DOI:10.48047/IJFANS/V11/I12/198

Intrusion Detection Using an Ensemble Deep Learning Approach

Dr. V. Ramachandran[1], Professor and HOD, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Yamparala Anuhya[2], **Vutukuri Venkata Lakshmi Raga Pravallika**[3], **Vamsi Makke**[4], **Vasireddy Venkata Leela Sai Srikar**[5]
[2,3,4,5]UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
[1] vrc.bhatt@vvit.net, anuhyayamparala@gmail.com[2],
pravallikavutukuri365@gmail.com[3], vamsimakke@gmail.com[4],
srikarvasireddy5@gmail.com[5]

Abstract

Today's cyber society faces a serious intrusion detection security issue. Recent years have seen a sharp rise in network intrusion attacks, raising severe privacy and security concerns. The complexity of cyber-security threats is increasing due to technological improvement, making it impossible for the current detection methods to handle the problem. So, creating an intelligent and efficient network intrusion detection system would be crucial to resolving this problem. In this paper, we created an intelligent intrusion detection system that can detect different networking attacks using deep learning approaches, specifically Convolutional Neural Networks (CNN) and Deep Neural Networks (DNN). We used an ensemble model of CNN and DNN which provides us with great accuracy. Before being used for model training and testing, the obtained data is analysed and pre-processed. Also, in order to choose the optimum model for the network intrusion detection system, we compared the outcomes of our proposed solution and evaluated the performance of the proposed solution using several evaluation matrices.

Keywords: Convolutional Neural Network, Deep Learning, Deep Neural Network, Intrusion Detection, Network Security

1. Introduction

New and sophisticated cyberattacks are being utilised to get around defences as technology develops, taking advantage of vulnerabilities and other unethical behaviour. Network infrastructure is one of the primary targets of numerous cyberattacks, including Denial of Service (DoS) and Distributed Denial-of-Service (DDoS) attacks, TCP SYN Flood assaults, Ping of Death attacks, Teardrop attacks, Scan attacks, etc. In order to stop these attacks, ensure the network's safety and security, and maintain high availability for the network's genuine users, it has been seen that considerable effort was put into identifying and putting various approaches and tactics into practise.

A well-known method for enhancing network security is the introduction of intrusion detection systems. In network intrusion detection system, the detection system continuously monitors all hosts' incoming and outgoing network traffic. and, based on specific criteria, it can identify the attack and detect it. The danger of network damage is then significantly reduced as the system takes the required security precautions to stop or prevent the assault. Unfortunately, the present NIDS solutions are unable to adequately handle this issue because of the quick development in the complexity of cyber-security assaults. Therefore, in this paper, we aim to develop an intelligent detection system using the popular deep learning algorithms such as DNN and CNN such that they are highly capable enough in recognizing and differentiating between different network intrusion attacks.

2.Literature Survey

In earlier investigations, numerous authors developed efficient intrusion detection systems using a variety of methodologies. In [1] and [2],authors developed an intrusion detection system by performing with some ML algorithms and DL algorithms there by differentiating the accuracy of ML and DL algorithms. Whereas the authors in [3] developed an intrusion detection system using convolutional recurrent neural networks. To enhance the performance and prediction of the ID system, they employed RNN to capture temporal characteristics and CNN to conduct convolution to gather local featuresin which the dataset they used is CSE-CIC-DS2018. However, in [4], used algorithms such as CNN, RNN-LSTM, RNN-GRU and evaluated the performance of each, thereby concluding that CNN is the best algorithm to develop an intelligent IDS whereas authors in [5] compared all the deep learning algorithms and concluded that DNN is the best one.

In addition, authors in [6] proposed two deep neural networks (DNN) with multiple fully connected layers (Multilayer Perceptron) in order to classify the network traffic of an SME into normal or malicious for DDoS and malware threats. By conducting intrusion detection training on numerous data sets, authors in [7] demonstrated how the DBN-based feature learning approach outperforms the standard feature learning approach.

