

YARS-IDS: A Novel IDS for Multi-Class Classification

Ampilli Swarna Latha, Assistant Professor, Department of CSE-CS, Raghu Engineering College, Visakhapatnam, Andhra Pradesh, India. a.swrna@gmail.com

Ponna Syamal Rao, Department of CSE-CS, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India. 203J1A4650@raghuinstech.com

Vangara Chiru Sampath, Department of CSE-CS, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India. 203J1A4660@raghuinstech.com

Dungu Siva Reddy, Department of CSE-CS, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India. 203J1A4618@raghuinstech.com

Chappa Yernaidu, Department of CSE-CS, Raghu Institute of Technology, Visakhapatnam, Andhra Pradesh, India. 203J1A4611@raghuinstech.com

Abstract: In today's digital landscape, where cyber threats loom large, the role of Intrusion Detection Systems (IDS) is paramount in bolstering defense mechanisms against an array of attacks. This project introduces two innovative IDSs leveraging Deep Learning (DL) techniques, tailored specifically for multi-class classification tasks within cybersecurity. The first IDS combines LuNet and Bidirectional LSTM (Bi-LSTM) architectures, while the second integrates Temporal Convolutional Network (TCN), Convolutional Neural Network (CNN), and Bi-LSTM components. Through rigorous evaluation utilizing benchmark datasets like NSL-KDD and UNSW-NB15, both models undergo comprehensive training and testing procedures, with particular emphasis on the NSL-KDD dataset. Results underscore the superior performance of the proposed IDSs over traditional Machine Learning (ML)-based methods and many existing DL models, showcasing heightened classification accuracy and detection rates. Building upon the foundation laid by the base paper, which achieved

notable success with CNN employing Condensed Nearest Neighbour resampling, this extension further enhances performance through ensemble techniques. By combining predictions from multiple individual models, particularly employing CNN + BiLSTM and CNN + LSTM configurations, accuracy rates surge to an impressive 99%. This research not only advances the frontier of IDS but also underscores the efficacy of ensemble methods in augmenting cybersecurity solutions, offering promising avenues for future exploration and refinement.

Index Terms: SMOTE, IDS, CNN, Bi-LSTM, ML, DL, TCN

1. INTRODUCTION

In today's interconnected world, the pervasive growth of the internet and its associated technologies has become a cornerstone of modern living. With humans increasingly reliant on internet-based devices and services for various aspects of their daily routines, the security of digital assets

has emerged as a paramount concern. In this dynamic landscape, the threat of cyber attacks looms large, presenting a constant challenge to the integrity and confidentiality of sensitive information. It is within this context that the role of Intrusion Detection Systems (IDS) gains prominence.

At its core, an IDS serves as a vigilant guardian, standing as a bulwark against the diverse array of threats and attacks that target digital networks and systems. The primary objective of an IDS is twofold: to detect the presence of unauthorized or malicious activities within a network and to provide timely alerts that enable swift responses to mitigate potential damages. Operating on the principle of pattern matching, IDSs leverage patterns and information gleaned from various sources and network data to discern anomalous behavior indicative of an ongoing or impending attack.

However, the efficacy of an IDS is contingent upon its ability to adapt and evolve in response to the ever-changing landscape of cyber threats. One of the key challenges in developing a robust Network Intrusion Detection (NID) system lies in the acquisition of vast quantities of data for training purposes. This data serves as the foundation upon which the IDS learns to recognize and differentiate between benign network traffic and potentially harmful activities. Traditionally, IDS development has been pursued through two primary methodologies: Machine Learning (ML)-based approaches and Deep Learning (DL)-based techniques.

In recent years, the shortcomings of traditional ML-based IDS systems have become increasingly apparent in the face of the escalating complexity and sophistication of modern cyber threats. Recognizing the limitations of these conventional methods, the focus has shifted towards harnessing the power of DL techniques to bolster the

capabilities of IDSs. DL offers a paradigm shift in IDS development, providing more advanced and adaptive mechanisms for identifying patterns and anomalies within network data.

