ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

A Survey on Pecnnormance Evaluation of Supervised Machine Learning algorithms for Intrusion Detection

Yash Oza[1], DishaThalwar[2], Mrunalinee Patole[3]

U.G. Student, Department of Computer Engineering, R.M.D Sinhgad School of Engineering Warje, Pune , India[1,2]

Associate Professor, Department of Computer Engineering, R.M.D Sinhgad School of Engineering Warje, Pune , India[3]

ABSTRACT: Intrusion detection system plays a vital role in network security. Intrusion detection model may be a prophetic model accustomed predicts the network knowledge traffic as traditional or intrusion. Machine Learning algorithms square measure accustomed build correct models for cluster,Classification and prediction. During this paper classification and prophetic models for intrusion detection square measure engineered by victimization machine learning classification algorithms specifically provision Regression, Gaussian Naive mathematician, Support Vector Machine and Random Forest. These algorithms square measure tested with NSL-KDD knowledge set. Experimental results shows that Random Forest Classifier out performance other strategies in characteristic whether or not the info traffic is traditional or associate attack.

KEYWORDS: Deep and machine learning, intrusion detection, Auto-encoders, Network security, Novel Approach, RCNN.

I. INTRODUCTION

One of the major challenges in network security is the provision of a robust and effective Network Intrusion Detection System (NIDS). Despite the significant advances in NIDS technology, the majority of solutions still operate using less-capable signature-based techniques, as opposed to anomaly detection techniques. The current issues are the existing techniques leads to ineffective and inaccurate detection of attacks. There are three main limitations like, volume of network data, in-depth monitoring and granularity required to improve effectiveness and accuracy and finally the number of different protocols and diversity of data traversing. The main focus of NIDS research has been the application of machine learning and shallow learning techniques. The initial deep learning research has demonstrated that its superior layer-wise feature learning can better or at least match the performance of shallow learning techniques. It is capable of facilitating a deeper analysis of network data and faster identification of any anomalies. In this paper, we propose a novel deep learning model to enable NIDS operation within modern networks.

II. RELATED WORK

The paper [1] focuses on deep learning strategies that square measure impressed by the structure depth of human brain learn from lower level characteristic to higher levels construct. it's due to abstraction from multiple levels, the Deep Belief Network (DBN) helps to be told functions that square measure mapping from input to the output. the method of learning doesn't enthusiastic about human-crafted options. DBN uses AN unattended learning algorithmic rule, a Restricted Boltzmann Machine (RBM) for every layer. blessings are: Deep writing is its ability to adapt to dynamical contexts regarding knowledge that ensures the technique conducts thoroughgoing knowledge analysis. Detects abnormalities within the system that has anomaly detection, traffic identification. Disadvantages are: Demand for quicker and economical knowledge assessment.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

The main purpose of [2] paper is to review and summarize the work of deep learning on machine health observance. The applications of deep learning in machine health observance systems square measure reviewed principally from the subsequent aspects: Autoencoder (AE) and its variants, Restricted {boltzmann|Boltzmann|LudwigBoltzmann|physicist} Machines and its variants together with Deep Belief Network (DBN) and Deep Boltzmann Machines (DBM), Convolutional Neural Networks (CNN) and continual Neural Networks (RNN). blessings are: DL-based MHMS don't need in depth human labor and professional data. The applications of deep learning models aren't restricted to specific varieties of machines. Disadvantages are: The peCNNormance of DL-based MHMS heavily depends on the dimensions and quality of datasets.

Proposes the utilization of a stacked denoisingautoencoder (SdA), that may be a deep learning algorithmic rule, to determine AN FDC model for synchronic feature extraction and classification. The SdA model [3] will establish international and invariant options within the device signals for fault observance and is strong against mensuration noise. AN SdA is consisting of denoisingautoencoders that square measure stacked layer by layer. This multilayered design is capable of learning international options from advanced input file, like variable time-series datasets and high-resolution pictures. blessings are: SdA model is helpful in real applications. The SdA model proposes effectively learn traditional and fault-related options from device signals while not preprocessing. Disadvantages are: got to investigate a trained SdA to spot the method parameters that almost all considerably impact the classification results.

