. Therefore, we have employed the most important and required data preprocessing steps. For instance, the selected training sets from UNSW-NB15 do not contain redundancy and duplication records. Also, we applied data normalization by applying a scaling effect on values for the data within a relative range of every feature. We have applied this step to evade dataset's bias toward features that naturally have superior values. Based on the Eq. (1) almost all the data set has been normalized and has the range into [0, 1]. Then, in the following step, we made a conversion for the symbolic data to numeric numbers. Furthermore, all the attack kinds in the data set have been converted into two labels, either 1 to the abnormal instances or 0 label for the set of the normal instances. After that, we extracted only the normal data samples from the training data sets. The main reason for this extraction step is because the proposed model is developed to be in usage in the early phases of the lifetime of the network. This is because the suggested IDs are invented to be utilized in the initial phases of a network period and in the early stage the model can learn only the normal traffic. After that, in the testing stage, the proposed model must have the ability to differentiate between normal and attack traffic. Lastly, to ensure the proposed model has a perfect training speed, during the training phase, the acquired regular trials are clustered to decrease the number of training records.

$$X_{Normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

In addition, regarding performance metrics that have been implemented to assess the several techniques matched with the proposed model based on the Accuracy level for classification. The Accuracy Equation is represented in Eq. (2),

$$Accuracy = \frac{TP + TN}{TN + TP + FN + FP} \tag{2}$$

However, the accuracy metric is not enough for the evaluation process, especially for the classification model. Therefore, we have applied additional evaluation metrics such as confusion matrix- error matrix. It's a specific table that visualizes the performance for the algorithm. Each row shows the actual condition – class while the column represents the instanced in the predicted condition-class as shown in Table 3.

Table 3
Confusion Matrix metric

		Predicted condition	
	Total population = P + N	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True positive (TP)	False negative (FN)
	Negative (N)	False positive (FP)	True negative (TN)

6. The proposed model basics and architecture

6.1 Attention based LSTM Model, AT-LSTM Model

At the same time, there is no definite computerized model, neither a ML nor a DL model, which will be able to certify the network intrusion detection systems (NIDS). Hence, the fundamental influence for this research mostly comprises the subsequent: we implicate the attention model to powerfully mine the features of UNSW-NB15 dataset. The presented model employed a special influential layer termed LSTM layer. This layer is positioned in the offered model and achieves preservative interactions; it can aid in enhancement gradient flow through extended orders in training. According to standard models, AT-LSTM can conserve and effort with the non-stationary sequences, also it can sense the nonlinear relations (Elman , 1991). Besides, when compared to DL models such as RNN, AT-LSTM can effectively circumvent issues related to long-term dependencies and contribute to enhanced interpretability (Alsharaiah et al., 2022; Vaswani, 2017). The technique for the attention in the suggested model makes it easy to distinguish how the data in the input sequence impacts the last generated arrangement by the model outcome process (Kim, 2017). This step helps in finding the interior action instrument of the model and repairs specific exact inputs and outputs. Supplementary, the consequences from the experiments on UNSW-NB15 datasets clarifies that AT-LSTM succeeds improved tasks than classical models.

1. The Design of the AT-LSTM Model

The proposed AT-LSTM model for binary class prediction on UNSW-NB15 datasets comprises two components: the LSTM is DL model and the attention model. The attention layer can dynamically choose the most distantly associated input features and assign greater importance to the corresponding original feature sequence. Subsequently, herein utilized the LSTM prototype as input for the attention prototypical to make predictions for binary classification.

