# Comparison of Machine Learning Algorithms for Detection of Network Intrusions

Zhida Li, Prerna Batta, and Ljiljana Trajković
Communication Networks Laboratory, Simon Fraser University, Vancouver, British Columbia, Canada

## Introduction

- Detecting and analyzing network anomalies and intrusions are important topics in cyber security.
- Network intrusions may be classified using machine learning algorithms such as Recurrent Neural Networks (RNNs) and Broad Learning System (BLS).
- Classification models are trained and tested using the NSL-KDD dataset containing information about intrusion and regular network connections.
- Performance results indicate that the BLS algorithm shows comparable performance and has shorter training time.

## Intrusion Detection

- Various detection systems have been designed using machine learning techniques that help detect malicious intentions of network users.
  - Classification algorithms: J48, naive Bayes (NB), NB Tree, Random Forests (RF), Random Tree (RT), Multilayer Perception (MP), Support Vector Machine (SVM)
  - Deep learning algorithms: Network (NIDS) and Recurrent Neural Network (RNN-IDS) Intrusion Detection Systems
  - Hybrid framework: Binary Classifier (BC) modules based on the C4.5 algorithm, aggregation module, and k-NN module
  - Recurrent Neural Networks (RNNs): Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (Bi-LSTM)
  - Broad Learning System (BLS): an alternative to deep learning networks with increased number of mapped features and enhancement nodes

## Data Processing

### NSL-KDD Dataset: Types of Intrusion Attacks

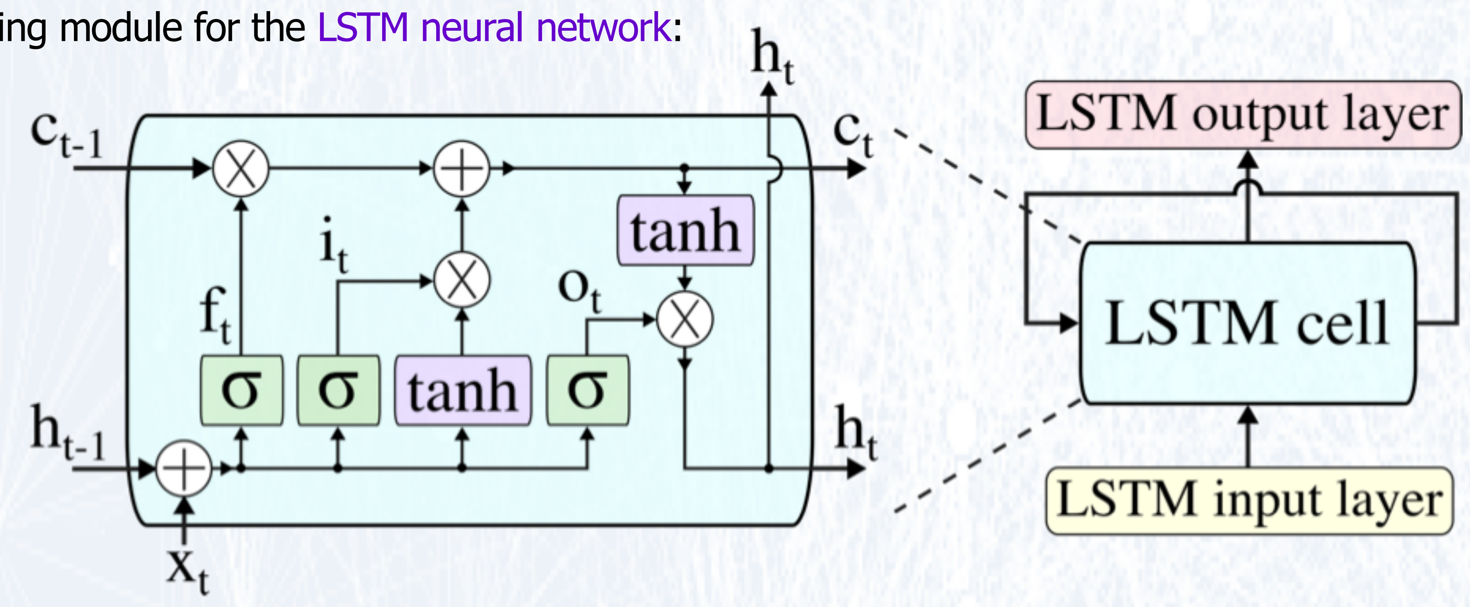
| Type | Intrusion attacks |
|---|---|
| DoS | back, land, neptune, pod, smurf, teardrop, mailbomb, processtable, udpstorm, apache2, worm |
| U2R | buffer-overflow, loadmodule, perl, rootkit, sqlattack, xterm, ps |
| R2L | fpt-write, guess-passwd, imap, multihop, phf, spy, warezmaster, xlock, xsnoop, snmpguess, snmpgetattack, httptunnel, sendmail, named |
| Probe | ipsweep, nmap, portsweep, satan, mscan, saint |