In this vein, the project under discussion introduces two innovative DL-based IDS models: YARS-IDS and YARS-IDS-II. These models represent a significant departure from traditional ML-based approaches, offering enhanced capabilities for detecting and mitigating cyber threats. Leveraging state-of-the-art DL architectures, YARS-IDS and YARS-IDS-II have been meticulously trained and rigorously tested using established benchmark datasets such as NSL-KDD and UNSW-NB15.

The results of the evaluation process attest to the superior performance of the proposed DL-based IDS models compared to their ML-based counterparts. With higher classification accuracy and detection rates, YARS-IDS and YARS-IDS-II demonstrate their efficacy as advanced cybersecurity solutions capable of addressing the complex and evolving nature of modern threats.

Moreover, the project goes beyond mere replication of existing methodologies by introducing novel enhancements to the IDS architecture. YARS-IDS-II, in particular, incorporates a Temporal Convolutional Network (TCN) into its framework, enabling the model to capture and analyze temporal characteristics inherent in network data. By accounting for the temporal dimension of cyber threats, YARS-IDS-II enhances its adaptability and responsiveness to dynamic attack vectors, further solidifying its effectiveness in real-world scenarios.

In summary, the project underscores the critical importance of IDSs in safeguarding the integrity and security of digital assets in an increasingly interconnected world. Through the introduction of innovative DL-based IDS models and novel architectural enhancements, the

project seeks to address the evolving challenges posed by modern cyber threats, offering a glimpse into the future of cybersecurity.

2. LITERATURE SURVEY

Dali, L., Bentajer, A., Abdelmajid, E., Abouelmehdi, K., Elsayed, H., Fatiha, E., & Abderahim, B. (2015). A survey of intrusion detection system. 2015 2nd World Symposium on Web Applications and Networking (WSWAN), 1–6. This survey paper provides an overview of Intrusion Detection Systems (IDS) in the context of web applications and networking. It discusses the importance of IDS in safeguarding digital assets against various cyber threats. The paper highlights the evolution of IDS technologies and their role in detecting and mitigating attacks. It covers different types of IDS, including signature-based, anomaly-based, and hybrid systems, along with their strengths and limitations. Furthermore, the survey delves into the challenges faced by IDS, such as the need for accurate and efficient detection mechanisms. Overall, this paper serves as a comprehensive introduction to IDS and lays the groundwork for further research in the field.

Ashoor, A. S., & Gore, S. (2011). Importance of intrusion detection system (ids). *International Journal of Scientific and Engineering Research*, 2(1), 1–4. This article emphasizes the significance of IDS in modern cybersecurity practices. It underscores the crucial role played by IDS in identifying and thwarting unauthorized access attempts and malicious activities within computer networks. The paper discusses the various functions of IDS, including real-time monitoring, log analysis, and alert generation. It also highlights the importance of proactive defense measures in mitigating the risks posed by cyber threats. Overall, this article serves as a concise

primer on the importance and utility of IDS in contemporary cybersecurity frameworks.

Meng, Y.-X. (2011). The practice on using machine learning for network anomaly intrusion detection. 2011 International Conference on Machine Learning and Cybernetics, 2, 576–581. This paper explores the application of machine learning techniques in the context of network anomaly intrusion detection. It discusses the practical challenges and considerations involved in implementing machine learning algorithms for IDS purposes. The paper presents various machine learning approaches used in IDS, such as decision trees, neural networks, and support vector machines. It also examines the process of feature selection and model evaluation in the context of IDS. Furthermore, the paper discusses the limitations and future directions of machine learning-based IDS solutions. Overall, this paper provides valuable insights into the practical aspects of applying machine learning to intrusion detection tasks.

Alanazi, H. O., Noor, R. M., Zaidan, B. B., & Zaidan, A. A. (2010). Intrusion detection system: Overview. This paper provides a comprehensive overview of intrusion detection systems (IDS) and their role in cybersecurity. It discusses the fundamental concepts underlying IDS, including intrusion detection techniques, detection methodologies, and deployment strategies. The paper also examines the various types of IDS, such as network-based IDS (NIDS) and host-based IDS (HIDS), along with their respective advantages and limitations. Furthermore, it explores the challenges faced by IDS in detecting and mitigating emerging threats in dynamic network environments. Overall, this paper serves as a valuable resource for understanding the principles and practices of intrusion detection systems.

