The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2024

Volume 29, Pages XXX-XXX

ICRETS 2024: International Conference on Research in Engineering, Technology and Science

Network Anomaly Detection Using Hybrid Multiscale Residual Features Approach

Salman Mohammed JIDDAH
Cyprus International University

Dilnoza ASROROVA
Cyprus International University

Balen Kamal HAMA
Cyprus International University

Abstract: This study proposes a novel hybrid multiscale residual feature approach for network traffic anomaly detection. Utilizing the NSL-KDD dataset, we employed random forest feature selection to identify the 10 most efficient feature sets for classification. These features were then classified using a hybrid deep learning model incorporating multiscale residual blocks and Bidirectional Long Short-Term Memory (Bi-LSTM) layers. The performance evaluation revealed a significant improvement in accuracy compared to previous state-of-the-art methods. Specifically, the proposed approach achieved an accuracy of 99.19% for binary classification and 96.67% for multiclass classification. The lower performance in multiclass classification highlights the challenges posed by unbalanced and insufficient data for certain network anomaly classes within the dataset. Our findings suggest that this hybrid model provides a robust and effective solution for network anomaly detection. We recommend future research to conduct additional experiments using this approach on other network anomaly datasets to further validate its efficacy and robustness across diverse network anomaly classes. This study contributes to advancing the field of network security by presenting a highly accurate and innovative method for detecting network traffic anomalies.

Keywords: Anomaly Detection, Multiscale Residual Blocks, Bidirectional Long Short-Term Memory (Bi-LSTM), Random Forest Feature Selection, NSL-KDD Dataset

Introduction

Information and communication technology has become a critical part of human development, and is the backbone of the current digital age. The internet as we know it is primarily run on computer networks, and it is crucial there is security and uptime at all times (Singh et al., 2024). However, the internet, like all technologies, has its challenges regarding the desired ideal functions. Networks often face security breaches that can be devastating on the network users, and the frameworks of the networks (Zehra et al., 2023). Network security is an approach of ensuring security for computer networks, this includes fraud detection, cyber security, network surveillance for military purposes and much more (Yaseen, 2023). Network security is often done with the observation of network traffic, and when there is an unusual pattern in the network otherwise known as anomalies it is investigated. Hence, network anomaly has become a crucial part of network security. Network anomaly detection is a process that has seen recent advancements with the development of machine learning and artificial intelligence.

The field of network security and network anomaly detection has seen increased interests with the evolution of the internet and communication technologies. The growth of the technology also implies an increased challenge

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

of security of the infrastructure (Diro et al., 2024). Several datasets and methods have over the years been proposed for advancing the discipline or research of network anomaly detection and security. Artificial intelligence application for network anomaly detection has recently been the trending approach due to the robustness of the computing tools (Abdssalam & Jiddah, 2024).

Research Contribution

This study is aimed at contributing to the research on network anomaly detection and network security. The study proposes a hybrid approach of network anomaly detection using a fusion of multiscale residual features block and layers of Bidirectional Long Short Term Memory (Bi-LSTM) variation of recurrent neural network (RNN). The proposed approach is applied and tested on the NSL-KDD network anomaly dataset. The contributions of this study are highlighted as follows:

- Proposing a novel approach for classification and detection of network anomalies with significant performance improvement.
- Approaching the anomaly detection problem from a binary and multiclass approach to ensure the robustness and versatility of proposed approach in real time applications.

Literature Review

Network anomaly detection research has become an essential research area as it involves ensuring there is optimal network security. Network anomaly detection has evolved over the years, the process has seen the use of statistical algorithms and has now pivoted into the use of machine learning and artificial intelligence techniques (Zehra et al., 2023). Early network anomaly detection as the study of Denning (1987) proposed the use of rule based systems for network intrusion detection. The significant development in machine learning has significantly complemented the processes and techniques of network anomaly detection. Machine learning techniques such as support vector machines (SVM), decision trees, and k-nearest neighbors (k-NN) have been widely adopted with much success in processes of anomaly detection (Cui et al., 2023). The demonstration of the efficiency of such machine learning techniques were seen as early as 2007 in the work of Peddabachigari et al. (2007) which highlighted the efficiencies of these techniques. These early machine learning techniques are however challenged by complex network patterns and high dimensional data (Li et al., 2023).