### NSL-KDD Dataset: Number of Data Points

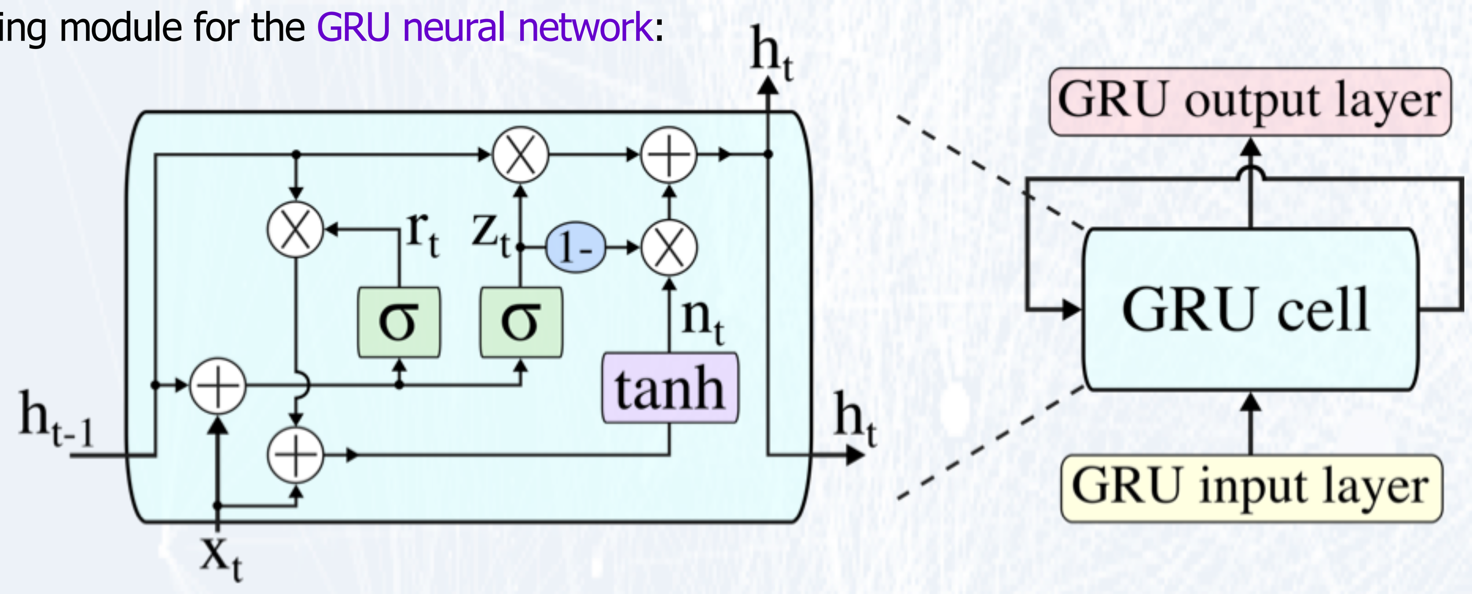
| | Regular | DoS | U2R | R2L | Probe | Total |
|---|---|---|---|---|---|---|
| $KDDTrain^+$ | 67,343 | 45,927 | 52 | 995 | 11,656 | 125,973 |
| $KDDTest^+$ | 9,711 | 7,458 | 200 | 2,754 | 2,421 | 22,544 |
| $KDDTest^{-21}$ | 2,152 | 4,342 | 200 | 2,754 | 2,402 | 11,850 |