2. The LSTM Model

To definite input raw, $X = (x^1, x^2, \dots, x^n)^T = (x_1, x_2, \dots, x_m) \in R^{(n \times m)}$, n denotes the quantity of sequences for feature - orders, also m represents the measurement of the window. $x^k = (x_1{}^k, x_2{}^k, \dots, x_m{}^k)^T \in R^m$ is involved to indicate a sequence – vector- of length m. We utilize $x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$ to denote a set of groups for the vectors n features on time t. Further, the group model is recognized as follows: Let x_t, h_t *and* C_t characterize the input, controller state, and the cell state through time step t. Transmission an arrangement – sequence- of inputs $(x_1, x_2, \dots, x_m)$ also, the LSTM guesses the collection of sequence $(h_1, h_2, \dots, h_m)$ and the C-sequence $(C_1, C_2, \dots, C_m)$ as shown below:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \quad (3)$$
$$c_t = tanh\,(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$
$$C_t = f_t * C_{t-1} + i_t * c_t) \quad (5)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (6)$$
$$h_t = o_t \times \tanh\,(C_t) \quad (7)$$

For illustration, every equation has a fixed of singular symbols, and recognizes several functions. For incidence, σ denotes the function of logistic sigmoid, $*$ is a factor wise product, also C_t is indicates if the state of the cell that is needed to be altered. Similarly, W_f, W_i, W_C, W_o and b_f, b_i, b_C, b_o are a group of parameters for the model. Moreover, parameters can be learned during the treatment. Furthermore, f_t, i_t and o_t are similarly sanctified as gates for the forgotten, alongside an input gate as well as output gate. Surely, every LSTM unit takes a memory cell because each LSTM unit that committed a memory cell takes state **C***t* at time *t*, and it is measured through the 3 overhead gates as exposed in Fig. 2 for the architecture of the suggested model.

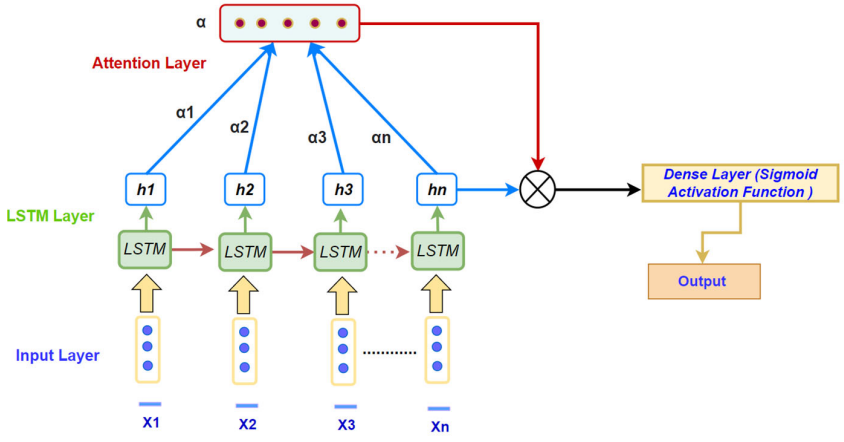

Fig. 2. The Architecture of the Attention-LSTM Model.

3. The Attention Model

An important portion for artificial human is that there's no straight deal through all responses from the external realm. As a supernumerary, human's starting attention is on the important segments to obtain the info they necessitate. It is likewise crucial to highlight key features first and eliminate recurrent features. So, through the operational material motivated by the directly above evidence, we recommend an attention model, and this model knows how to perform the optimization to enhance the input feature sequence forecast of suspicious hacking activities. The mechanism of attention (Kim et al., 2017) can be clear as mapping a query. Furthermore, a group of key-value pairs to an output, and the other model component is defined as vectors. For instance, the query, keys, values, and the containing output are demarcated as well clear vectors. The model's output is determined by calculating a weighted sum of values. These weighted sums are computed using a function that is linked to the compatibility between the query and its corresponding key, as signified in Fig. 2. The technique of generating the new input features and the attention weights and recognition on attention is clarified in Fig. 2. In the introductory section, x_t maps to h_t over the subsequent.