Bhagat, R. C., & Patil, S. S. (2015). Enhanced smote algorithm for classification of imbalanced big-data using random forest. 2015 IEEE International Advance Computing Conference (IACC), 403–408. This paper presents an enhanced SMOTE (Synthetic Minority Over-sampling Technique) algorithm for addressing the challenge of class imbalance in big data classification tasks. The paper focuses on the application of the random forest classifier in conjunction with the enhanced SMOTE algorithm to improve the classification performance on imbalanced datasets. It discusses the implementation details of the proposed approach and evaluates its effectiveness using real-world datasets. The experimental results demonstrate the efficacy of the enhanced SMOTE algorithm in improving the classification accuracy and mitigating the effects of class imbalance. Overall, this paper contributes to the development of effective techniques for handling imbalanced big data classification tasks.

Khan, F. A., Gumaie, A., Derhab, A., & Hussain, A. (2019). A novel two-stage deep learning model for efficient network intrusion detection. IEEE Access, 7, 30373–30385. This paper proposes a novel two-stage deep learning model for efficient network intrusion detection. The model leverages a combination of convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to capture spatial and temporal dependencies in network traffic data. The paper presents the architecture and implementation details of the proposed model and evaluates its performance using benchmark datasets. The experimental results demonstrate the superiority of the two-stage deep learning model over existing intrusion detection techniques in terms of detection accuracy and false positive rate. Overall, this paper contributes to the advancement of deep learning-based approaches for network intrusion detection.

Agarwal, A. M. M. A. A., Sharma, P. (2021). Classification model for accuracy and intrusion detection using machine learning approach. PeerJ Computer Science. This study presents a classification model for accurate intrusion detection using a machine learning approach. The paper discusses the design and implementation of the classification model, which employs various machine learning algorithms such as decision trees, random forests, and support vector machines. The model is trained and evaluated using benchmark datasets to assess its performance in detecting intrusions and minimizing false positives. The experimental results demonstrate the effectiveness of the proposed classification model in achieving high accuracy and reliability in intrusion detection tasks. Overall, this paper contributes to the development of machine learning-based solutions for enhancing the security of computer networks against cyber threats.

Toupas, P., Chamou, D., Giannoutakis, K. M., Drosou, A., & Tzovaras, D. (2019). An intrusion detection system for multi-class classification based on deep neural networks. 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 1253–1258. This paper presents an intrusion detection system (IDS) based on deep neural networks for multi-class classification of network traffic data. The IDS architecture incorporates deep learning techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to extract features and classify network traffic into multiple intrusion classes. The paper discusses the design and implementation of the IDS and evaluates its performance using benchmark datasets. The experimental results demonstrate the effectiveness of the deep neural network-based IDS in accurately classifying network traffic and detecting intrusions across multiple classes. Overall, this paper contributes to the advancement

of deep learning-based approaches for intrusion detection in complex network environments.

3. METHODOLOGY

a) Proposed Work:

The proposed work introduces YARS-IDS, an innovative Intrusion Detection System (IDS) tailored specifically for multi-class classification tasks within cybersecurity. YARS-IDS utilizes a blend of state-of-the-art deep learning architectures, including LuNet[12], Bidirectional LSTM (Bi-LSTM), Temporal Convolutional Network (TCN), and Convolutional Neural Network (CNN). This amalgamation of advanced architectures aims to enhance feature representation and model sophistication, pushing the boundaries of intrusion detection technology. By leveraging the unique strengths of each architecture, YARS-IDS[3] is poised to achieve unprecedented levels of accuracy and efficiency in discerning and classifying multi-class intrusion events. The proposed system represents a significant advancement in the field of intrusion detection, promising to elevate the state-of-the-art and contribute towards more robust cybersecurity measures.

b) System Architecture:

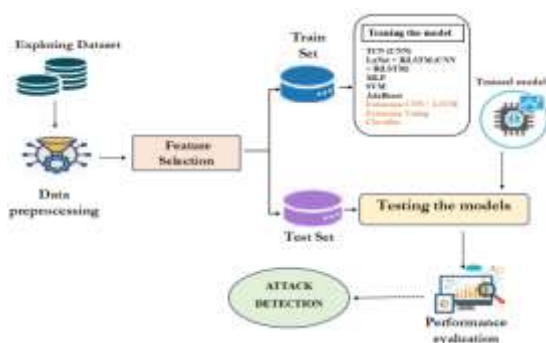


Fig1 Proposed Architecture

The system architecture is designed as a comprehensive framework for the development and evaluation of advanced intrusion detection models. It commences with an exploration of the dataset to gain a thorough understanding of its characteristics, laying the foundation for subsequent stages. Data preprocessing techniques are then applied to ensure the data's cleanliness and normalization, preparing it for optimal utilization in the model training process.