## Classification Algorithms

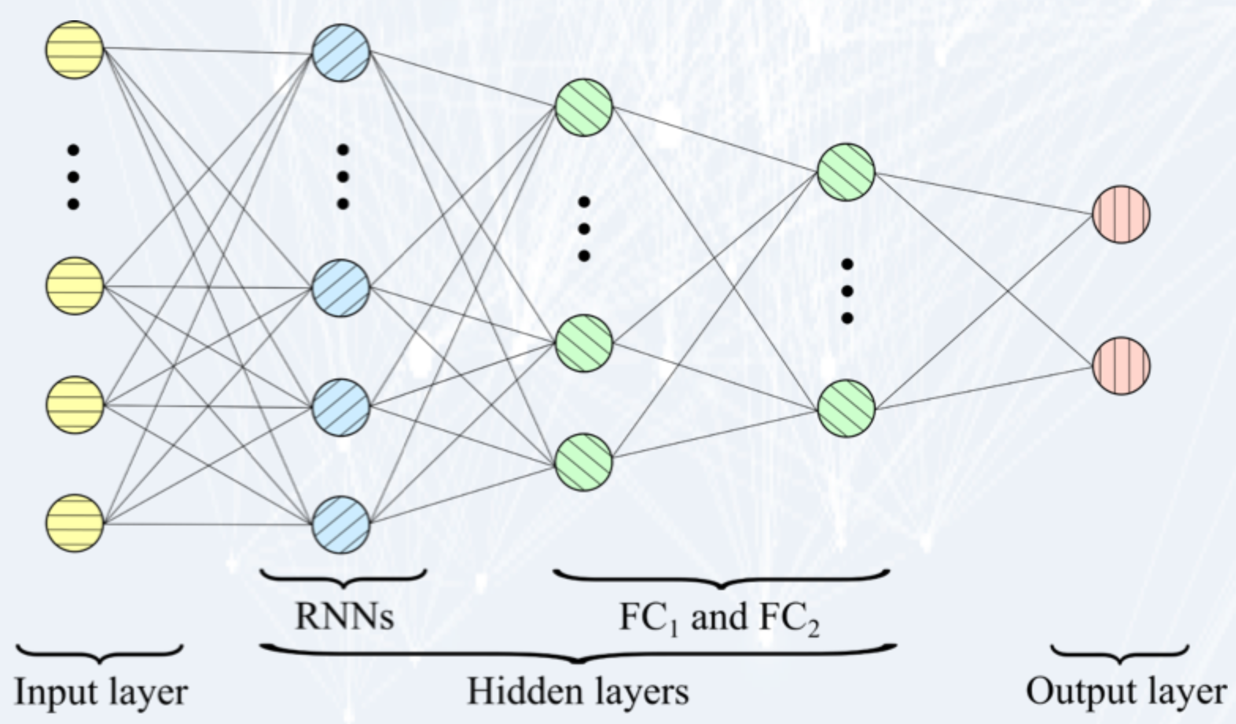
- Repeating module for the LSTM neural network:



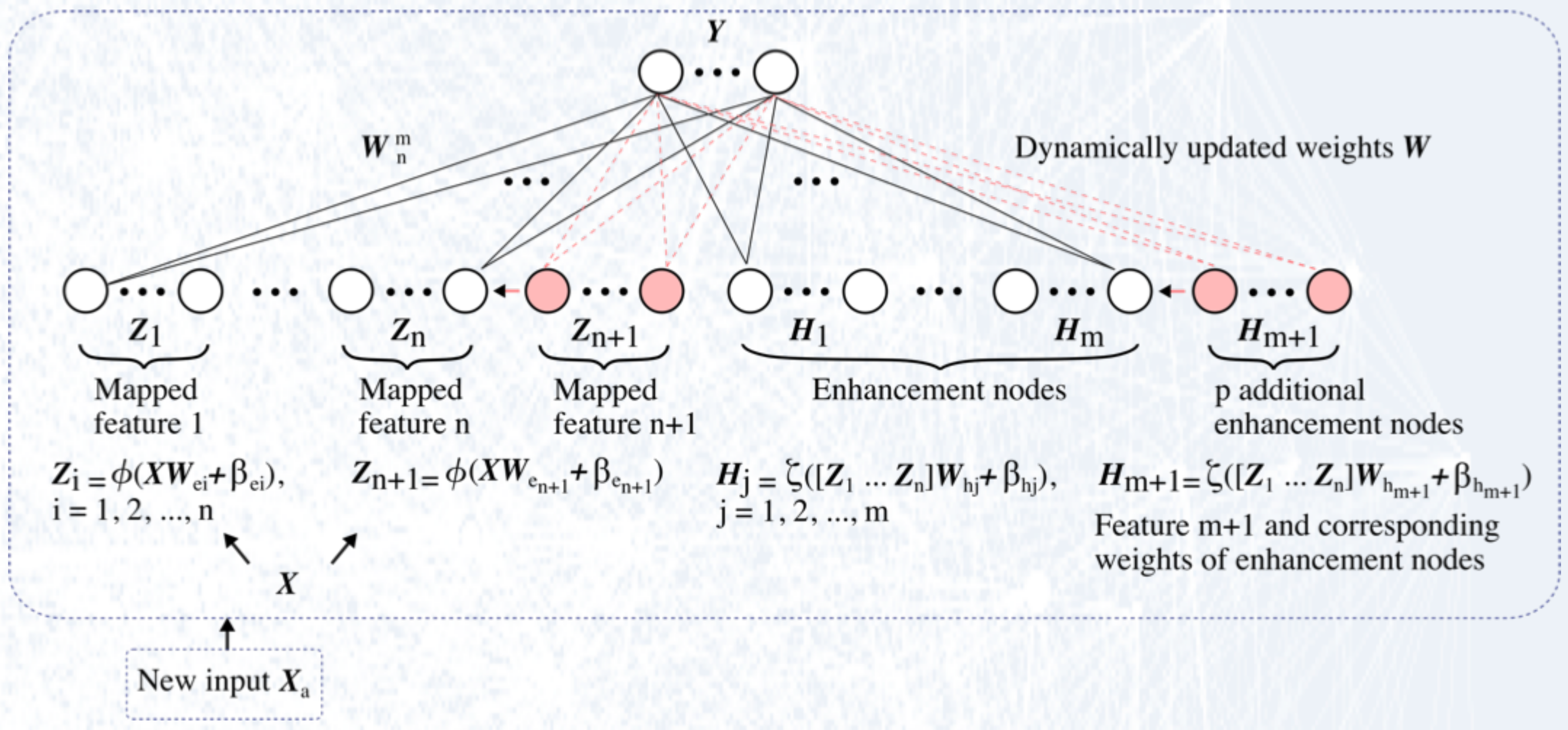

- Repeating module for the GRU neural network:




- Deep learning neural network model:




- Module of the Broad Learning System algorithm with increment of mapped features, enhancement nodes, and new input data:




## Experimental Procedure

- Step 1: Convert categorical into numerical features using dummy coding for training and test datasets
- Step 2: Normalize training and test datasets
- Step 3: Tune the model parameters during the 10-fold validation
- Step 4: Test LSTM, GRU, Bi-LSTM, and BLS models using $KDDTest^+$ and $KDDTest^{-21}$ datasets
- Step 5: Evaluate derived models based on accuracy and F-Score for binary and multiple classes