$$h_t = f_1(h_{t-1}, x_t) \quad (8)$$

The f_1 stands for activation function (non-linear), while the state with hidden format on time t, and s specifies the size of the state with hidden format while represented by $h_t \in R^s$. LSTM is fulfilled as f_1 also, the implementation for the LSTM model keens to avoid the long-term dependence concern, which usually rises with technique of data prediction. During the subsequent fragment, we produce an attention instrument via utilizing specially the deterministic feature inside attention

prototypical. For the meticulous feature sequence similar $x^k = (x_1^k, x_2^k, \dots, x_m^k)^T \in R^m$, by depending on to the aforesaid hidden state h_{t-1} and the cell state C_{t-1} in the LSTM unit, we prompt

$$\alpha_t^k = v^T \tanh(W_1 \cdot [\, h_{t-1},\, C_{t-1}] + W_2 x^k) \tag{9}$$

$$\beta_t^k = softmax(\alpha_t^k) = \frac{\exp(\alpha_t^k)}{\sum_{i=1}^{n} \alpha_t^k} \tag{10}$$

The binary matrices W_1, W_2 and the vector *v* indicate the learning factors of the recommended model. a^k is a vector with a size termed *m* and its i-th element dealings the importance of the k-th input feature order at time t. The abovementioned items need be standardized over softmax. β^k It signifies the attention weight, comprising a score that indicates the level of attention allocated to the k-th feature sequence. We can also derive the results of the attention model at time t, which corresponds to the arrangement of the weighted input feature referred to as z_t would be offered as:

$$z_t = (\beta_t^1 x_t^1, \beta_t^2 x_t^2, \dots, \beta_t^n x_t^n)^T \tag{11}$$

In all the equations, x_t exchanged through a new computed z_t to retain up the attention in the model. Nevertheless, traditional prediction models that encompass RNN typically applied dataset features as input, also handling all sequences for the input feature in a comparable manner. Yet, the lately obtained z_t be able to guide auxiliary attention to the certain sequence for the input feature, extracting the sequences key feature competently, and depend on the weight in attention, we squeezed the effect of the repeated feature orders. Theoretically, there is an enhancement in expectation accuracy with z_t for the input in softmax layer.

7. Outcomes and deliberations

Data exploration and imitation Experiments

This research accomplished numerous trials on the suggested AT-LSTM model for UNSW-NB15 Binary classification. The suggested model is trained in the Denial of services (DS) dataset which is mentioned in Table 1, so, the proposed model is trained on 82,323 samples and it is tested on 175,341 samples. To validate the model's performance, various established techniques can be employed. Cross-validation methods are routinely used to assess the predictor's accuracy. These methods include K-fold cross-validation, the jackknife method, and sub-sampling. The jackknife method is often regarded as the most objective and least arbitrary among these, and it has been widely adopted by researchers to evaluate the effectiveness of different predictors. However, it is computationally intensive and time-consuming. To address this, herein, we have implemented an initial prevention mechanism to prevent model overfitting, setting the patience parameter to 3 epochs. Additionally, we have utilized K-fold cross-validation, with K becoming 1, as we only performed a train/test split to assess the attention-based LSTM model for binary classification in the UNSW-NB15 dataset.

7.1 Training

The proposed network intrusion detection systems (NIDS) framework has been implemented via Keras and Python. They were utilized to train the AT-LSTM model for UNSW-NB15 Binary classification. Regarding the mission for classifying data, a powerful Adam optimization procedure has been involved through learning rate amount, and it was set to 0.01 and momentum fixed to 0.0. supplementary, the batch size for the model was also set to 90. The model began with incorporating 342,106 parameters. The model utilized 256 LSTM units, likewise the attention dimensions were fixed to 150 and 300. As well, for the classification task the suggested model size showed to necessitate 55 second for each epoch. Also, over in each epoch the training data would be arbitrarily shuffled. The AT-LSTM model for forecasting suspicious hacking activities was trained to decrease the validation loss in binary_crossentropy.

7.2 Results investigation

Most of machine learning model that deliver a binary classification for the dataset like UNSW-NB15 utilized feature engineering because the dimenationality. For instance, Chen and Guestrin (2016) apply filter-inspired feature reduction technique and Zhu et al. (2007) also applied wrapper-based feature extraction approach. However, as mentioned before the proposed model has been built without applying feature engineering and the result is comared with other machine larning models over the same data set.