Proposes a completely unique deep learning-based continual neural networks (RNNs) model [4] for automatic security audit of short messages from prisons, which may classify short messages(secure and non-insecure). during this paper, the feature of short messages is extracted by word2vec that captures ordination info, and every sentence is mapped to a feature vector. particularly, words with similar that means square measure mapped to the same position within the vector house, so classified by RNNs. blessings are: The RNNs model achieves a median ninety two.7% accuracy that is over SVM. Taking advantage of ensemble frameworks for group action totally different feature extraction and classification algorithms to spice up the general peCNNormance. Disadvantages are: it's apply on solely short messages not large-scale messages.

Signature-based options technique as a deep convolutional neural network [5] during a cloud platform is projected for plate localization, character detection and segmentation. Extracting vital options makes the LPRS to adequately acknowledge the car place during a difficult scenario like i) engorged traffic with multiple plates within the image ii) plate orientation towards brightness, iii) further info on the plate, iv) distortion thanks to wear and tear and v) distortion regarding captured pictures in weather like as hazy pictures. blessings are: the prevalence of the projected algorithmic rule within the accuracy of recognizing platter instead of different ancient LPRS. Disadvantages square measure: There are some unrecognized or miss-detection pictures.

In [6] paper, a deep learning approach for anomaly detection employing a Restricted Boltzmann Machine (RBM) and a deep belief network square measure enforced. This methodology uses a one-hidden layer RBM to peCNNorm unattended feature reduction. The resultant weights from this RBM square measure passed to a different RBM manufacturing a deep belief network. The pre-trained weights square measure passed into a fine calibration layer consisting of a supplying Regression (LR) classifier with multi-class soft-max. blessings are: Achieves ninety seven.9% accuracy. It produces an occasional false negative rate of two.47%. Disadvantages are: got to improve the tactic to maximise the feature reduction method within the deep learning network and to enhance the dataset.

The paper [7] proposes a deep learning based mostly approach for developing AN economical and versatile NIDS. A distributed autoencoder and soft-max regression based mostly NIDS was enforced. Uses Self-taught Learning (STL), a deep learning based mostly technique, on NSL-KDD - a benchmark dataset for network intrusion. blessings are: STL achieved a classification accuracy rate quite ninety eight for every type of classification. Disadvantages are: got to implement a time period NIDS for actual networks victimization deep learning technique.

In [8] paper select multi-core CPU's moreover as GPU's to judge the peCNNormance of the DNN based mostly IDS to handle Brobdingnagian network knowledge. The parallel computing capabilities of the neural network build the Deep Neural Network (DNN) to effectively check the network traffic with AN accelerated peCNNormance. blessings are: The DNN based mostly IDS is reliable and economical in intrusion detection for distinguishing the particular attack categories with needed range of samples for coaching. The multicore CPU's was quicker than the serial coaching mechanism. Disadvantages are: got to improve the detection accuracies of DNN based mostly IDS.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

In [9] paper, proposes a mechanism for sleuthing giant scale network-wide attacks victimization Replicator Neural Networks (RNNs) for making anomaly detection models. Our approach is unattended and needs no labeled knowledge. It conjointly accurately detects network-wide anomalies while not presuming that the coaching knowledge is totally freed from attacks. blessings are: The projected methodology is ready to with success discover all outstanding DDoS attacks and SYN Port scans injected. projected methodology is resilient against learning within the presence of attacks, one thing that connected work lacks. Disadvantages are: got to improve projected methodology by victimization stacked autoencoder deep learning techniques.

Based on the flow-based nature of SDN, we have a tendency to propose a flow-based anomaly detection system victimization deep learning. In [10] paper, apply a deep learning approach for flow-based ANomaly detection in an SDN setting. blessings are: It finds AN best hyper-parameter for DNN and confirms the detection rate and warning rate. The model gets the peCNNormance with accuracy of seventy five.75% that is sort of affordable from simply victimization six basic network options. Disadvantages are: it'll not work on real SDN setting.

III. SYSTEM OVERVIEW

In this paper, propose a novel deep learning model to enable NIDS operation within modern networks. The model proposes is a combination of deep and shallow learning, capable of correctly analyzing a wide-range of network traffic. More specifically, we combine the power of stacking our proposed Non-symmetric Deep Auto-Encoder (NDAE) (deep learning) and the accuracy and speed of Random Forest (FAST-RCNN) (shallow learning). This paper introduces our NDAE, which is an auto-encoder featuring non-symmetrical multiple hidden layers. NDAE can be used as a hierarchical unsupervised feature extractor that scales well to accommodate high-dimensional inputs. It learns non-trivial features using a similar training strategy to that of a typical auto-encoder. Stacking the NDAEs offers a layer-wise unsupervised representation learning algorithm, which will allow our model to learn the complex relationships between different features. It also has feature extraction capabilities, so it is able to refine the model by prioritizing the most descriptive features.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

