****Application of MaCHINE LEARNING methods IN CYBERSECURITY****

Mesut Guven\*1

1Cyber Defense Command, Gendarmerie General Command HQ, Ankara/TURKEY

\* Corresponding author: mesutguven@ieee.org

**Abstract:** The Internet grid and computer networks have already transformed the way we live. Today, governments and commercial firms deployed their services to the internet. In everyday life, we enjoy using internet-connected devices thanks to the high speed and available internet grid. Since modern life is intertwined with computers and connected devices, cybersecurity tools and products have been developed to maintain the stable and integral operation of cyber assets. Most of those cybersecurity products make use of machine learning because classical signature-based methods are unable to detect zero-day attacks or slightly deformed versions of known attacks. In this work, we explain how machine learning, deep learning, and transfer learning methods are being used in cybersecurity.

**Keywords**: Cybersecurity, machine learning, deep learning, transfer learning, malware detection

# Introduction

The Internet has been the main medium for communication and nearly 60% of the world population can now have the opportunity to access to internet cloud [1]. Internet-based technologies have transformed our daily job routine, our way of entertainment, and even human interactions. Today, people are more reliant on their smartphones and computer devices. Moreover, the objects we use in our daily life such as vehicles, home appliances, etc. are fully connected to the internet which is described as the Internet-of-Things (IoT) phenomenon [2]. Modern telecommunication techniques and computer networks enabled individuals, governments, and institutions to migrate their functions to the cyber domain. Therefore, securing the data produced by the connected devices and ensuring stable and harmonious operation is of utmost importance. From this point of view, internet infrastructure and cyberspace are considered important instruments of modern warfare. Most countries and military security alliances such as North Atlantic Treaty Organization (NATO) are now accepting the cyber assets and internet as “a vital national infrastructure, and a key driving force for national growth” [3].

Cybersecurity is generally defined as the preservation of confidentiality, integrity, and availability of information in cyberspace and consists of several concepts, respectively: information security, network security, internet security, critical information infrastructure security, etc. Information security is maintained by using endpoint protection tools such as Data Leakage Prevention (DLP) software and Endpoint Detection Response (EDR, XDR) tools such as antivirus software. DLP is used to detect and prevent unwanted breaches, exfiltration, or unwanted destruction of sensitive data. On the other hand, EDR and XDR provide detection, analytics, and response across endpoint clients, servers, networks, SIEM, and much more. Network security contains firewalls, denial of distributed service attack mitigation solutions, proxy servers, network security monitoring tools, etc. Internet security concerns with internet-related services.

Machine learning has been used to solve classification and regression problems in various domains such as image processing and classification tasks, natural language processing, autonomous vehicles, etc. In addition, one of the main application fields of machine learning algorithms is cybersecurity. Machine learning models are useful, especially where classical signature-based malware detection methods are insufficient to discover zero-day attacks or slightly changed versions of the known attacks [4]. Both classical machine learning techniques such as k-nearest neighbor, decision trees, support vector machines, etc., and state-of-art deep transfer learning models can be implemented for malware detection. In this review, deep learning models are the focal point since they skip the time-consuming feature extraction process and have surpassed the classical machine learning methods in image classification tasks [5]. Deep transfer models have been trained over large datasets. Those pre-trained models are very successful in image classification tasks and can be used for static analysis of malware [6]. There are some successful experiments made by using PyTorch and pre-trained models based on Generative Pre-trained Transformer 2 which performs over assembly codes for static malware analysis [7].

# Datasets

Machine learning methods perform better with a high amount of data, especially, the performance of deep neural networks gradually increases concerning the data size. The correct choice of the dataset is very important since the realistic representation of the real world in the training dataset increases the classification accuracy. Moreover, especially in deep learning models, the classification accuracy increases with respect to the data size.

## DARPA Intrusion Detection Dataset

It is one of the earliest datasets in network traffic. It was collected and published by Cyber System and Technology Group, formerly DARPA. This data consists of several weeks-long training and testing data with ground truth labels about intrusion detection assessment [8].

## KDD CUP 99 Dataset

KDD CUP 99 dataset is the most widely used and publicly available [9]. It is created by replicating the original DARPA 1998 dataset and contains 4 900 000 replicated attacks. Each sample contains 41 input features divided into two, respectively: basic features and higher-level features. The basic features are directly extracted or derived from the header information of IP packets and TCP/UDP segments in the tcpdump files of each session. High-level features contain several failed logins and root access attempts during the session.

## NSL-KDD Dataset

This dataset was derived from KDD CUP 99. NSL-KDD was formed in the 3rd international knowledge discovery and data mining contest. Each sample has network data characteristics of the TCP/IP model and is labeled with 22 different attack types and 4 main clusters.