For instance, Khammassi and Krichen (2017) implemented a hybrid mode consisting of Genetic algorithm (GA) in combination per logistic regression established on the UNSW-N15 and KDDCup99 datasets. The outcomes indicated that the model could provide an inclusive correctness 81.42% by using the whole features set. Ambusaidi et al. (2016) developed a classifier model based on Least Square SVM (LS-SVM) . The LS-SVM FMI acquired an inclusive accuracy of 78.86% . Janarthanan and Zargari (2017) applied a wide study on the UNSW-N15 dataset utilizing the Weka tool, the authors implemented several algorithms such as the attribute evaluator, Ranker Method, Greedy Stepwise and Information Gain. In

addition, the authors assess the efficiency of every subset by using the Kappa Statistic. However, the results showed that tRF classifier was carefully chosen as the finest method in terms of its total interpretation with accuracy of 75.66%. Another IDS system implemented such as Kumar et al. (2020) based on the UNSW-NB15 dataset. The writers involved the information Gain methodology in their model, and they also used integrated rule based models that engage multiple tree classifiers. The results from their IDS models obtained the accuracy with 57.01%. However, they reported that their model can be enhanced if an alternative machine learning procedure could be included in alteration to a severe devotion to Tree based approaches. Khan et al. (2020) developed an IDS system based on machine learning algorithms such as RF algorithm. The outcome from their model shows that The RF algorithm acquired the finest outcomes with an Accuracy of 75.56% . Gao et al. (2019) recommended a hybrid IDS system utilizing an Advanced Principal Component (APCA) procedure implemented with the increased method of the Extreme Learning Machine (IELM). The proposed model was trained and tested over the UNSW-N15 data set. However, the outcomes disclosed that the IELM-APCA achieved an accuracy of 70.51%. Furthermore, Jiang et al. (2020) offered a combined NIDS framework. The framework consists of Bidirectional - Long-Short Term Memory (Bi-LSTM) and (CNN). The mixture of CNN-Bi-LSTM models trained and tested with the UNSW-N15 data set. However, The outcomes exposed that the CNN-Bi-LSTM accomplished precisions of 77.16%.

On execution experiments, it is possible to be confident that the investigators have completed using trials on the AT-LSTM model by trying diverse hyper-parameters. The offered model was likewise tested with 3 dissimilar arrangements. For instance, the BiLSTM using Attention layer, LSTM using Attention Layer with dimension 150 and LSTM with Attention dimension 300. Also, the performances for the classification process was estimated will be recorded in an accurateness score involuntary metric assessment. The information that is represented in Table 4 demonstrates the effectiveness of the suggested model for the binary classification. Note that according to Table 4, the suggested model acquired an admirable consequence when we employed LSTM through the attentional technique. For instance, the proposed model achieved an accurateness of 92.2. Compared to additional settings recorded in Table 1, the suggested model (LSTM with Attention dimension 300) gained an improved accuracy than the last conformations. The outcomes recommend that the suggested model is operative and precise in binary classification in a justification dataset. However, as exposed in Table 3, the attention-based BiLSTM configuration has not acquired a reasonable outcome because it achieved an accuracy of 71.60, which identifies that the attention based-BiLSTM configuration did ont notice suspicious hacking activities accurately. On other hand, our proposed Attention based LSTM model with dimension 150 has achieved competitive outcomes, it achieved an accuracy of 82.27. As exemplified in Table 4, it can be determined that the suggested model outdoes the Attention-based BiLSTM. Further significantly, the results offered in Table 4 and Fig. 3 illustrate the interpretation of the proposed model that uses the attention methodology and LSTM was greater than the model that utalized BiLSTM. Furthermore, as illustrated in Fig. 3, it becomes evident that the fault in the training data decreases as the learning process progresses, and simultaneously, the error in the validation data also diminishes as the training proceeds. This trend serves to indicate that the proposed AT-LSTM model is free from overfitting concerns.

Table 4
Experimental outcome.