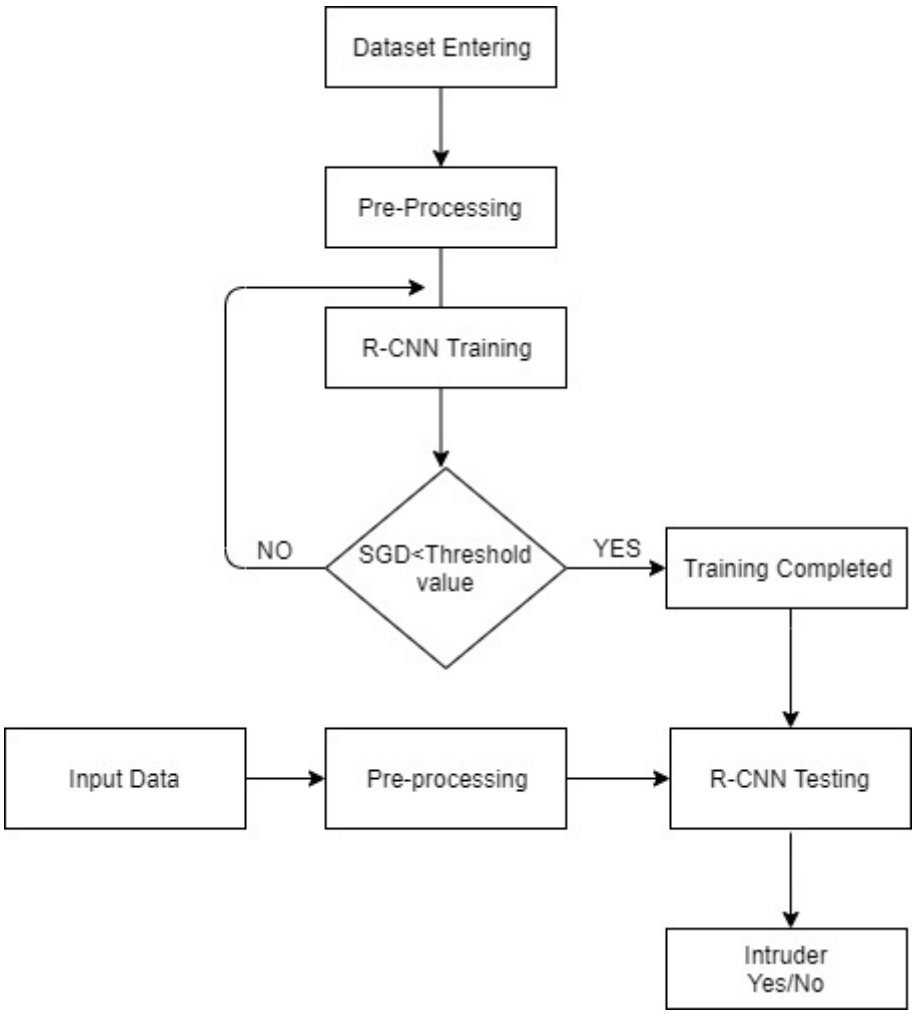

Fig. 1 Proposed System Architecture

The proposed system architecture of Network Intrusion Detection and Prevention System (NIDPS). The input traffic data is uses for NSL KDD dataset with 41 features. The training dataset contains data preprocessing which includes two steps: Data transformation and data normalization. After uses two NDAEs arranged in a stack, which uses for selecting number of features.After that apply the Random Forest Classifier for attack detection. Intrusion Prevention Systems (IPS) contains IDS functionality but more sophisticated systems which are capable of taking immediate action in order to prevent or reduce the malicious behavior.

Advantages are:

- Due to deep learning technique, it improves accuracy of intrusion detection system.
- The network or computer is constantly monitored for any invasion or attack.
- The system can be modified and changed according to needs of specific client and can help outside as well as inner threats to the system and network.
- It effectively prevents any damage to the network.
- Any alterations to files and directories on the system can be easily detected and reported.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

IV. ALGORITHM

Let's quickly summarize the different algorithms in the R-CNN family (R-CNN, Fast R-CNN, and Faster R-CNN). This will help lay the ground for our implementation part later when we will predict the bounding boxes present in previously unseen new data.