The emergence of deep learning techniques has presented tools and algorithms to mitigate the challenges that were faced by the early machine learning techniques due to their ability to handle complex and high dimension data (Yaseen, 2023). Recent studies have proposed and implemented deep learning algorithms for network anomaly detection. Li et al. (2020) proposed a hybrid deep learning network combining convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to capture both spatial and temporal features of network traffic. Similarly, Yin et al. (2021) utilized a deep belief network (DBN) combined with an SVM to enhance detection accuracy, demonstrating significant improvements over traditional machine learning methods. Residual networks (ResNet) as proposed by He et al. (2016), have been used for anomaly detection and have shown significant performance improvement as seen in Sun et al. (2021). Sun et al. implemented a ResNet based deep learning model for the enhancement of network feature extraction which they reported improved accuracy performance. Bi-LSTM deep learning networks have been reported as powerful deep learning networks with a long term sequential data sequence capture ability (Redhu & Kumar, 2023). A demonstration of the effectiveness of Bi-LSTM networks is demonstrated in the study of Wang et al. (2021).

Recent studies have emphasized the importance of feature selection in improving model performance. Liu et al. (2020) employed a random forest-based feature selection method to identify the most relevant features from the NSL-KDD dataset, resulting in improved detection accuracy. Similarly, hybrid approaches combining feature selection techniques with deep learning models have been shown to effectively balance computational efficiency and detection accuracy. The NSL-KDD dataset remains a benchmark for evaluating network anomaly detection methods, providing a comprehensive set of labeled network traffic data. Despite its limitations, such as class imbalance, it continues to serve as a valuable resource for benchmarking and comparative studies. Recent works have also explored the use of other datasets, such as UNSW-NB15 and CICIDS2017, to validate the robustness of proposed methods across diverse network environments.

Method

This section presents the proposed methodology for network traffic detection, the presented method uses a hybrid multiscale residual feature network and Bi-LSTM network. The section presents the step by step methodological design and tools used in this study as illustrated in Figure 1.

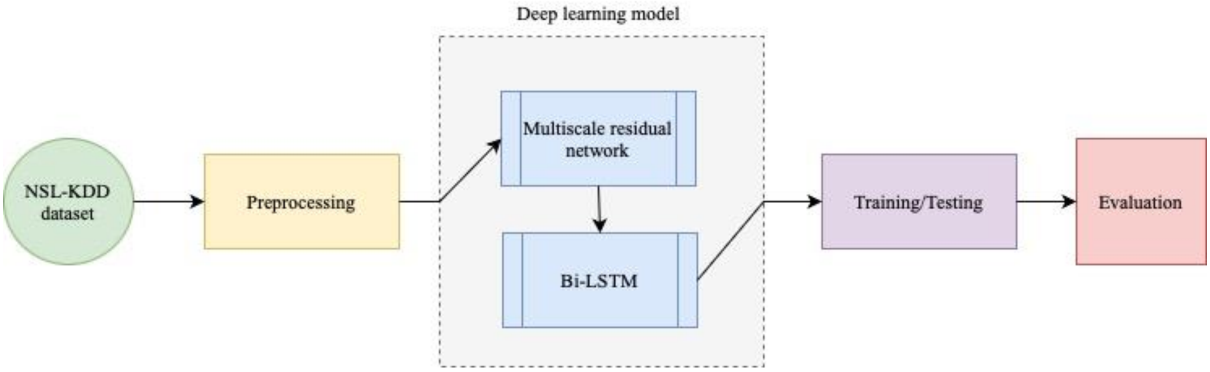

Figure 1. Proposed approach framework

NSL-KDD Dataset

The NSL-KDD dataset was first introduced by the Canadian Institute of Cyber Security in 2009 (Bala & Nagpal, 2019). This dataset consists of network based intrusion detection data, this dataset was introduced as a derivation and update to the famous KDD99 dataset introduced by DARPA intrusion detection systems which contains raw military network traffic data (Tavallaee et al., 2009). The NSL-KDD is a robust dataset which contains a total of 41 sets of classes, where 40 of those classes are different types of network intrusion classes, and one class is the normal network data (Meena & Choudhary, 2017). The classes of network intrusion classes in the dataset include Denial of Service (DoS) attacks, user to root (U2R) attacks, remote to local (R2L) attacks, and probing attacks.