Authors in [8] and [9]provided a thorough investigation of intrusion detection systems and CNN-based IDS techniques whereas authors in [10] implemented IDS using CNN algorithm. [11-19]

3. Existing Methodology

The development of intrusion detection systems (IDS) that can quickly and automatically identify and categorise cyberattacks at the host- and network-levels has made substantial use of machine learning techniques. Hence, a variety of conventional machine learning techniques were developed in order to create smart intrusion detection systems. The algorithms include like SVM, Random forest, Naive bayes, Bayesian Belief Networks etc.. But all these classical machine learning methods were failed to predict dynamic attacks and they should be trained in advance to detect such attacks. Later, many ensemble models of machine learning were developed to obtain higher accuracy. But all those models were failed to produce the best accuracy than deep learning algorithms. The publicly available malware datasets ought to be upgraded and standardised often because of the ever-evolving nature of malware and its rapidly changing attacking mechanisms. Since harmful threats are happening at an extremely high pace and are constantly changing, a comprehensive approach is needed.

The machine learning algorithms SVM produces around 50, Naïve Bayes produces around 50, Random Forest produces around 77 in terms of accuracy. Not only these famous ML algorithms, but also no ML algorithm had achieved the accuracy exceeding 80 percent of accuracy.

Some of the disadvantages of the existing system are namely:

a. They cannot predict the attack if attacker introduces any change in the attack parameters.
b. They should be well-trained in advance before prediction.
c. Less accuracy when compared to deep learning algorithms.

4. Proposed Methodology

After studying and reviewing many papers, we have developed a model based on DNN and an ensemble model of both DNN and CNN to outperform traditional machine learning algorithms in terms of accuracy.

Convolutional Neural Networks (CNN) have exhibited promising outcomes in the field of attack detection, which deals with detecting and blocking illegal access to computer networks. CNNs are a subset of deep learning algorithms with the capacity to automatically detect data-related trends and make use of such patterns to classify new data.

The CNN's performance in detecting intrusions is significantly influenced by its architectural design. In a typical CNN architecture for intrusion detection, pooling layers are added to a number of convolutional layers to reduce the dimensionality of the data. Following the pooling layers, fully linked layers that carry out the classification process receive the output.

Deep Neural Networks, which are similar to Convolutional Neural Networks are a type of deep learning algorithm that enable independent feature learning from data. DNNs however can demand more training data and processing resources than CNNs since they are often more complicated.

To improve the overall performance of the system, one strategy known as ensemble learning integrates different machine learning or deep learning algorithms. In this study, we combined CNN and DNN to create an ensemble model. The benefits of the CNN and DNN models can be combined to improve the system's overall performance. CNNs excel at extracting spatial components from data, while DNNs are better at capturing temporal information. By combining the two models, the ensemble may be able to capture more characteristics and perform better than either model operating alone.

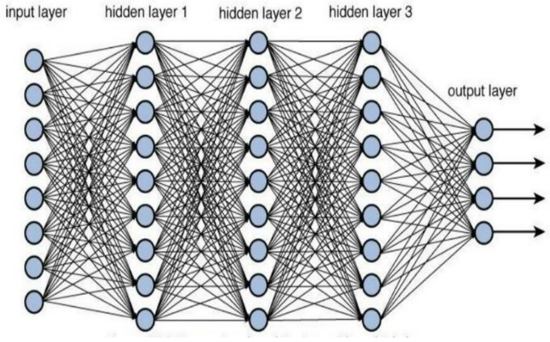


Figure 1: Architecture of Neural Network

Figure 1 is how a neural network looks like with two or more hidden layers.

5. Implementation

The proposed intrusion detection systems uses the concepts of two deep learning algorithms namely DNN and an ensemble of DNN and CNN.

A sufficient amount of dataset that closely reflects the real-world environment is needed to train and test the IDS with Deep Learning algorithm. NSL-KDD, a standard dataset is used in our project for training and testing purposes. Out of all our dataset, 80% of the data is allocated to the training purposes and 20% of the data is allocated to testing and validation purposes. In the NSL-KDD dataset, each record encompasses 41 attributes of different features and a label to identify the type of attack.

