

Anomaly Detection in Network Traffic Using Machine Learning - IDSSNet

- Sharen H

1. INTRODUCTION:

The rapid growth of the Internet and the steadfast backing of state regulations have both helped propel China's cloud computing sector into the fast lane of development in recent years. Additionally, more and more businesses are choosing to move their operations to the cloud. Cloud security challenges are becoming more and more prevalent as a result of the quick growth of cloud technology. Criminals have employed several cutting-edge cloud open network technologies as a new class of criminal tactics that substantially jeopardise the steady functioning of the cloud open network, result in economic losses for the country, and even pose a threat to national security.

Anomaly identification, also known as outlier analysis, is a data mining procedure that finds data points, occurrences, and/or observations that differ from a dataset's typical pattern of activity. Unusual data might point to serious occurrences, like a technological malfunction, or to promising possibilities, like a shift in customer behaviour. Automated anomaly detection is increasingly being done using machine learning.

Authentication, access control, data encryption, and intrusion detection are only a few of the techniques used to defend networks from security threats. This study employs deep learning and falls under the category of intrusion detection. It should be noted that the assault traffic data differs significantly from the regular traffic data. Anomaly network traffic detection may effectively detect and intercept network assaults in advance and limit the losses brought on by network attacks by finding changes in the characteristic of the two traffic data. The set criteria created manually for matching and classification served as the foundation for early identification and categorization of aberrant traffic [1].

Network security is provided by anomaly detection systems, which can precisely identify malicious network activity. The complexity and breadth of network assaults are increasing as internet technology advances, making it challenging for conventional anomaly detection systems to monitor and detect aberrant information. Deep neural network (DNN) technology has now shown excellent results in anomaly detection and is capable of automatic detection. Deep neural networks' prediction outputs do nevertheless contain misclassified traffic, which leads to duplicate warning data [2].

In order to do this, we utilise convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) to build reliable multi-class classifiers that can categorise each new network log instance that enters our system. A number of quantitative tests are conducted to assess the performance of our method, and it is then compared against cutting-edge formulations.

2. LITERATURE SURVEY:

To improve the effectiveness of anomaly detection, many machine learning techniques, including supervised, unsupervised, and semi-supervised methods, have been developed. Anomaly detection techniques under supervision include neural networks, support vector machines (SVM), and k-nearest neighbour (k-NN). On the KDDCup99 Dataset, Gao et al. [3] suggested an IDS architecture based on DBNs employing energy-based reduced Boltzmann machines (RBMs). Aygun and Yavuz [27] reported accuracy of 88.28% and 88.6% on the NSLKDDTest+ dataset by using vanilla and denoising deep Autoencoders on NSLKDD. A taxonomy and survey of deep and conventional structures for intrusion detection were presented by Hodo et al. [4]. A thorough analysis of supervised and unsupervised learning strategies for anomaly detection was presented by Horbani et al. in their paper from the year 22. For the purpose of detecting anomalies, Solanas and Martinez-Balleste [5] introduced clustering techniques.

Thing et al. [6] approach was built on deep learning-based anomaly detection and classification. The deep learning technique is capable of effectively classifying attacks and self-learns the attributes required to detect network anomalies. We treated the classification in our trials as a multi-class problem (i.e., genuine traffic, flooding type assaults, injection type attacks, and impersonation type attacks), and were able to classify the attacks with an overall accuracy of 98.6688% using the proposed approach. A deep neural network (DNN) for an IoT network was proposed by Ahamed et al. [7] as an effective anomaly detection technique utilising mutual information (MI). The IoT-Botnet 2020 dataset is used to do a comparative examination of several deep-learning models, including DNN, Convolutional Neural Network, Recurrent Neural Network, and its various versions, such as Gated Recurrent Unit and Long Short-Term Memory. Experimental findings demonstrate the efficiency of the DNN-based NIDS model in comparison to the well-known deep learning models by demonstrating an accuracy increase of the model of 0.57-2.6% and a reduction in the FAR of 0.23-7.98%.