Preprocessing

The NSL-KDD dataset was used as the primary dataset to carry out experiments to test the proposed network anomaly detection approach presented in this paper. The dataset was subjected through a process of data preprocessing to ensure the dataset is suitable for the research experiments based on the research approach. The dataset is processed for missing values and inconsistent data to ensure redundant data and noise is not carried into the experiments. Categorical feature sets were also identified in the dataset and were converted into numerical sets using the one not encoding approach of data normalization (Jiddah & Yurtkan, 2023). Data normalization is implemented with numbers between 0 and 1 based on the min-max data normalization technique as illustrated in equation 1. Figure 2, and Figure 3 illustrate the distribution of classes in the processed dataset for binary classes of normal and anomaly network classes, and multiple classes of network anomaly respectively.

$$A_{Norm} = \frac{\mathrm{A} - A_{min}}{A_{max} - A_{min}} \qquad (1)$$

Where A and A_{norm} are normalization of the data and the normalized data respectively.

Feature Selection

The NSL-KDD dataset has a total of 40 different network data features, in our proposed technique of network anomaly a feature selection is proposed to enhance the deep learning accuracy. This study implements the random forest feature selection technique to reduce the features from 40 to the best 10 feature sets to train the proposed model. Random forest is a machine learning algorithm which implements numerous decision trees using ensemble learning to classify the best features from a set of features (Kosaraju et al., 2023). This feature selection approach also ensures the reduction of imbalance component ranks using what is referred to as information gain (Ma et al., 2023). The final 10 features selected for the use in the model training and testing are as follows: logged_in, count, same_srv_rate, dst_host_count, dst_host_srv_count, dst_host_same_srv_rate, dst_host_serror_rate, level, service_domain_u, flag_SF.