This is how the implementation takes place:

After uploading a dataset, a DNN model is created to extract the features from the test data. Then, the DNN features were splitted into train and tests and a convolutional2D layer is added to the CNN model with 32 neurons to filter the dataset 32 times. After defining output and prediction layers, CNN is used to predict the accuracy.

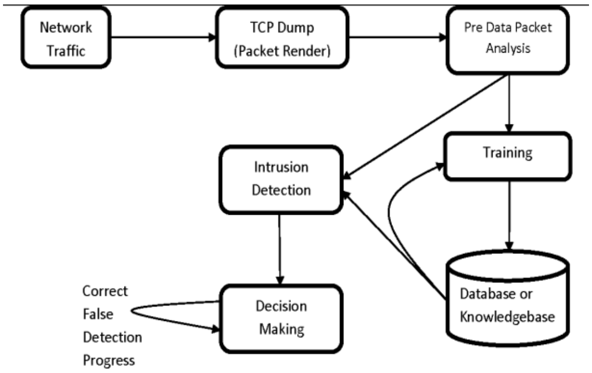


Fig 2. System Architecture

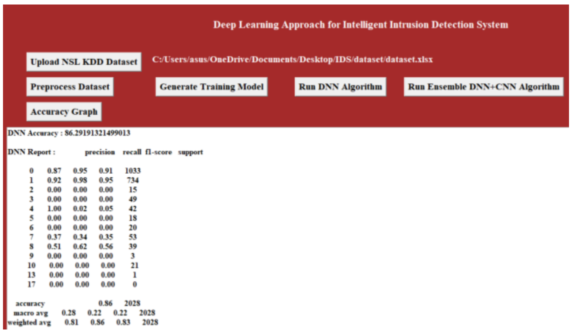
The modules in the implementation phase were:

- Upload NSL KDD dataset: using this module we will load data into system
- Process dataset: This module is used for preprocessing the data.
- Generate training model: using this module training model is generated on loaded data
- Run DNN algorithm – Here, DNN algorithm is executed by providing its accuracy, confusion matrix and ROC AUC curve.
- Run Ensemble DNN+CNN algorithm – This module helps in running this ensemble model thereby providing its accuracy, confusion matrix and ROC AUC curve.
- Accuracy graph: This module provides the algorithms accuracy comparison graph.

6. Results and Discussions

In this experiment, after uploading the dataset, the data is pre-processed and a training model is generated. Then, DNN algorithm is executed thereby providing its accuracy and performance metrics such as precision, recall, f1-score and support for the network layers.

As shown in Figure 3, the accuracy achieved for DNN algorithm is 86%.

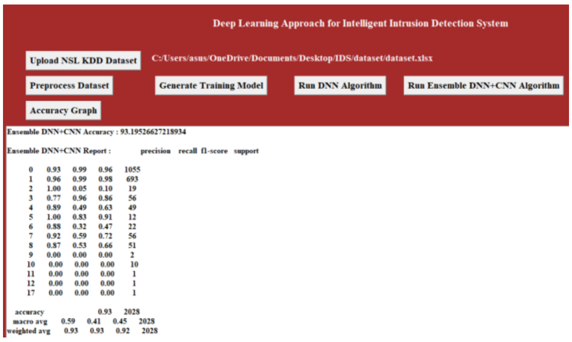

Deep Learning Approach for Intelligent Intrusion Detection System
Upload NSL KDD Dataset
C:/Users/asus/OneDrive/Documents/Desktop/IDS/dataset/dataset.xlsx
Preprocess Dataset
Generate Training Model
Run DNN Algorithm
Run Ensemble DNN+CNN Algorithm
Accuracy Graph

```
DNN Accuracy : 86.29191321499013

DNN Report :          precision   recall  f1-score   support

      0     0.87    0.95    0.91    1033
      1     0.92    0.98    0.95     734
      2     0.00    0.00    0.00      15
      3     0.00    0.00    0.00      49
      4     1.00    0.02    0.05      42
      5     0.00    0.00    0.00      18
      6     0.00    0.00    0.00      20
      7     0.37    0.34    0.35      53
      8     0.51    0.62    0.56      39
      9     0.00    0.00    0.00       3
     10     0.00    0.00    0.00      21
     13     0.00    0.00    0.00       1
     17     0.00    0.00    0.00       0

   accuracy                 0.86    2028
  macro avg     0.28    0.22    0.22    2028
weighted avg    0.81    0.86    0.83    2028
```

Figure3: DNN performance metrics

After the execution of DNN algorithm, ensemble model is executed thereby providing its accuracy and performance metrics respectively.