Model Configuration	Accuracy	The Epochs Numbers	Attention Dimension
Bi-LSTM +Attention	71.60	9	150
The proposed model (LSTM with Attention)	92.2	12	300
The proposed model (LSTM with Attention	82.27	15	150

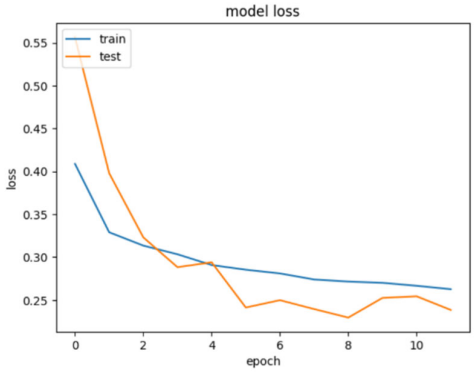

Fig. 3. Model Error

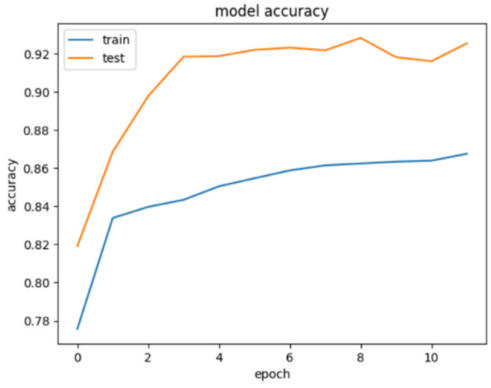

Fig. 4. The proposed AT-LSTM model accuracy

In addition, as revealed in Fig. 4 the accuracy is plotted and it can be realized that the proposed model is trained right fine. The tendency for accuracy in the training and testing datasets are quite increasing starting in epoch number 2 until epoch number 8. This is an indication of the AT-LSTM model performance and accurate classification. Furthermore, we have

applied additional metrics to measure the affectivity of the proposed model. We have applied the confusion matrix metric as exposed in Fig. 5.

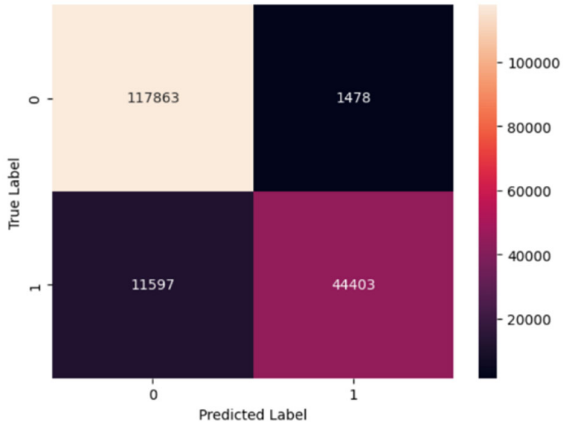

Fig. 5. Confusion Matrix for the proposed model.

The confusion metric exposes how the model performs and explains when the model is confused in prediction. For example, the confusion matrix provides insight into the count of true positives, which are correctly predicted in the correct class and genuinely belong to that category. It also accounts for true negatives, which represent instances that belong to another class and are correctly classified as such. As shown in Fig. 5, the prediction of the true positive number is 117863 and the true negative prediction number is 44403. On other hand, the confusion matrix shows the false in prediction when predicting and classifying samples in classes, but in the actual they belong to different classes. The false prediction is presented in false positives and negatives. Fig. 5 also exposes the false positive with 11597 samples and false negative 1478.

Table 4
A comparison between the proposed AT-LSTM model Accuracy with further available Machine learning and Deep Learning models from the literature

The IDEs Model	The core Machine learning and deep learning algorithm in the model	Accuracy
Zhu et al., 2007	CNN-Bi-LSTM	77.16%.
Chen and Guestrin (2016)	IELM-APCA	70.51%.
Kim et al., 2017	RF algorithm	75.56%
Kim et al., 2017	rule based model	57.01%.
Hochreiter & Schmidhuber, 1997	tRF classifier	75.66%.
Alsharaiah et al. (2022)	Least Square SVM (LS-SVM)	78.86%
Ambusaidi et al. 2016	Genetic algorithm with logistic regression	81.42%
	The proposed AT-LSTM model	92.2%

Also, Fig. 6 disclose the accuracy for the offered AT-LSTM model per additional available machine learning and deep learning models for binary classification tasks.