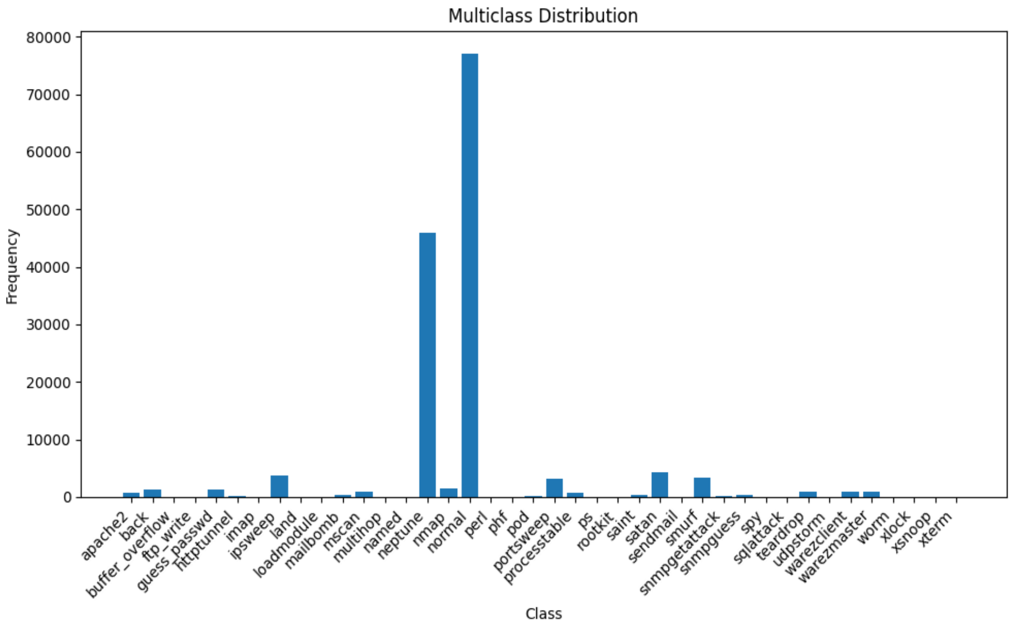

Figure 2. Multiclass distribution

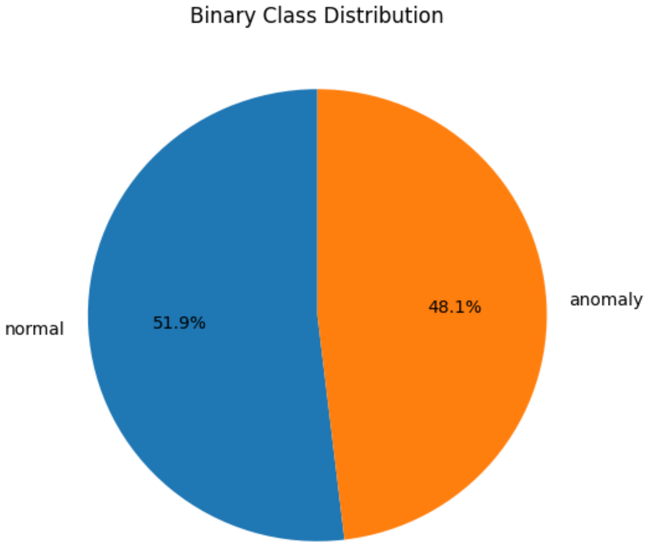

Figure 3. Binary class distribution

Deep Learning Model

This study proposes the use of a hybrid deep learning architecture which uses an integration of multiscale residual blocks and layers of Bi-LSTM recurrent network layers. The use of multiscale residual blocks in this approach is done to address challenges of gradient vanishing, and the Bi-LSTM layers are used to enable the capture of inherent network traffic temporal dependencies in the input data.

Multiscale Residual Block

Residual blocks are a key component of the ResNet deep learning architecture which were introduced by He et al. (2016). They were introduced to address vanishing gradients which are known challenges of deep neural networks; they function by allowing a direct flow of gradients using shortcut connections that allow for deeper networks (Duan et al., 2023). Residual blocks implemented in models have been reported to enable faster convergence which require fewer training epochs to reach network performance (Wu et al., 2023). A residual block is designed to learn a residual mapping, F(x)F(x), which represents the difference between the input and the desired output. The key idea is to allow the network to fit the residual mapping instead of directly learning the desired underlying mapping. Let x be the input to the residual block and H(x)H(x) be the desired underlying mapping. The residual block aims to learn F(x)=H(x)−xF(x)=H(x)−x. Thus, the output of the residual block is:

$$y=F(x)+xy=F(x)+x \qquad (2)$$

This can be implemented as:

$$y=\sigma(W2\sigma(W1x+b1)+b2)+xy=\sigma(W2\sigma(W1x+b1)+b2)+x \qquad (3)$$

where:

W1 and W2 are the weight matrices of the two dense layers.

b1 and b2 are the bias vectors.

σ denotes the ReLU activation function.

This study uses a multiscale residual block as part of the deep learning model which varies in sizes, the model uses 128, 64, and 32 units which are sequentially stacked. The multiscale residual blocks ensure a hierarchical multiscale feature capture which enables the robust learning of diverse points in the feature set of the input data. Figure 4 illustrates the residual block model, and its connection to the Bi-LSTM module of the deep learning model.