As shown in Figure 4, the accuracy achieved for ensemble model is 93% which is actually greater than the DNN algorithm.


Deep Learning Approach for Intelligent Intrusion Detection System
Upload NSL KDD Dataset
C:/Users/asus/OneDrive/Documents/Desktop/IDS/dataset/dataset.xlsx
Preprocess Dataset
Generate Training Model
Run DNN Algorithm
Run Ensemble DNN+CNN Algorithm
Accuracy Graph

```
Ensemble DNN+CNN Accuracy : 93.19526627218934

Ensemble DNN+CNN Report :          precision   recall  f1-score   support

      0     0.93    0.99    0.96    1055
      1     0.96    0.99    0.98     693
      2     1.00    0.05    0.10      19
      3     0.77    0.96    0.86      56
      4     0.89    0.49    0.63      49
      5     1.00    0.83    0.91      12
      6     0.88    0.32    0.47      22
      7     0.92    0.59    0.72      56
      8     0.87    0.53    0.66      51
      9     0.00    0.00    0.00       2
     10     0.00    0.00    0.00      10
     11     0.00    0.00    0.00       1
     12     0.00    0.00    0.00       1
     17     0.00    0.00    0.00       1

   accuracy                 0.93    2028
  macro avg     0.59    0.41    0.45    2028
weighted avg    0.93    0.93    0.92    2028
```

Figure 4: Ensemble DNN+CNN performance metrics

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

InFigure 5, x-axis represents predicted labels and y-axis represents true Labels and all counts in diagonal for predicted and true matching labels represents correct prediction count and other boxes represents incorrect count.

InFigure 6, x-axis represents False positive rate and y-axis represents true positive rate and if blue line comes on top of orange line then predictions are correct and if come down then predictions are incorrect and in above graph we can see only few predictions are incorrect.

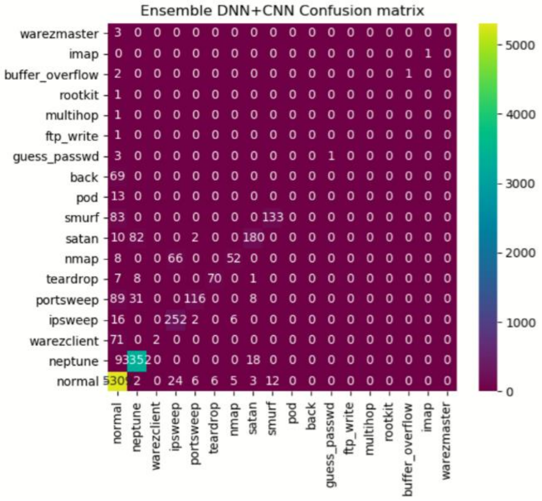


Figure 5: Ensemble DNN+CNN confusion matrix

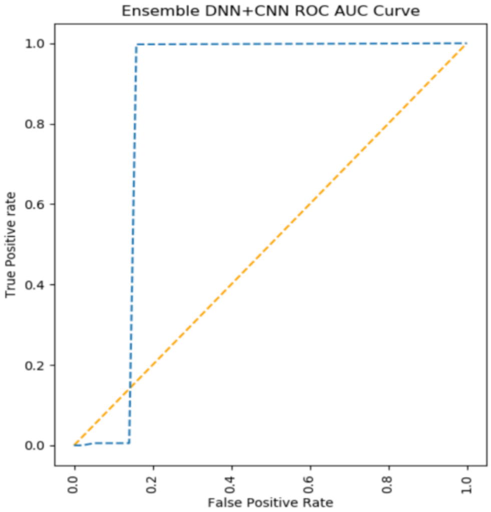


Figure 6: Ensemble ROC AUC Curve

To plot the difference between DNN algorithm and ensemble algorithm, we also presented a bar graph which is seen in fig 5. This actually provides a great understanding that how these algorithms achieved their respective accuracies in developing an intelligent intrusion detection system.