# Traditional Machine Learning Algorithms in Cybersecurity

Traditional machine learning algorithms consist of k-nearest neighbor classifiers, support vector machines, decision trees, neural networks, and regression models. Traditional machine algorithms are not good to deal with big amounts of data. Depending on the throughput value, millions of packets go through over the network. So, this makes cyberspace a big data environment. Even though there is some research, traditional algorithms are not the best fit for cybersecurity.

## KNN

KNN is the acronym for the k-nearest neighbor classifier. This method is known as the lazy classifier since it doesn't require prior training and easy to implement. It is a supervised method and utilizes metric distances of known samples for predicting the new samples' labels. There are different types of metrics. But, in general, the Euclidian model, which is a very straightforward method is used to measure the distance between the new sample and its closest neighbors. KNN has been applied to network datasets. For example, in a study conducted by Sharifi et al., NSL-KDD dataset was divided into clusters for labeling, then, the KNN algorithm was implemented over the set samples [10].

## Support Vector Machines

Support vector classifiers are one the most successful and robust machine learning methods. It had been the state-of-art method before deep learning algorithms surpassed all the traditional machine learning algorithms. The support vectors are the closest points between different classes. In two-dimensional space, support vectors are lines but in three or higher dimensional spaces support vectors become hyperplanes that enable optimal separation. For datasets that are not linearly separable, some kernel functions are used to make them separable in a higher-dimensional space [11]. In an experiment conducted over KDD Cup 99 dataset, the effectiveness of an intrusion detector was tested with support vector machines [12]. In another work, KDD Cup 99 dataset is divided into subsets of attack types, respectively: DoS, Probe, R2L, and U2R [13]. In this work, the trained model has good accuracy metrics over test data but failed to detect the abnormal packets in the actual network.

## Decision Trees

Decision trees are supervised machine learning models which have decision nodes and leaves. Each leaf represents a category, and the leaves are the decisions or the outcomes. The main goal is to create a tree-like structure for predicting the value of a target variable. The most common forms of decision tree methods are ID3, C4.5, and CART. Unlike other methods, decision trees require little data preparation, and they can handle both numerical and categorical data. In a study conducted by Ingre et al., the NSL-KDD dataset was used with 14 features. These features were then used in a decision tree structure [14]. The developed model is designed as an intrusion detection system to detect five different classes. The accuracy rate for the five different classes is 83.7% and for the two different test data, 90.3% accuracy was acquired.

# Deep Learning Algorithms in Cybersecurity

Neural networks are based on a single perceptron unit. They were first proposed in 1957 and for a very long time, it was believed that perceptron-based networks can only solve linear problems just like a logistic regression classifier or linear support vector machines. But, in 1987, it was corrected that with the help of multi-layer perceptrons and activation functions such as sigmoid, exponential linear unit, etc. artificial neural networks can be used for non-linear problems like XOR problems [15]. In Fig. 1, it can be seen that multi-layer perceptrons can solve XOR problems. In this example, the muti-layer perceptron produces 0, and 1 for the input values of (0,0) and (1,1) respectively. All connections have a weight equal to 1, except the four connections where it is seen with red arrows.

Chart

Description automatically generated

Fig. 1: XOR classification problem and a multi-layer perceptron that solves it.

Artificial neural networks generally consist of three segments, the input layer, hidden layers, and output layer, respectively. The information taken from the input layer goes thorough by multiplication with the weight and bias values among the tensors in the hidden layer. In the output layer, a result is then produced. This process is called forward propagation. If the produced result is different than the ground-truth label, the error is compensated by refining each connection weight and each bias term. This process is called backpropagation. The backpropagation works well with the activation functions such as logistic function, hyperbolic tangent function, linear unit function, etc. since they enable gradient descent to make some progress at every step.

Before the 1990s, artificial neural networks with more than two hidden layers were considered deep. But after the increase in computational power, today’s networks have hundreds of hidden layers. Thus, the term “deep” is now being defined with dozens or hundreds of hidden layers. Convolutional neural networks followed by some dense layers or recurrent neural networks are the most commonly used forms of deep learning algorithms. In the following parts, the working mechanism of the convolutional networks, pre-trained models, and some of their implementation on cybersecurity datasets will be explained.

## Convolutional Neural Networks

Convolutional neural networks managed to achieve superhuman performance on visual tasks since we have big-sized classified image datasets and computation power for training deep networks. Convolution is an operator that slides one function over another to measure the integral of their element-wise multiplication. An example of the convolution operation is presented in Fig. 2. It is impossible to compute the convolution value of the edge elements, that’s why the convoluted image in the Fig. 2 is a 4 by 4 matrices. In general, to mitigate information loss, some zeros or ones are added to the edges for padding.