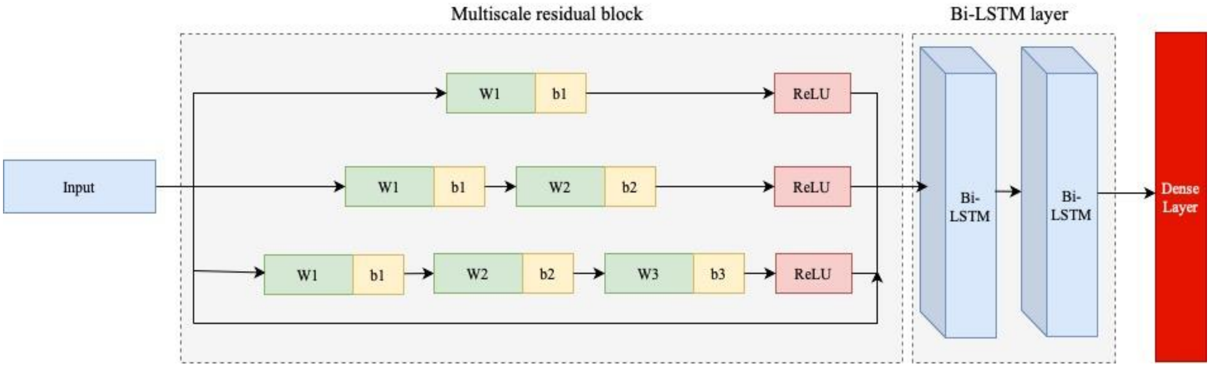

Figure 4. Deep learning model

Bi-LSTM Layers

Bidirectional LSTM (Bi-LSTM) are a variation of recurrent neural networks under the categories of Long Short Term Memory (LSTM). Bi-LSTMs are designed as neural networks that learn and remember features over a long sequence of data input (Abubaera & Jiddah, 2023). Bi-LSTMs are considered enhancements to RNNs and LSTMs, because of their bidirectional learning and remembering over a sequence of data (Ranawat et al., 2023). This research integrates a Bi-LSTM layer to the proposed model presented in this study as illustrated in Figure 4. The Bi-LSTM is integrated into the model by first reshaping the output of the multiscale residual blocks preceding it into a three dimensional tensor suited for the layers of the Bi-LSTM. Two layers of Bi-LSTM are used to enable capture of temporal and complex dependencies. The first layer processes forward and backward input sequences to form a representation, which is further refined by the second Bi-LSTM layer.

Experimental Setup

The study implements experiments of the proposed approach using the Google collab virtual computing platform. This platform is suitable for the study because of its enormous computing resources. The platform also supports the use of python programming language with a wide array of machine learning libraries predefined for experiments like this. The dataset is split using a 80:20 ratio which uses 80% of the dataset for training and validation, and uses 20% for testing the trained model. The proposed model was also implemented to test its performance on a binary classification and multi class classification. Performance evaluation is carried out for the model using accuracy, recall, precision, and f1-score.

Results and Discussion

The training process for the binary classification of the proposed model initiated an early stoppage before exhausting all 100 preset epochs of training, while the multiclass classification exhausted all 100 epochs preset. Figure 5 and 6 illustrate the training accuracy and loss charts for binary classification and multiclass classification respectively. The binary classification ended its training and validation with accuracies of 99.22% and 99.08% respectively. The multiclass classification ended its training session with accuracies of 97% and 96.74% respectively. The test results of the binary classification achieved a performance accuracy of 99.19%, and the accuracy of the multiclass classification achieved a performance accuracy of 96.57%. The accuracy score of the binary classification is considerably higher than the multiclass due to the complexity of the multiclass classification of having to carry out a differentiation learning of 40 classes of network anomaly types. Table 1 shows a complete performance evaluation carried out for the two classification experiments using the proposed model of the study. Figure 7 illustrates the confusion matrix of the binary classification test experiment. A look into the multiclass classification results also showed the model was able to classify the normal network data with 99% accuracy. The classes which had a lower classification accuracy were observed to have a limited number of data samples in the dataset which impedes deep learning performance. Table 2 and 3 show the higher accuracy and lower accuracy classes as classified by the proposed model,

Table 1. Evaluation report for study classification experiments

Classification	Accuracy	Precision	Recall	F1-score
Binary	99.18%	99.19%	99.18%	99.18%
Multiclass	96.57%	96.74%	96.57%	96.38%

Table 2. Multiclass top accuracy score classes

Class	Accuracy	Support samples
Neptune attack	100%	9147
Normal	99%	15450
Port sweep attack	97%	585

Table 3. Multiclass worse accuracy score classes

Class	Accuracy	Support samples
Butter overflow attack	27%	11
Imap attack	33%	3
Warezmaster attack	57%	193