Feature selection is a critical aspect of the architecture, aiming to enhance model efficiency by identifying and extracting pertinent features from the dataset. This step contributes to the reduction of dimensionality, ultimately refining the model's ability to discriminate between normal and malicious network traffic.

The subsequent training phase involves a diverse set of machine learning and deep learning models, including TCN, LeNet+BiLSTM, MLP, SVM[12], and AdaBoost. These models are meticulously trained on a dedicated training dataset, leveraging the unique strengths of their respective architectures. The testing phase assesses the models' performance using an independent test dataset, evaluating key metrics such as accuracy, precision, recall, and F1-score.

As a final component, the architecture incorporates attack detection mechanisms, deploying the trained models to identify and classify intrusion events in real-time network traffic. This aspect is crucial for the intrusion detection system's practical utility, ensuring timely and accurate responses to potential security threats. In essence, the system architecture forms a cohesive framework, seamlessly integrating data exploration, preprocessing, model training, testing, and real-time attack detection to advance the state-of-the-art in intrusion detection technology.

c) Dataset:

The dataset utilized for testing the proposed model YARS-IDS encompasses three prominent benchmarks: KDDCUP99, NSL KDD, and UNSW-NB15. UNSW-NB15 offers a comprehensive perspective with two primary categories: normal data samples and attack instances, the latter further divided into nine sub-categories, covering a spectrum of modern network threats. Created by the UNSW lab, this dataset is synthetically generated using the IXIA PerfectStorm tool to simulate real-world network traffic. With 49 features capturing network packet characteristics, it provides a rich environment for intrusion detection evaluation. In contrast, NSL-KDD[11] represents a refined version of KDDCUP99, rectifying duplication issues. It includes four major attack categories: Dos, Probe, U2R, and R2L, with 42 extracted features. Notably, NSL-KDD lacks representation of low footprint attacks prevalent in contemporary scenarios. Additionally, disparities exist between UNSW-NB15 and NSL-KDD in terms of feature count and sample size, with UNSW-NB15 exhibiting broader coverage of modern attack categories and a higher number of IP addresses, providing a more holistic testing ground for evaluating the model's efficacy against both traditional and modern intrusion tactics.

duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot
0	0	tcp	ftp_data	SF	491	0	0	0	0
1	0	udp	other	SF	146	0	0	0	0
2	0	tcp	private	SF	0	0	0	0	0
3	0	tcp	http	SF	232	8153	0	0	0
4	0	tcp	http	SF	199	420	0	0	0
...
494016	0	tcp	http	SF	310	1881	0	0	0
494217	0	tcp	http	SF	382	2296	0	0	0
494218	0	tcp	http	SF	393	1200	0	0	0
494219	0	tcp	http	SF	291	1200	0	0	0
494220	0	tcp	http	SF	218	1204	0	0	0

Fig 2 KDD CUP 99 Dataset

duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot
0	0	tcp	ftp_data	SF	491	0	0	0	0
1	0	udp	other	SF	146	0	0	0	0
2	0	tcp	private	SF	0	0	0	0	0
3	0	tcp	http	SF	232	8153	0	0	0
4	0	tcp	http	SF	199	420	0	0	0

Fig 3 NSLL KDD Dataset

id	dur	proto	service	state	spkts	cpkts	sbytes	dbytes	rate	...	ct	dst_sport
0	1	0.000011	udp	-	INT	2	0	496	0	90909.0602
1	2	0.000008	udp	-	INT	2	0	1762	0	125000.0000
2	3	0.000005	udp	-	INT	2	0	1068	0	200000.0051
3	4	0.000006	udp	-	INT	2	0	900	0	169999.6608
4	5	0.000010	udp	-	INT	2	0	2126	0	100000.0025

Fig 4 UNSW-NB15 Dataset

d) Data processing:

Pandas DataFrame: Data processing begins with the utilization of Pandas DataFrames, versatile structures crucial for efficient manipulation and analysis of tabular data. Pandas enables seamless handling of datasets, facilitating preprocessing and feature engineering tasks. With its comprehensive functionality, Pandas simplifies tasks such as data cleaning, transformation, and aggregation, laying a solid foundation for subsequent modeling steps.