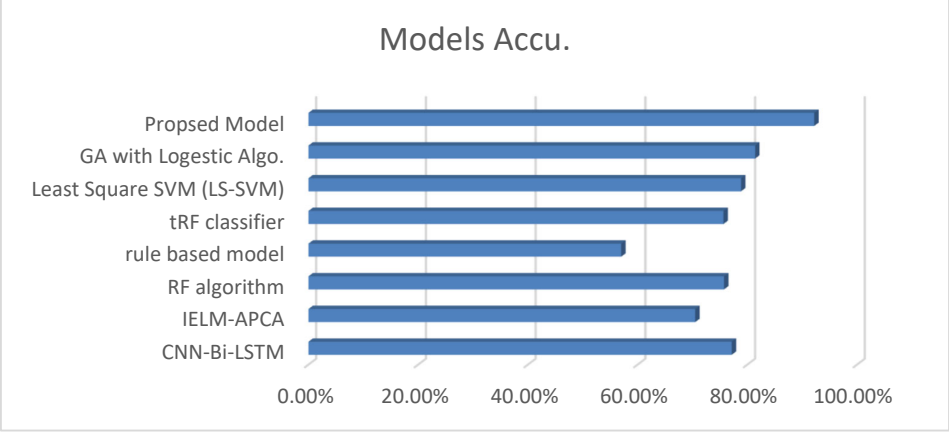

Fig. 6. The accuracy illustration between the proposed AT-LSTM model with other available machine learning and deep learning models for binary classification.

8. Discussion

The proposed model has presented an integrating LSTM and an attention layer, and it has attained a binary classification accuracy of 92.2% on the UNSW-NB15 dataset. This outperformed several existing methods applied to the same dataset, such as the CNN-Bi-LSTM model (77.16%) (Jiang et al., 2020), IELM-APCA (70.51%) (Gao et al., 2019), RF algorithm (75.56%) (Sultana et al., 2019), rule-based models (57.01%) (Kumar et al., 2020), LS-SVM (78.86%) (Ambusaidi et al., 2016).

A key factor contributing to the advantages of the suggested model lies in its proficiency in accurately capturing prolonged correlations and temporal trends within the network traffic data. LSTM, categorized as a recurrent neural network, demonstrates excellence in modeling sequential data and preserving information across extensive sequences. The LSTM layer in the proposed model enables it to learn from the sequential nature of the network traffic, capturing important patterns that might be missed by other methods.

Firstly, the LSTM component of the proposed model enables it to effectively model the temporal dependencies and sequential patterns present in network traffic data. Traditional machine learning algorithms, such as rule-based models or SVMs, often struggle to capture the complex and dynamic nature of network traffic. LSTM, as a member of the recurrent neural network family, stands out in its ability to grasp extended connections and uphold data throughout lengthy sequences. By incorporating LSTM, the proposed model can effectively learn from the temporal features of the dataset, leading to improved accuracy in detecting network intrusions.

Furthermore, Within the proposed model, the attention layer introduces an added dimension of precision and concentration on pivotal features. The attention mechanism empowers the model to dynamically allocate significance to diverse sections of the input sequence, amplifying the emphasis on the most pertinent information. This adaptable feature selection elevates the model's aptitude to differentiate between regular and malicious network traffic, culminating in an enhanced classification accuracy. By prioritizing the most informative elements of the data, the model can more effectively uncover meaningful patterns and furnish more precise predictions.

Compared to traditional machine learning algorithms, which often rely on handcrafted feature engineering, the advanced deep learning model we recommend autonomously acquires feature representations straight from the unprocessed data. This holistic learning method eradicates the necessity for labor-intensive manual feature extraction, which can prove demanding and time-consuming in intricate datasets such as network traffic. By leveraging the power of deep learning and allowing the model to learn hierarchical representations, the proposed model achieves better performance and avoids the limitations imposed by manual feature engineering.