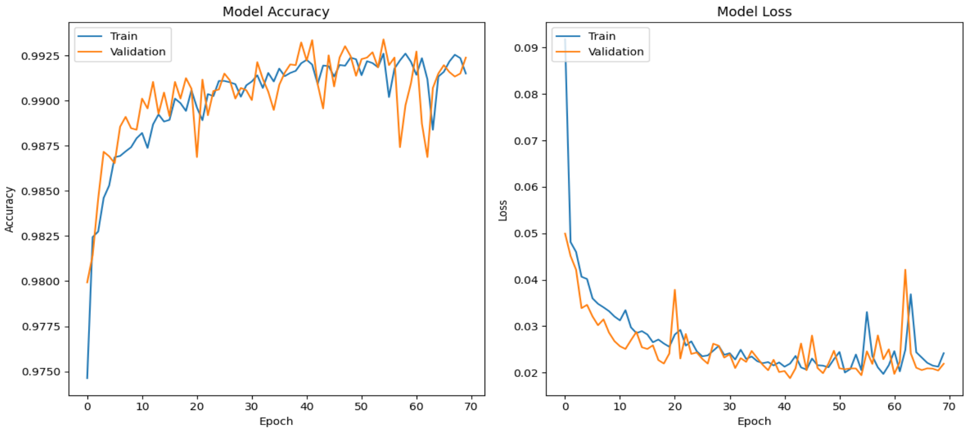

Figure 5. Training and validation charts of binary classification

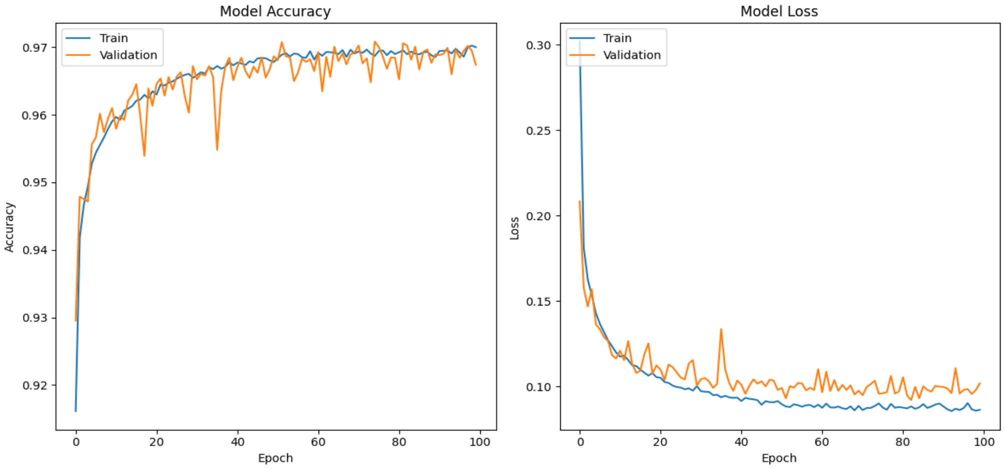

Figure 6. Training and validation charts of multiclass classification

The reported performance of the proposed approach reported in this study shows a significant performance gain on the NSL-KDD dataset. Comparing the accuracy performance of the proposed method, this study shows a significant improvement in accuracy gain compared to other state of the art approaches implemented on the NSL-KDD dataset. Some of the recent studies that approached the classification of the NSL-KDD dataset include the study of Duan et al. (2023), they proposed the use of multiscale residual blocks for feature training, they reported a binary classification performance accuracy of 89.14%. Another study used a Random Forest feature selection coupled with a classification using KNN, they reported a performance accuracy of 98.24% (Vibhute et al., 2024). A proposed approach using Fuzzy based feature selection and SVM classification reported an accuracy of 96.89% (Shiravani et al., 2023). Table 4 shows the performance comparison of the proposed method and other state of the art methods proposed in recent studies.

Table 4. Comparison of proposed approach with recent state of the art methods

Study	Proposed approach	Accuracy
Duan et al. (2023)	Multiscale residual traffic feature training	89.14%
Vibhute et al. (2024)	Random forest feature selection and KNN classification	98.24%
Shiravani et al. (2023)	Fuzzy based feature selection and SVM classification	96.89%
This study	Random forest feature selection and hybrid multiscale residual feature and Bi-LSTM	99.18%