R-CNN extracts a bunch of regions from the given data using selective search, and then checks if any of these boxes contains an object. We first extract these regions, and for each region, CNN is used to extract specific features. Finally, these features are then used to detect objects. Unfortunately, R-CNN becomes rather slow due to these multiple steps involved in the process.

Fast R-CNN, on the other hand, passes the entire data to ConvNet which generates regions of interest (instead of passing the extracted regions from the data). Also, instead of using three different models (as we saw in R-CNN), it uses a single model which extracts features from the regions, classifies them into different classes, and returns the bounding boxes.

All these steps are done simultaneously, thus making it execute faster as compared to R-CNN. Fast R-CNN is, however, not fast enough when applied on a large dataset as it also uses selective search for extracting the regions.

Faster R-CNN fixes the problem of selective search by replacing it with Region Proposal Network (RPN). We first extract feature maps from the input data using ConvNet and then pass those maps through a RPN which returns object proposals. Finally, these maps are classified and the bounding boxes are predicted.

Summarized below the steps followed by a Faster R-CNN algorithm:

- Take an input data and pass it to the ConvNet which returns feature maps for the data
- Apply Region Proposal Network (RPN) on these feature maps and get object proposals
- Apply ROI pooling layer to bring down all the proposals to the same size.

Data Exploration

It's always a good idea to first explore the data we have. This helps us not only unearth hidden patterns, but gain a valuable overall insight into what we are working with. D

Dataset are:

Train.csv: Contains the name, class for each record. There can be multiple rows for one record a single record can have more than one object.

V. CONCLUSION

The problems faced by existing NIDS techniques. In response to this we have proposed our novel NDAE method for unsupervised feature learning. We have then built upon this by proposing a novel classification model constructed from stacked NDAEs and the FAST-RCNN classification algorithm. Also we implemented the Intrusion prevention system. The result shows that our approach offers high levels of accuracy, precision and recall together with reduced training time. The proposed NIDS system is improved only 5% accuracy. So, there is need to further improvement of accuracy. And also further work on real-time network traffic and to handle zero-day attacks.

REFERENCES

1. B. Dong and X. Wang, "Comparison deep learning method to traditional methods using for network intrusion detection," in Proc. 8th IEEE Int.Conf. Commun. Softw. Netw, Beijing, China, Jun. 2016, pp. 581–585.
2. R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring: A survey," Submitted to IEEE Trans. Neural Netw. Learn. Syst., 2016. [Online]. Available: http://arxiv.org/abs/1612.07640
3. H. Lee, Y. Kim, and C. O. Kim, "A deep learning model for robust wafer fault monitoring with sensor measurement noise," IEEE Trans. Semicond. Manuf., vol. 30, no. 1, pp. 23–31, Feb. 2017.
4. L. You, Y. Li, Y. Wang, J. Zhang, and Y. Yang, "A deep learning based RNNs model for automatic security audit of short messages," in Proc. 16th Int. Symp. Commun. Inf. Technol., Qingdao, China, Sep. 2016, pp. 225–229.
5. R. Polishetty, M. Roopaei, and P. Rad, "A next-generation secure cloud based deep learning license plate recognition for smart cities," in Proc. 15th IEEE Int. Conf.Mach. Learn. Appl., Anaheim, CA, USA, Dec. 2016, pp. 286–293.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

6. K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," in Proc. 15th IEEE Int. Conf. Mach. Learn. Appl., Anaheim, CA, USA, Dec. 2016, pp. 195–200.
7. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in Proc. 9th EAI Int.Conf. Bio-Inspired Inf. Commun. Technol., 2016, pp. 21–26. [Online]. Available: http://dx.doi.org/10.4108/eai.3-12-2015.2262516
8. S. Potluri and C. Diedrich, "Accelerated deep neural networks for enhanced intrusion detection system," in *Proc. IEEE 21st Int. Conf. Emerg. Technol. Factory Autom.*, Berlin, Germany, Sep. 2016, pp. 1–8.
9. C. Garcia Cordero, S. Hauke, M. Muhlhauser, and M. Fischer, "Analyzing flow-based anomaly intrusion detection using replicator neural networks," in *Proc. 14th Annu. Conf. Privacy, Security. Trust*, Auckland, New Zeland, Dec. 2016, pp. 317–324.
10. T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for network intrusion detection in software defined networking," in *Proc. Int. Conf. Wireless Netw. Mobile Commun.*,Oct. 2016, pp. 258–263.