Graphical user interface, application

Description automatically generated

Fig. 2: Convolution operation over the grey levels of an image.

The images are fed to convolutional layers directly. Convolutional layers are used consecutively and through these layers, features are automatically extracted. At the end of the convoluted layers, the information is flattened and forwarded to neural layers such as recurrent type or fully connected dense type. The convolutional neural structure can easily be implemented with a few lines of code by using modern deep learning environments such as TensorFlow [16].

In a work carried out by Wang et al., network traffic is converted to image shape and sent to a convolutional neural network [17]. With the help of the convolutional layers, raw network traffic data is processed and some features extracted. In another study, the NSL-KDD dataset of log files is used as an image for feature extraction [18]. In another study, binary classification of the NSL-KDD dataset is also carried out by a one-dimensional convolutional network [19]. Samples in the NSL-KDD are also tested by converting them to an RGB-like formation, then fed to convolutional networks [20]. In this work, it is reported to achieve a better classification performance in RGB-like encoding than grayscale.

Several studies show the effectiveness of using convolutional neural networks and recurrent neural networks together [21]. In this study, first convolutional networks are used to map the network traffic samples of KDD99. Then, extracted features are fed to recurrent networks of type RNN and LSTM. The convolutional networks with LSTM configuration reached 99,7% accuracy in a two-class scenario of abnormal traffic versus normal traffic. In another study, convolutional networks and LSTM recurrent layers are used together for detecting the intrusions in the DARPA 1998 dataset [22].

## Using Pre-Trained Models with Transfer Learning

Deep learning algorithms have trained on very large datasets and the classification accuracy of the pre-trained models has already surpassed all other traditional algorithms. The idea of using a model trained on a different dataset is called transfer learning. Generally, this is achieved by first, training the model on a large dataset then, freezing a portion of the layers and adding some dense layers at the end as seen in Fig. 3.

Chart, line chart

Description automatically generated

Fig. 3: Transfer Learning structure, pre-trained part weights belong to the ResNet.

Transfer learning speeds up the training time and enables to use of deep networks on small-sized datasets where fine-tuning the optimal weights is not possible. Even though there are relatively large cyber datasets for network traffic and malware, they are still small when compared to image datasets such as ImageNet. Thus, using pre-trained models in malware detection scenarios produces effective classification accuracy.

In a work conducted by Li Z. et al., the features in NSL-KDD dataset were converted to binary vectors by using n-bits vectorization and created an eight-by-eight grayscale image for representing each sample [23]. Using pre-trained models of ResNet50 and GoogLeNet, they acquired a classification accuracy of 80% on NSL-KDD dataset. In another study, the VGGNet is used by converting KDD99 dataset samples into eleven-by-eleven image format [24]. The accuracy in this work is 98.34% over the KDD99 dataset.

In addition to studies conducted on public cybersecurity datasets, Intel Corporation carried out an experiment in which transfer learning is used on their malware dataset [25]. According to the published white paper, the method first converts application binaries to malware images. Conversion is achieved by reading eight bits and noticing them as pixel values between 0-255. Secondly, the malware images need to be resized since the transfer learning algorithms accept only inputs of fixed sizes. For example, ResNet expects images of size 224x224. It is stated that resizing does not cause information loss if bilinear interpolation or nearest-neighbor algorithms are used.

In the evaluation part, a Microsoft dataset of 2.2 million hashes of malware binaries and 10 columns of data information is used. The dataset is divided into three, respectively; 60% for training, 20% for validation, and 20% for testing. To mitigate the overfitting, the early stopping parameter aborted the fit process after the 10th epoch. It is stated in the study that the classification result on the test set is 99,07% with a 2,58% false-positive rate.

# Discussion and Conclusion

Classical malware detection relies on signature-based detection. The signature matching method is not suitable since these kinds of algorithms can not detect zero-day attacks. Moreover, the exponentially growing numbers of signatures make this process cumbersome. Thus, using deep learning in cybersecurity enhances classification accuracy and helps to create more agile products. Malware analysis consists of two approaches, static and dynamic analysis. Static analysis can suffer from the obfuscation of the codes and dynamic analysis can be time-consuming. The malware analysis tool which has both static and dynamic analysis capacity can perform better.

As a result, machine learning techniques have been used to detect abnormal network traffic, malware, etc. Especially, deep learning algorithms can detect abnormalities with a very high accuracy rate. The pre-trained models of computer vision field such as GoogLeNet, ResNet, etc. have been used in the detection of the image-like representations of malware and the classification accuracy of those models are promising.