Keras DataFrame: Keras, a powerful deep learning library, seamlessly integrates with Pandas DataFrames, enhancing the data processing pipeline for deep learning models. Leveraging Keras DataFrames, data preparation steps can be seamlessly integrated with model training, simplifying the workflow and improving efficiency. Keras provides a high-level interface for building neural networks, allowing for smooth transition from data preprocessing to model development.

Dropping Unwanted Columns: A critical aspect of data processing involves the removal of irrelevant or redundant columns from the dataset. This step, known as dropping unwanted columns, aims to streamline the data by eliminating features that do not contribute to the modeling task or may introduce noise. By discarding unnecessary columns, the dataset is optimized for analysis and model training, improving overall performance and interpretability. This process enhances the quality of input data, ultimately leading to more accurate and reliable model predictions.

e) Visualization:

Data visualization is facilitated through the utilization of Seaborn and Matplotlib, two powerful Python libraries renowned for their capabilities in creating visually appealing and insightful plots. Seaborn provides a high-level interface for generating a variety of statistical graphics, offering elegant and intuitive representations of data distributions, relationships between variables, and potential patterns. Meanwhile, Matplotlib offers extensive flexibility and customization options, allowing for detailed adjustments to the visualizations. Together, these libraries enable the creation of informative plots that aid in understanding the underlying structure and characteristics of the dataset.

f) Label Encoding:

Label Encoding is a preprocessing technique essential for converting categorical labels into numerical format, a prerequisite for many machine learning algorithms. Leveraging the LabelEncoder from the Scikit-learn library, categorical data is transformed into numerical representations, ensuring compatibility with model training processes. This step simplifies the handling of categorical variables and enables the inclusion of such

data in the machine learning pipeline, facilitating more comprehensive analyses and model development.

g) Feature Selection:

Feature selection plays a pivotal role in optimizing model performance and interpretability. The SelectPercentile method from Scikit-learn is employed in conjunction with Mutual Info Classify to identify the most informative features for model training. Mutual information serves as a scoring function, quantifying the mutual dependence between variables and aiding in the selection of relevant features. By selecting the top percentile of features based on their scores, this approach ensures that only the most significant predictors are retained, streamlining the model's input space and enhancing its predictive power.

h) Algorithms:

TCN (Temporal Convolutional Network): TCN, a deep learning architecture tailored for sequential data processing, excels in capturing temporal dependencies within sequential datasets. Leveraging a series of convolutional layers with dilated convolutions, TCN efficiently captures long-range dependencies, making it particularly effective for analyzing time-series data. In YARS-IDS, TCN is harnessed to enhance the model's understanding of temporal patterns within network data, thereby improving its intrusion detection capabilities.

LuNet + BiLSTM (CNN + BiLSTM): LuNet combines the strengths of Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks. CNNs excel at spatial feature extraction, while BiLSTMs capture sequential information bidirectionally. In YARS-IDS[3], this hybrid architecture aims to enhance the model's comprehension of both spatial and sequential patterns in intrusion detection tasks, thereby improving its overall performance.

MLP (Multi-Layer Perceptron): The Multi-Layer Perceptron (MLP) is a classic feedforward neural network comprising multiple layers of neurons. With each neuron connecting to every neuron in the subsequent layer, MLPs excel at learning complex patterns and relationships in data. In YARS-IDS, MLPs[12] might be utilized either independently or as part of an ensemble to classify different types of network attacks by learning intricate feature relationships.