A semi-supervised detection framework built on Unsupervised DL techniques was introduced by Dawoud et al. [8]. The study examines the potential and difficulties of using autoencoders as a non-probabilistic approach to use DL to detect abnormalities. They provide AE a thorough examination for finding abnormalities. Our findings indicate that the USDL would improve detection with a 99% accuracy rate. Miao et al. [9] proposed a method for predicting river water levels and detecting anomalies by combining the Conv-GRU model with the multivariate Gaussian distribution approach. In this potent time series prediction model, the GRU model was used to learn the long-term dependant features in a time series, while CNN was utilised to acquire features from the time series data. Based on the data sets from the water level stations, the combined CNN and GRU model was used to forecast water levels. Finally, the chance of abnormal water level behaviour was calculated using the resultant prediction error, which was represented as a multivariate Gaussian distribution.

3. The Proposed Methodology

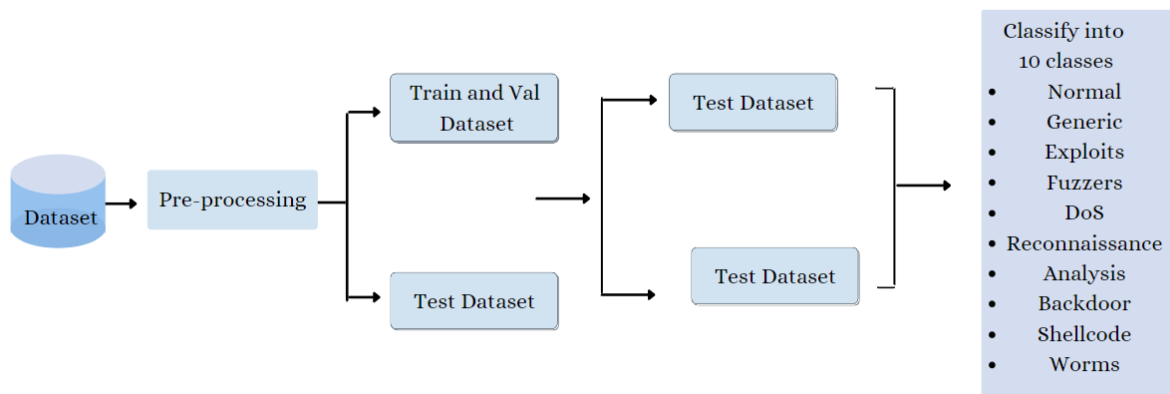


Fig 1 Describing the overall workflow.

The following sections explain the methodology

3.1 Dataset

The IXIA PerfectStorm tool in the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) produced the raw network packets of the UNSW-NB 15 dataset in order to produce a mix of genuine current normal activities and synthetic contemporary attack behaviours.

100 GB of the raw traffic are captured using the Tcpdump programme (e.g., Pcap files). Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, and Worms are among the nine attack categories in this dataset. To produce a total of 49 characteristics with the class label, the Argus and Bro-IDS tools are utilised, and twelve methods are built.

Class	Count
Normal	37000
Generic	18871
Exploits	11132
Fuzzers	6062
DoS	4089
Reconnaissance	3496
Analysis	677
Backdoor	583
Shellcode	378
Worms	44

Table 1. Data classes and its count

3.2 Data Processing

Normalization. First, we use the logarithmic scaling approach to limit the scaling scope for specific characteristics, where the gap between the maximum and lowest values has a very big scope. Second, each feature's value is linearly translated to the [0,1] range using the formula below, where Max and Min stand for the feature's maximum and minimum values, respectively.

$$y_i = \frac{y_i - Min}{Max - Min}$$

3.3 Proposed Model

IDSSNet architecture included one 1D convolution layer followed by a relu activation layer , max pooling layers with ReLu activation function and dropout. The initial layer included 64 filters of kernel size 3 processing the 1D signal resulting in optimal features. The obtained features are fed to relu activation layer and fed to convolution layer with 64 filters and kernel size of 3 and max pooling layers which reduced the features for next level. These reduced elite features are fed to a convolution followed by a max pooling layer. These features are flattened before proceeding to a fully connected layer. Instead of directly processing the data by a fully connected layer the features are dropped at 30% to avoid overfitting. Finally two dense layers with ReLu activation layer and softmax function are used for multi class classification. To analyze and test the efficiency of the IDSSNet model, it was trained and tested.