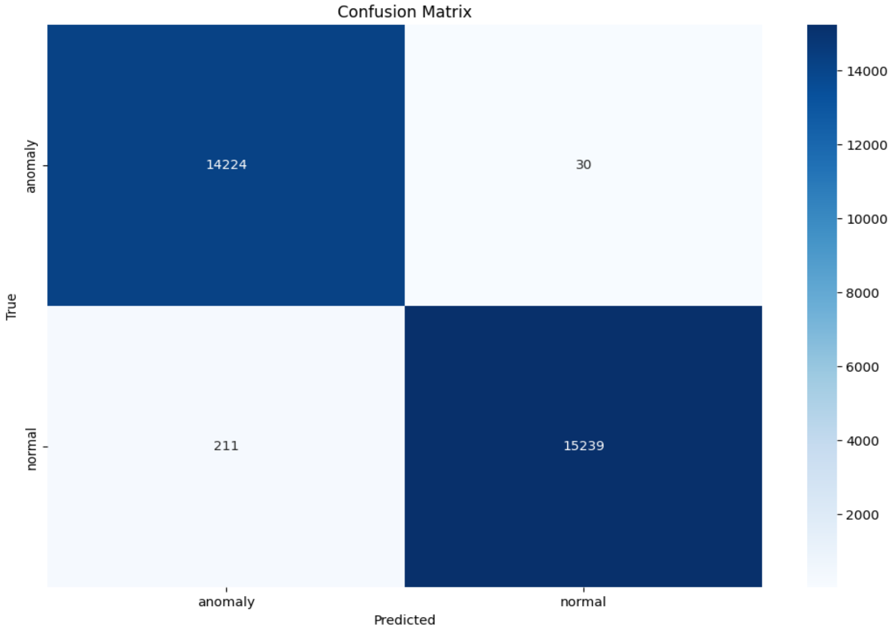

Figure 7. Binary classification test confusion matrix

Conclusion

This study proposes the use of hybrid multiscale residual feature approach for network traffic anomaly detection. The study uses random forest feature selection on the NSL-KDD dataset to select 10 feature sets that are most efficient for classification. The features were then classified over a proposed hybrid deep learning model using a multiscale residual block fused with layers of Bi-LSTM layers. The results of the performance accuracy show a significant performance improvement on the NSL-KDD dataset in comparison to previously proposed state of the art approaches, where binary classification using this approach achieved 99.19%, and 96.67% for multiclass classification. The lower performance in the multiclass classification is indicative of the unbalanced and low data for some of the classes of the network anomalies in the dataset. Based on the findings in this paper, we recommend feature directions of carrying out more experiments using the proposed approach on other network anomaly datasets to evaluate its robust performance on other network anomaly classes.

Scientific Ethics Declaration

The author(s) declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the author(s).

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Research in Engineering, Technology and Science (www.icrets.net) held in Thaskent/Uzbekistan on August 22-25, 2024.

References

Abdssalam, M.A.A.M & Jiddah, S. M. (2024). Wireless Sensor Network Anomaly Detection Using Recurrent Neural Network. The International Journal of Engineering & Information Technology (IJEIT)

Abubaera, M. M., & Jiddah, S. M. Natural Language Processing and Bi-Directional LSTM for Sentiment Analysis. *International Journal of Computer Applications*, *975*, 8887.

Bala, R., & Nagpal, R. (2019). A review on kdd cup99 and nsl nsl-kdd dataset. *International Journal of Advanced Research in Computer Science*, *10*(2).

Cui, Y., Liu, Z., & Lian, S. (2023). A survey on unsupervised anomaly detection algorithms for industrial images. *IEEE Access*, *11*, 55297-55315.

Diro, A., Kaisar, S., Vasilakos, A. V., Anwar, A., Nasirian, A., & Olani, G. (2024). Anomaly detection for space information networks: A survey of challenges, techniques, and future directions. *Computers & Security*, *139*, 103705.

Duan, X., Fu, Y., & Wang, K. (2023). Network traffic anomaly detection method based on multi-scale residual classifier. *Computer Communications*, *198*, 206-216.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778.

Jiddah, S. M., & Yurtkan, K. (2023). Dominant and complementary emotion recognition using hybrid recurrent neural network. *Signal, Image and Video Processing*, *17*(7), 3415-3423.

Kosaraju, N., Sankepally, S. R., & Mallikharjuna Rao, K. (2023, February). Categorical data: Need, encoding, selection of encoding method and its emergence in machine learning models—a practical review study on heart disease prediction dataset using pearson correlation. In *Proceedings of International Conference on Data Science and Applications: ICDSA 2022, Volume 1* (pp. 369-382). Singapore: Springer Nature Singapore.