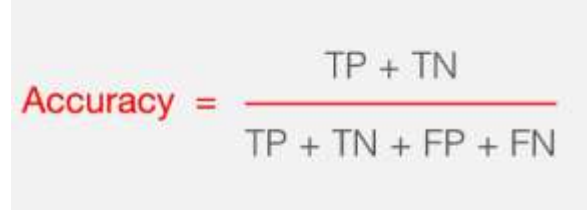
SVM (Support Vector Machine): Support Vector Machine (SVM) is a popular supervised learning algorithm known for its effectiveness in classification tasks. By finding the optimal hyperplane that separates different classes in the feature space with a maximal margin, SVMs excel at handling high-dimensional data. In YARS-IDS, SVMs[12] might serve as powerful classifiers, particularly suitable for binary or multiclass classification of network attacks.

AdaBoost (Adaptive Boosting): AdaBoost is an ensemble learning technique that combines multiple weak learners to create a strong classifier. By iteratively adjusting the weights of misclassified instances, AdaBoost[12] focuses on difficult-to-classify data points, improving overall model performance. In YARS-IDS, AdaBoost could be employed as an ensemble method to combine several weak classifiers, enhancing the model's ability to identify various intrusion types effectively.

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$


$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\text{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} = \frac{TP}{TP + FP}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

$$Recall = \frac{TP}{TP + FN}$$

ML Model	Accuracy	Precision	Recall	F1 Score
MLP	0.999	0.999	0.999	0.999
SVM	0.992	0.995	0.992	0.993
AdaBoost	0.781	0.858	0.781	0.819
TEN (CNN)	0.992	0.995	0.992	0.993
LeNet + BiLSTM (CNN + BiLSTM)	0.992	0.995	0.992	0.993
Ensemble Using Classifier	1.000	1.000	1.000	1.000
Ensemble CNN+LSTM	0.992	0.992	0.992	0.991