## Performance Evaluation

| Model | | LSTM | GRU | Bi-LSTM | BLS |
|---|---|---|---|---|---|
| Two-way classification | | | | | |
| Accuracy (%) | $KDDTest^+$ | 82.68 | 82.87 | 81.03 | 84.14 |
| | $KDDTest^{-21}$ | 64.32 | 65.42 | 64.31 | 72.64 |
| F-Score (%) | $KDDTest^+$ | 82.76 | 83.05 | 81.23 | 84.68 |
| | $KDDTest^{-21}$ | 73.18 | 74.60 | 73.49 | 80.61 |
| Five-way classification | | | | | |
| Accuracy (%) | $KDDTest^+$ | 79.56 | 80.17 | 79.44 | 82.47 |
| | $KDDTest^{-21}$ | 60.51 | 60.75 | 60.80 | 70.30 |

| Model | | NB Tree [1] | RT [1] | NIDS [2] | RNN-IDS [3] | BC+k-NN [4] |
|---|---|---|---|---|---|---|
| Two-way classification | | | | | | |
| Accuracy (%) | $KDDTest^+$ | 82.02 | 81.59 | 75.75 | 83.28 | 94.92 |
| | $KDDTest^{-21}$ | 66.16 | 58.51 | N/A | 68.55 | 91.35 |

| Model | | J48 [3] | NB [3] | NB Tree [3] | MP [3] | RNN-IDS [3] |
|---|---|---|---|---|---|---|
| Five-way classification | | | | | | |
| Accuracy (%) | $KDDTest^+$ | 74.60 | 74.40 | 75.40 | 78.10 | 81.29 |
| | $KDDTest^{-21}$ | 51.90 | 55.77 | 55.40 | 58.40 | 64.67 |

| Model | LSTM | GRU | Bi-LSTM | BLS | RNN-IDS [3] |
|---|---|---|---|---|---|
| Training time (s) | 355.86 | 345.04 | 497.66 | 21.92 | 5,516.00 |

## Conclusion

- Three types of RNNs and a BLS have been employed to detect network intrusions.
- $KDDTest^+$ and $KDDTest^{-21}$ datasets: BLS shows better performance than LSTM, GRU, Bi-LSTM, and most reported results.
- BLS performance depends on the number of mapped features and enhancement nodes.
- While additional mapped features and enhancement nodes improve BLS performance, they require more memory and longer training time.
- Advantage of the BLS model is that it requires considerably shorter time for training than the conventional deep learning networks.

## References

[1] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. IEEE Symp. Comput. Intell. in Security and Defense Appl. (CISDA)*, Ottawa, ON, Canada, July 2009, pp. 1–6.

[2] 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. Wireless Netw. Mobile Commun. (WINCOM)*, Fez, Morocco, Oct. 2016, pp. 258–263.

[3] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *IEEE Access*, vol. 5, pp. 21954–21961, Nov. 2017.

[4] L. Li, Y. Yu, S. Bai, Y. Hou, and X. Chen, "An effective two-step intrusion detection approach based on binary classification and k-NN," *IEEE Access*, vol. 6, pp. 12060–12073, Mar. 2018.

[5] C. L. P. Chen and Z. Liu, "Broad learning system: an effective and efficient incremental learning system without the need for deep architecture," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 1, pp. 10–24, Jan. 2018.

[6] NSL-KDD Data Set [Online]. Available: https://web.archive.org/web/20150205070216/http://nsl.cs.unb.ca/NSL-KDD/. Accessed: July 18, 2018.

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

SMC 2018
IEEE International Conference on Systems, Man, and Cybernetics
October 2018, Miyazaki, Japan

Communication Networks Laboratory
http://www.ensc.sfu.ca/~ljilja/cnl