However, it is important to consider the restrictions and limitations of the proposed model and other methods. Deep learning models, including the proposed LSTM-based model, can be computationally demanding and require substantial computational resources for training and inference. The training phase frequently entails optimizing a substantial number of parameters, demanding significant time and resources. Furthermore, implementing deep learning models within real-time intrusion detection systems may necessitate efficient hardware architectures to guarantee swift and responsive detection.

Furthermore, it is crucial to thoroughly assess the proposed model's resistance to adversarial attacks. Adversarial examples are meticulously crafted inputs designed to mislead the model and evade detection. It is also imperative to evaluate the model's ability to withstand such attacks and develop strategies for bolstering its resilience. Exploring techniques like adversarial training, defensive distillation, or other methods can enhance the model's capacity to resist adversarial manipulation.

The comparison of the proposed model with known methods highlights the benefits of leveraging LSTM and attention mechanisms in network intrusion detection. The proposed model's ability to capture long-term dependencies, focus on important features using attention mechanisms, and leverage the power of deep learning contributed to its superior performance. However, the restrictions of computational complexity and robustness against adversarial attacks should be carefully addressed in practical deployments.

9. Conclusion

This study has introduced the AT-LSTM model as an innovative approach for achieving precise intrusion detection in IDSs. The model's performance was assessed using the widely recognized UNSW-NB15 dataset, which serves as a benchmark for network intrusion detection. A comprehensive review and comparison of various methods and classifiers for binary classification on the UNSW-NB15 dataset were conducted through an extensive literature analysis.

The experimental outcomes clearly illustrated the superiority of the proposed AT-LSTM model in terms of detection accuracy when compared to other machine learning methods on the test data. This underscores the efficacy of incorporating an attention-based LSTM architecture in network intrusion detection tasks. Nonetheless, there is room for further improvement in the model's performance.

In future work, the application of synthetic oversampling techniques can be explored to address the class imbalance issue within the UNSW-NB15 dataset, potentially yielding even more favorable results. Additionally, investigating the use of an XGBoost-based feature selection method with the NSL-KDD dataset can help evaluate its impact on the model's performance compared to existing state-of-the-art approaches. These enhancements can bolster the proposed AT-LSTM model's capacity to accurately detect and categorize network attacks.

Overall, this research contributes to the field of network intrusion detection by introducing a novel model that enhances accuracy in binary classification tasks. The findings underscore the potential of integrating attention-based LSTM with advanced machine learning techniques to create more effective and efficient intrusion detection systems.

References

Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE access*, *6*, 33789-33795.

Alsharaiah, M. A., Baniata, L. H., Adwan, O., Abu-Shareha, A. A., Alhaj, M. A., Kharma, Q., ... & Baniata10, M. (2022). Attention-based Long Short Term Memory Model for DNA Damage Prediction in Mammalian Cells. *development*, *13*(9).

Ambusaidi, M. A., He, X., Nanda, P., & Tan, Z. (2016). Building an intrusion detection system using a filter-based feature selection algorithm. *IEEE transactions on computers*, *65*(10), 2986-2998.

Bengio, Y., Delalleau, O., & Roux, N. (2005). The curse of highly variable functions for local kernel machines. *Advances in neural information processing systems*, *18*.

Biswas, S. K. (2018). Intrusion detection using machine learning: A comparison study. *International Journal of pure and applied mathematics*, *118*(19), 101-114.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Cisco. (2003). Guide to Secure Intrusion Detection Systems. Cisco Security Professional's Guide to Secure Intrusion Detection Systems.

da Costa, N. L., de Lima, M. D., & Barbosa, R. (2021). Evaluation of feature selection methods based on artificial neural network weights. *Expert Systems with Applications*, *168*, 114312.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine learning*, *7*, 195-225.