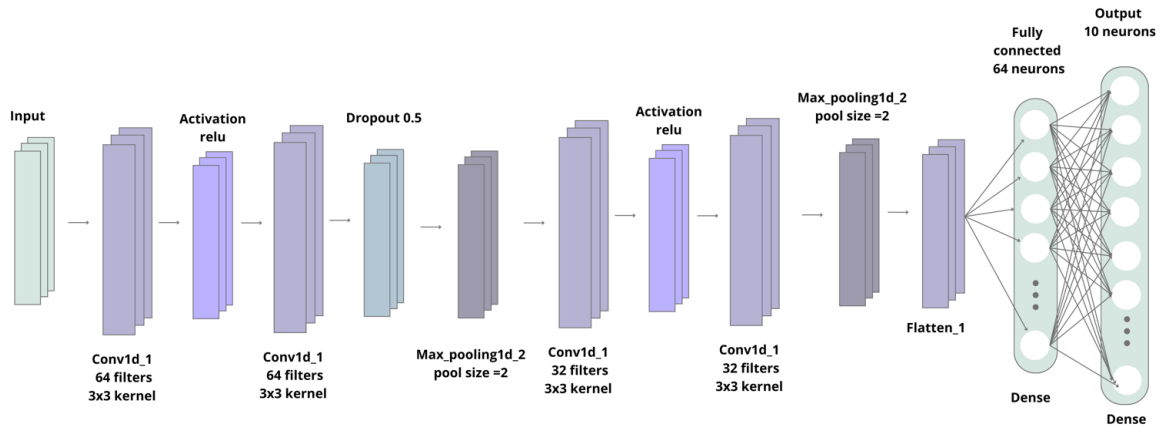


Fig 2. Architecture of the proposed IDSSNet

3.4 Experimental Setup

The proposed work was implemented using the cloud computing platform Kaggle. A free graphics processing unit (GPU) is available from Kaggle for creating deep learning models. Using Kaggle Notebook, the IDSSNet model was trained and put to the test.

3.5 Hyperparameter tuning

The learning rate, dropout, batch size, and number of epochs are the hyperparameters that were employed in this model to enhance performance. For improved performance, 30 epochs with a batch size of 1028 were chosen. The effectiveness of the 30% dropout rate was examined. Adam is an optimization strategy that may be used to update weights in the network periodically depending on training data where Adam employs the value of AdaGrad and RMSProp. Adam replaces the standard stochastic gradient descent approach. The optimising function used, the Adams gradient optimizer, outperformed and improved accuracy.

4. Results

The accuracy has been estimated based on the outcomes of the anomaly detection on the 10 classes using CNN and LSTM. For IDSSNet and LSTM, the specificity was 85.04% and 79%, respectively. IDSSNet's training parameters include 114474. Table 2 displays the accuracy and loss graphs of the proposed model and LSTM learning rates employed in this investigation.