Li, Z., Zhu, Y., & Van Leeuwen, M. (2023). A survey on explainable anomaly detection. *ACM Transactions on Knowledge Discovery from Data*, *18*(1), 1-54.

Liu, Z., Liu, Y., Liu, X., & Wu, Y. (2020). Feature selection and deep learning for intrusion detection systems. *IEEE Access*, 8, 49367-49376.

Ma, H., Peng, T., Zhang, C., Ji, C., Li, Y., & Nazir, M. S. (2023). Developing an evolutionary deep learning framework with random forest feature selection and improved flow direction algorithm for NOx concentration prediction. *Engineering Applications of Artificial Intelligence*, *123*, 106367.

Meena, G., & Choudhary, R. R. (2017, July). A review paper on IDS classification using KDD 99 and NSL KDD dataset in WEKA. In *2017 International Conference on Computer, Communications and Electronics (Comptelix)* (pp. 553-558). IEEE.

Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference (MilCIS)*, 1-6.

Peddabachigari, S., Abraham, A., Thomas, J., & Muda, Z. (2007). Intrusion detection systems using decision trees and support vector machines. *International Journal of Advanced Networking and Applications*, 2(3), 372-379.

Ranawat, N. S., Prakash, J., Miglani, A., & Kankar, P. K. (2023). Performance evaluation of LSTM and Bi-LSTM using non-convolutional features for blockage detection in centrifugal pump. *Engineering Applications of Artificial Intelligence*, *122*, 106092.

Redhu, P., & Kumar, K. (2023). Short-term traffic flow prediction based on optimized deep learning neural network: PSO-Bi-LSTM. *Physica A: Statistical Mechanics and its Applications*, *625*, 129001.

Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSP*, 1-6.

Shiravani, A., Sadreddini, M. H., & Nahook, H. N. (2023). Network intrusion detection using data dimensions reduction techniques. *Journal of Big Data*, *10*(1), 27.

Singh, R., Srivastava, N., & Kumar, A. (2024). Network Anomaly Detection Using Autoencoder on Various Datasets: A Comprehensive Review. *Recent Patents on Engineering*, *18*(9), 63-77.

Sun, Y., Li, Y., & Liu, C. (2021). Network intrusion detection based on deep learning. *Journal of Physics: Conference Series*, 1885(2), 022012.

Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009, July). A detailed analysis of the KDD CUP 99 data set. In *2009 IEEE symposium on computational intelligence for security and defense applications* (pp. 1-6). Ieee.

Vibhute, A. D., Patil, C. H., Mane, A. V., & Kale, K. V. (2024). Towards detection of network anomalies using machine learning algorithms on the NSL-KDD benchmark datasets. *Procedia Computer Science*, *233*, 960-969.

Wang, J., Zhang, H., & Zhang, J. (2021). Network anomaly detection based on bidirectional LSTM. *IEEE Access*, 9, 66543-66553.

Wu, X., Shi, H., & Zhu, H. (2023). Fault diagnosis for rolling bearings based on multiscale feature fusion deep residual networks. *Electronics*, *12*(3), 768.

Yaseen, A. (2023). The role of machine learning in network anomaly detection for cybersecurity. *Sage Science Review of Applied Machine Learning*, *6*(8), 16-34.

Yin, C., Zhu, Y., Fei, J., & He, X. (2021). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, 9, 12319-12329.

Zehra, S., Faseeha, U., Syed, H. J., Samad, F., Ibrahim, A. O., Abulfaraj, A. W., & Nagmeldin, W. (2023). Machine learning-based anomaly detection in NFV: A comprehensive survey. *Sensors*, *23*(11), 5340.

Author Information

Salman Mohammed Jiddah
Cyprus International University
Nicosia, Cyprus.
Contact e-mail: *salman.m.jiddah@gmail.com*

Dilnoza Asrorova
Cyprus International University
Nicosia, Cyprus.

Balen Kamal Hama
Cyprus International University
Nicosia, Cyprus.

To cite this article:

Jiddah, S.M., Asrorova, D.. & Hama, B.K. (2024). Network Anomaly Detection Using Hybrid Multiscale Residual Feature Approach. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)*, *29*, Page X-Page Y.