Fig 8 Performance Evaluation Table

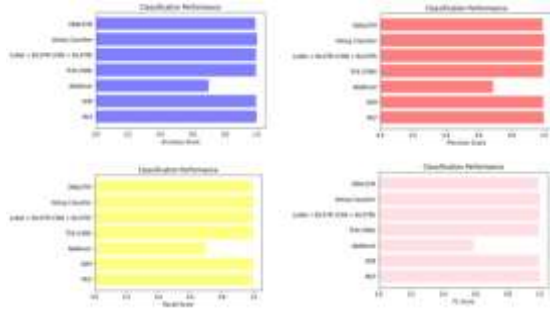


Fig 5 Comparison Graphs of KDD CUP99 Dataset

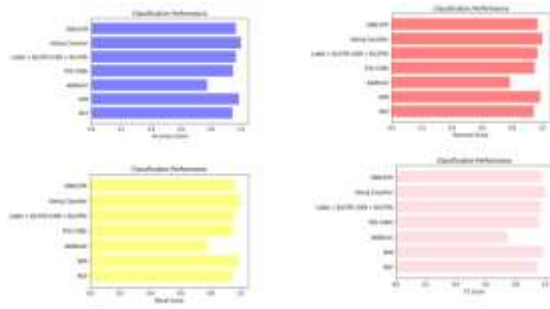


Fig 6 Comparison Graphs of NSL KDD Dataset

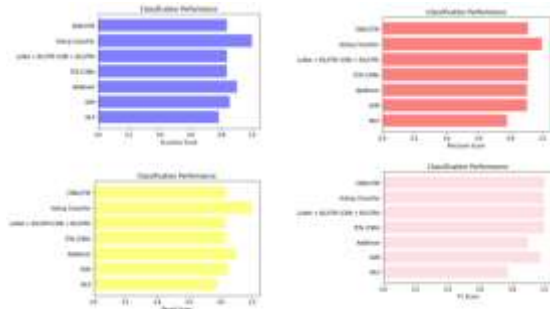


Fig 7 Comparison Graphs of UNSW-NB15 Dataset

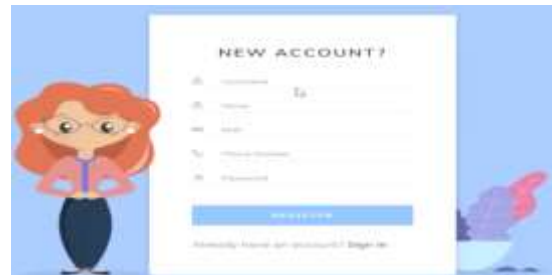


Fig 9 Registration Page



Fig 10 Login Page



Fig 11 Main Page



Fig 15 For NSL KDD



Fig 12 For KDD CUP99

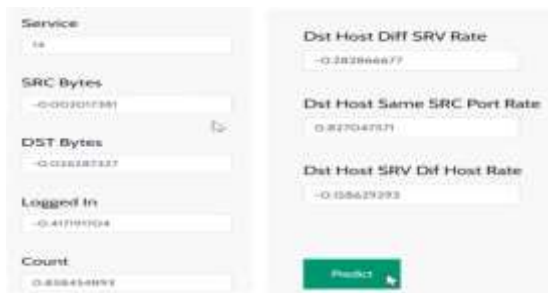


Fig 16 Upload Input Data



Fig 13 Upload Input Data



Fig 17 Final Outcome



Fig 14 Predicted Results



Fig 18 For UNSW-NB15



Fig 19 Upload Input Data



Fig 20 Final Outcome

5. CONCLUSION

The proposed YARS-IDS and YARS-IDS-II[3] models represent significant advancements in the field of intrusion detection, demonstrating superior performance compared to traditional machine learning and other deep learning approaches. Through rigorous training and testing on benchmark datasets like NSL-KDD[11] and UNSW-NB15[9], these models have showcased high classification accuracy and detection rates, reaffirming their efficacy in accurately categorizing various types of network intrusions. Moreover, the incorporation of ensemble methods such as the Voting Classifier (RF+AB) and hybrid models like CNN+LSTM has further enhanced the models' accuracy and robustness, particularly evident with the Voting Classifier achieving 100% accuracy for the KDDCUP99 dataset. Additionally, the inclusion of a Flask-based front end streamlines testing and interaction, offering users a user-friendly platform to assess model performance.

6. FUTURE SCOPE

While the proposed models exhibit promising results, there are avenues for future research and development. Addressing the performance limitations observed with specific attack classes, such as analysis and backdoor classes in UNSW-NB15 and U2R class in NSL-KDD, presents an opportunity for refinement. Furthermore, deploying the models in simulated or real-world applications to assess their effectiveness in practical scenarios remains an important future endeavor. Additionally, exploring techniques for further performance improvement and scalability, as well as enhancing the user interface for enhanced usability, are areas ripe for future exploration and development.

REFERENCES

- [1] L. Dali, A. Bentajer, E. Abdelmajid, K. Abouelmehdi, H. Elsayed, E. Fatiha, B. Abderahim, A survey of intrusion detection system, in: 2015 2nd World Symposium on Web Applications and Networking (WSWAN), 2015, pp. 1–6. doi:10.1109/WSWAN.2015.7210351.
- [2] A. S. Ashoor, S. Gore, Importance of intrusion detection system (ids), International Journal of Scientific and Engineering Research 2 (1) (2011) 1–4.
- [3] Y.-X. Meng, The practice on using machine learning for network anomaly intrusion detection, in: 2011 International Conference on Machine Learning and Cybernetics, Vol. 2, 2011, pp. 576–581. doi: 10.1109/ICMLC.2011.6016798.
- [4] H. O. Alanazi, R. M. Noor, B. B. Zaidan, A. A. Zaidan, Intrusion detection system: Overview (2010). doi:10.48550/ARXIV.1002.4047. URL <https://arxiv.org/abs/1002.4047>