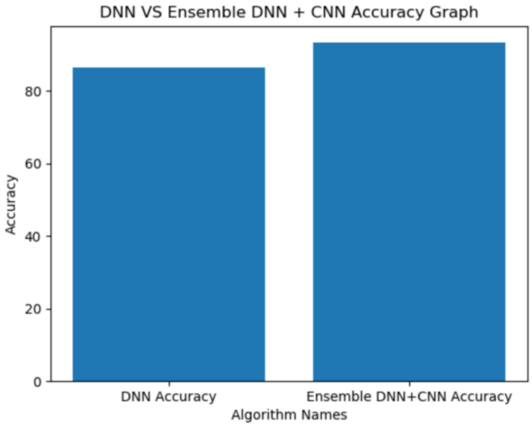


Figure 7: Graph of comparison between DNN and Ensemble DNN+CNN accuracy

7. Conclusion

In this paper, we have developed an intrusion detection model using an ensemble DNN and CNN in which DNN is used to extract features and CNN is used to predict on test data. However, this model required large amount of training data buthas exhibited great potential in the effort to detect intrusions. We would like to extend this project by modelling with multiple ensemble techniques to improve both performance and accuracy.

References

[1]. Ahmad Z, Shahid Khan A, Wai Shiang C, Abdullah J, Ahmad F. Network intrusion detection system: A systematic study of machine learning and deep learning approaches. Trans Emerging Tel Tech. 2020;e4150.

[2]. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," in IEEE Access, vol. 7, pp. 41525-41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[3]. Khan, M.A. HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System. Processes 2021, 9, 834. https://doi.org/ 10.3390/pr9050834

[4]. S. Al-Emadi, A. Al-Mohannadi and F. Al-Senaid, "Using Deep Learning Techniques for Network Intrusion Detection," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 171-176, doi: 10.1109/ICIoT48696.2020.9089524.

[5]. S. Osken, E. N. Yildirim, G. Karatas and L. Cuhaci, "Intrusion Detection Systems with Deep Learning: A Systematic Mapping Study," 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), Istanbul, Turkey, 2019, pp. 1-4, doi: 10.1109/EBBT.2019.8742081.

[6]. Chamou et al., "Intrusion Detection System Based on Network Traffic Using Deep Neural Networks," 2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), 2019, pp. 1-6, doi: 10.1109/CAMAD.2019.8858475.

[7]. W. Peng, X. Kong, G. Peng, X. Li and Z. Wang, "Network Intrusion Detection Based on Deep Learning," 2019 International Conference on Communications, Information System and Computer Engineering (CISCE), Haikou, China, 2019, pp. 431-435, doi: 10.1109/CISCE.2019.00102.

[8]. Mohammadpour, Leila & Ling, Teck Chaw & Liew, Chee Sun &Aryanfar, Alihossein. (2022). A Survey of CNN-Based Network Intrusion Detection. Applied Sciences. 12. 8162. 10.3390/app12168162.

[9]. Khraisat, A., Gondal, I., Vamplew, P. *et al.* Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecur 2, 20 (2019). https://doi.org/10.1186/s42400-019-0038-7

[10]. L. Chen, X. Kuang, A. Xu, S. Suo and Y. Yang, "A Novel Network Intrusion Detection System Based on CNN," 2020 Eighth International Conference on Advanced Cloud and Big Data (CBD), Taiyuan, China, 2020, pp. 243-247, doi: 10.1109/CBD51900.2020.00051.

[11]. Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[12]. Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[13]. Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[14]. Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

[15]. Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[16]. Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

[17]. Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[18]. Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[19]. Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.