Elsayed, R., Hamada, R., Hammoudeh, M., Abdalla, M., & Elsaid, S. A. (2022). A Hierarchical Deep Learning-Based Intrusion Detection Architecture for Clustered Internet of Things. *Journal of Sensor and Actuator Networks*, *12*(1), 3.

Fu J, Z. (2017). Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, 4476-4484.

Gao, J., Chai, S., Zhang, B., & Xia, Y. (2019). Research on network intrusion detection based on incremental extreme learning machine and adaptive principal component analysis. *Energies*, *12*(7), 1223.

Han, J., & Pak, W. (2023). Hierarchical LSTM-Based Network Intrusion Detection System Using Hybrid Classification. *Applied Sciences*, *13*(5), 3089.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Hu, W., Gao, J., Wang, Y., Wu, O., & Maybank, S. (2013). Online adaboost-based parameterized methods for dynamic distributed network intrusion detection. *IEEE Transactions on Cybernetics*, *44*(1), 66-82.

Janarthanan, T., & Zargari, S. (2017, June). Feature selection in UNSW-NB15 and KDDCUP'99 datasets. In *2017 IEEE 26th international symposium on industrial electronics (ISIE)* (pp. 1881-1886). IEEE.

Jiang, K., Wang, W., Wang, A., & Wu, H. (2020). Network intrusion detection combined hybrid sampling with deep hierarchical network. *IEEE access*, *8*, 32464-32476.

Khammassi, C., & Krichen, S. (2017). A GA-LR wrapper approach for feature selection in network intrusion detection. *computers & security*, *70*, 255-277.

Khan, N. M., Madhav C, N., Negi, A., & Thaseen, I. S. (2020). Analysis on improving the performance of machine learning models using feature selection technique. In *Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2* (pp. 69-77). Springer International Publishing.

Kim, Y., Denton, C., Hoang, L., & Rush, A. M. (2017). Structured attention networks. *arXiv preprint arXiv:1702.00887*.

Kim, Y., Denton, C., Hoang, L., & Rush, A. M. (2017). Structured attention networks. *arXiv preprint arXiv:1702.00887*.

Kumar, V., Sinha, D., Das, A. K., Pandey, S. C., & Goswami, R. T. (2020). An integrated rule based intrusion detection system: analysis on UNSW-NB15 data set and the real time online dataset. *Cluster Computing*, *23*, 1397-1418.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

Michalas, A., & Murray, R. (2017, October). MemTri: A memory forensics triage tool using bayesian network and volatility. In *Proceedings of the 2017 International Workshop on Managing Insider Security Threats* (pp. 57-66).

Moustafa, N., & Slay, J. (2015, November). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 military communications and information systems conference (MilCIS)* (pp. 1-6). IEEE.

Pal, S. K., & Mitra, S. (1992). Multilayer perceptron, fuzzy sets, and classification. IEEE Transactions on Neural Networks, 3, 683–697.

Panda, M., Abraham, A., Das, S., & Patra, M. R. (2011). Network intrusion detection system: A machine learning approach. *Intelligent Decision Technologies*, *5*(4), 347-356.

Pandya, P. (2013). *Computer and Information Security Handbook, 3rd ed.*

Sultana, N., Chilamkurti, N., Peng, W., & Alhadad, R. (2019). Survey on SDN based network intrusion detection system using machine learning approaches. *Peer-to-Peer Networking and Applications*, *12*, 493-501.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

The UNSW-NB15 Dataset. (n.d.). (The UNSW-NB15 Dataset) Retrieved from https://research.unsw.edu.au/projects/unsw-nb15-dataset

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Wang, K. (2019). Network data management model based on Naïve Bayes classifier and deep neural networks in heterogeneous wireless networks. *Computers & Electrical Engineering*, *75*, 135-145.

Zhang, J., Zulkernine, M., & Haque, A. (2008). Random-forests-based network intrusion detection systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *38*(5), 649-659.

Zhu, Z., Ong, Y. S., & Dash, M. (2007). Wrapper–filter feature selection algorithm using a memetic framework. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *37*(1), 70-76.