- [5] R. C. Bhagat, S. S. Patil, Enhanced smote algorithm for classification of imbalanced big-data using random forest, in: 2015 IEEE International Advance Computing Conference (IACC), 2015, pp. 403–408. doi: 10.1109/IADCC.2015.7154739.
- [6] F. A. Khan, A. Gumaiei, A. Derhab, A. Hussain, A novel two-stage deep learning model for efficient network intrusion detection, IEEE Access 7 (2019) 30373–30385. doi:10.1109/ACCESS.2019.2899721.
- [7] A. M. M. A. A. Agarwal A, Sharma P, Classification model for accuracy and intrusion detection using machine learning approach., PeerJ Computer Science (2021). doi:7:e437https://doi.org/ 10.7717/peerj-cs.437.
- [8] P. Toupas, D. Chamou, K. M. Giannoutakis, A. Drosou, D. Tzovaras, An intrusion detection system for multi-class classification based on deep neural networks, in: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), 2019, pp. 1253–1258. doi:10.1109/ICMLA.2019.00206.
- [9] N. Moustafa, J. Slay, Unsw-nb15: a comprehensive data set for network intrusion detection systems (unsw-nb15 network data set), in: 2015 Military Communications and Information Systems Conference (MilCIS), 2015, pp. 1–6. doi:10.1109/MilCIS.2015.7348942.
- [10] IXIA Perfectstorm Tool (2017). URL [http://downloads.ixiacom.com/library/user_guides/IxOS/6.60 EA/ EA 6.60 Rev B/IxiaReferenceGuide/PerfectStorm.html](http://downloads.ixiacom.com/library/user_guides/IxOS/6.60_EA/EA_6.60_Rev_B/IxiaReferenceGuide/PerfectStorm.html)
- [11] N. Moustafa, B. Turnbull, K.-K. R. Choo, An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of internet of things, IEEE Internet of Things Journal 6 (3) (2019) 4815–4830. doi:10.1109/JIOT.2018.2871719.
- [12] P. Wu, H. Guo, Lunet: A deep neural network for network intrusion detection, in: 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 617–624. doi:10.1109/SSCI44817.2019.9003126.
- [13] S. Wang, J. Cao, P. Yu, Deep learning for spatio-temporal data mining: A survey, IEEE Transactions on Knowledge and Data Engineering (2020) 1–1doi:10.1109/TKDE.2020.3025580.
- [14] K. Pal, B. V. Patel, Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques, in: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 83–87. doi: 10.1109/ICCMC48092.2020.ICCMC-00016.
- [15] F. Shakeel, A. S. Sabhitha, S. Sharma, Exploratory review on class imbalance problem: An overview, in: 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017, pp. 1–8. doi:10.1109/ICCCNT.2017.8204150.
- [16] T. E. Tallo, A. Musdholifah, The implementation of genetic algorithm in smote (synthetic minority oversampling technique) for handling imbalanced dataset problem, in: 2018 4th International Conference on Science and Technology (ICST), 2018, pp. 1–4. doi:10.1109/ICSTC.2018.8528591.
- [17] W. Satriaji, R. Kusumaningrum, Effect of synthetic minority oversampling technique (smote), feature representation, and classification algorithm on imbalanced sentiment analysis, in: 2018 2nd International Conference

on Informatics and Computational Sciences (ICICoS), 2018, pp. 1–5. doi:10.1109/ICICOS.2018.8621648.

[18] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, G. D. Hager, Temporal convolutional networks for action segmentation and detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[19] K. S. Rana, L.-W. Chen, L.-H. Tang, W.-T. Hong, A study on speech enhancement using deep temporal convolutional neural network, in: 2021 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), 2021, pp. 1–2. doi:10.1109/ICCE-TW52618.2021.9602920.

[20] W. Wang, Y. Sheng, J. Wang, X. Zeng, X. Ye, Y. Huang, M. Zhu, Hastids: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection, IEEE Access 6 (2018) 1792– 1806. doi:10.1109/ACCESS.2017.2780250.

[21] F. Z. Belgrana, N. Benamrane, M. A. Hamaida, A. Mohamed Chaabani, A. Taleb-Ahmed, Network intrusion detection system using neural network and condensed nearest neighbors with selection of nsl-kdd influencing features, in: 2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS), 2021, pp. 23–29. doi:10.1109/IoTaIS50849.2021.9359689.

[22] U. S. Musa, M. Chhabra, A. Ali, M. Kaur, Intrusion detection system using machine learning techniques: A review, in: 2020 International Conference on Smart Electronics and Communication (ICOSEC), 2020, pp. 149–155. doi:10.1109/ICOSEC49089.2020.9215333.

Dataset Links:

NSL – KDD:

<https://www.kaggle.com/datasets/kaggleprollc/nsl-kdd99-dataset>

KDD-CUP:

<https://www.kaggle.com/datasets/galaxyh/kdd-cup-1999-data>

UNSW-15-NB:

<https://www.kaggle.com/datasets/sweety18/unswnb15-training>