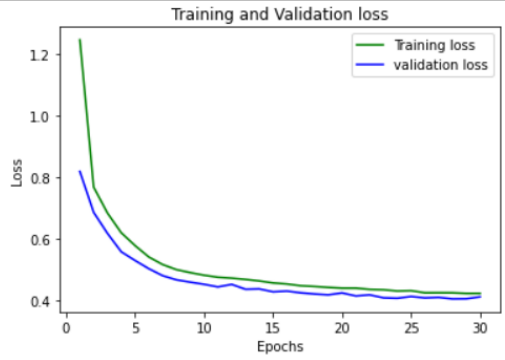
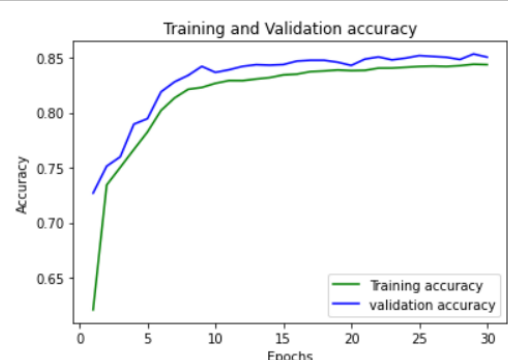
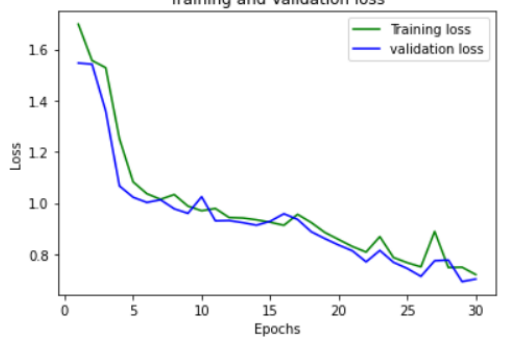
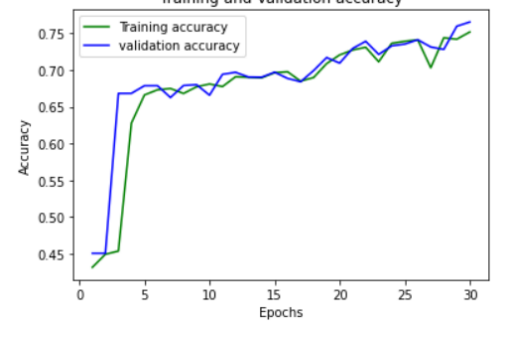
Model	Loss	Accuracy
IDSSNet		
LSTM		

Table 2. Learning process recorded by LSTM and the proposed model

Reference

1. Li, M., Han, D., Yin, X., Liu, H., & Li, D. (2021). Design and implementation of an anomaly network traffic detection model integrating temporal and spatial features. *Security and Communication Networks*, 2021.
2. Gao, M., Ma, L., Liu, H., Zhang, Z., Ning, Z., & Xu, J. (2020). Malicious network traffic detection based on deep neural networks and association analysis. *Sensors*, 20(5), 1452.
3. N. Gao, L. Gao, Q. Gao and H. Wang, "An intrusion detection model based on deep belief networks", *Proc. 2nd Int. Conf. Adv. Cloud Big Data*, pp. 247-252, Nov. 2014, [online] Available: <http://ieeexplore.ieee.org/document/7176101/>.
4. E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis and R. Atkinson, Shallow and deep networks intrusion detection system: A taxonomy and survey, 2017, [online] Available: <https://arxiv.org/abs/1701.02145>.
5. A. Solanas and A. Martinez-Balleste, *Advances in Artificial Intelligence for Privacy Protection and Security*, Hackensack, NJ, USA:World Scientific, 2010, [online] Available: <http://site.ebrary.com/id/10421991>.
6. Thing, V. L. (2017, March). IEEE 802.11 network anomaly detection and attack classification: A deep learning approach. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)* (pp. 1-6). IEEE.
7. Ahmad Z, Shahid Khan A, Nisar K, Haider I, Hassan R, Haque MR, Tarmizi S, Rodrigues JJPC. Anomaly Detection Using Deep Neural Network for IoT Architecture. *Applied Sciences*. 2021; 11(15):7050. <https://doi.org/10.3390/app11157050>
8. A. Dawoud, S. Shahristani and C. Raun, "Deep Learning for Network Anomalies Detection," *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, 2018, pp. 149-153, doi: 10.1109/iCMLDE.2018.00035.
9. Miao, S., & Hung, W. H. (2020). River flooding forecasting and anomaly detection based on deep learning. *IEEE Access*, 8, 198384-198402.

10. Link to code:

https://colab.research.google.com/drive/1KX-_dbcSNkyDFyeXEA5qS_w9M2YTDfhR?usp=sharing