# References

1. Individuals Using Internet Worldwide, https://data.worldbank.org/indicator/IT.NET.USER.ZS
2. Luong, N.C. et al, “Data Collection and Wireless Communication in Internet of Things (IoT) Using Economic Analysis and Pricing Models; A survey”, *Proceeding IEEE Community Surveys Tuts., Vol. 18*, 2016, pp. 2546-2590
3. Klimburg A., National Cyber Security Framework Manual. NATO CCD COE Publication Report, 2012
4. Handa, A. et al., Machine Learning in Cybersecurity. WIREs Data Mining Knowledge Discovery 2019, 9, 1306, https://doi.org/10.1002/widm.1306
5. Xin, Y. et al., Machine Learning and Deep Learning Methods for Cybersecurity
6. Chen L., et al., STAMINA: Scalable Deep Learning Approach for Malware Classification, White Paper, 2020
7. Şahin, N., Malware Detection Using Transformers-Based Model GPT-2, M.S. thesis, Middle East Technical University, Ankara, 2021
8. R. P. Lippmann et al., “Evaluating Intrusion Detection Systems: The 1998 DARPA off-line Intrusion Detection Evaluation” *Proceeding DARPA Inf. Survivability Conf. Expo. (DISCEX), vol. 2,* 2000, pp. 12-26
9. Revathi S. and Malathi, A., “A Detailed Analysis on NSL-KDD Dataset Using Various Machine Learning Techniques for Intrusion Detection” *Proceeding Int. J. Eng. Res. Technol.,* 2013, pp. 1848-1853
10. Sharifi, A.M., et al., Intrusion Detection Based on Joint of K-means and KNN. J. 2015, 10, 42-51,
11. Sharma, R. K., et al., “Analysis of Machine Learning Techniques-Based Intrusion Detection Systems” *Proceeding Int. Conf. Adv. Comput., Netw., Inform.,* 2016, pp. 485 493
12. Kotpalliwar, M. V. and Wajgi, R. “Classification of Attacks Using Support Vector Machine (SVM) on KDDCUP'99 IDS database” *Proceeding Int. Conf. Commun. Syst. Netw. Technol.,* 2015, pp. 987 990
13. A. M. Chandrasekhar and K. Raghuveer, “Confederation of FCM clustering, ANN and SVM techniques to implement hybrid NIDS using corrected KDD CUP 99 dataset” *Proceeding. Int. Conf. Commun. Signal Process., 2*014, pp. 672 676
14. Ingre, B., et al., “Decision Tree-Based Intrusion Detection System for NSL-KDD Dataset” *Proceeding* *Int. Conf. Inf. Commun. Technol. Intell. Syst.,* 2017, pp. 207-218
15. Minsky, M., and Papert, S.A., Perceptrons, Expanded Edition, MIT Press, Massachusetts, 1987
16. Geron, A., *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems,* O'Reilly Media, Inc., 2019
17. Wang, W. et al., “Malware Traffic Classification Using Convolutional Neural Network for Representation Learning'' *Proceeding Int. Conf. Inf. Netw.,* 2017, pp. 712-717.
18. Gayathri, R.G., et al., Image-Based Feature Representation for Insider Treat Classification, Applied Science, 2020, 10, 4945, doi:10.3390/app10144945
19. Song, J., et al., Statistical Analysis Of Honeypot Data and Building of Kyoto 2006+Dataset for NIDS Evaluation,Association for Computing Machinery, 2011, 29-36, doi:10.1145/1978672
20. Kim, T., et al., “An Encoding Technique for CNN-Based Network Anomaly Detection”, *Proceeding IEEE Int. Conf. Big Data, 2018,* pp 2960-2965
21. Vinayakumar, R., et al., “Applying Convolutional Neural Networks for Intrusion Detection”, *Proceeding Int. Conf. Adv. Comput. Commun. Informat., 2017,* pp 1222-1228
22. Wang, W., et al., Learning Hierarchical Spatial Temporal Features Using Deep Neural Networks to Improve Intrusion Detection. IEEE Access, 2018, 6, 1792-1806, https://doi.org/10.1109/ACCESS.2017.2780250.
23. Li. Z., et al., Intrusion Detection Using Convolutional Neural Networks Representation Learning. Neural Information Processing, 2017, 858-866. https://doi.org/[10.1007/978-3-319-70139-4\_87](http://dx.doi.org/10.1007/978-3-319-70139-4_87)
24. Zhang, L., et al., “An Improved Network Intrusion Detection Based-on Deep Neural Network”, *Proceeding IOP Conf. Ser. Mater. Sci. Eng.,* 2019, 563
25. Chen, L., “Deep transfer learning for static malware classification” arXiv preprint, 2018