Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

TabNet-Based Multi-Class Cyber Security Threat Classification Using Attention-Based Deep Learning

Rathod Maheshwar[1], Dr. Mahesh kandakatla[2]

[1]*Research Scholar, Department of CSE, Chaitanya Deemed to be University, Hyderabad Telangana, India*

[2]*Associate professor, Department of CSE, Chaitanya Deemed to be University, Hyderabad Telangana, India*

ARTICLE INFO

Received: 02 Feb 2023

Revised: 17 March 2023

Accepted: 28 March 2023

ABSTRACT

The growing complexity and the number of cyberattacks have made it imperative to have smart intrusion detection system that can detect multiple network threats in real-time with accuracy. Traditional machine learning techniques are typically highly complex and intricate, and require significant feature engineering, and are usually unable to model the high-dimensional and complex network traffic characteristics. This research presents a deep learning-based framework for cyber security threat classification based on TabNet multi-class intrusion detection model. By incorporating a sequential attention mechanism to automatically select features that are most informative, the proposed model not only achieves state-of-the-art classification results but also introduces intrinsic model interpretability, thus addressing the issue of model explainability. Experimental evaluation was performed on the CIC-IDS2017 benchmark dataset that consists of more than 2.23 million preprocessed network flow instances of benign traffic and various categories of attacks. To ensure the robustness of model training and evaluation, the dataset was preprocessed by removing data duplication, encoding categorical labels, standardizing features, and stratifying the data. For the proposed TabNet model, the overall classification accuracy was 97.92%, which is very high, with weighted precision, recall, F1-score and weighted ROC-AUC of 97.69%, 97.92%, 97.64% and 99.56%, respectively, showing excellent discrimination capability for multiple cyber threat classes. In addition, feature importance analysis revealed the role of important network traffic characteristics in the classification process, making the models more transparent and suitable for explainable artificial intelligence (XAI) in the field of cybersecurity. The experimental results show that the proposed framework is an accurate, scalable, and interpretable approach to intelligent cyber threat classification and has good potential for application in enterprise intrusion detection systems, cloud computing and next generation cyber defense platforms.

Keywords: Cyber Security; Intrusion Detection System (IDS); Multi-Class Threat Classification; TabNet; Deep Learning; Explainable Artificial Intelligence (XAI); CIC-IDS2017; Network Traffic Analysis; Intelligent Intrusion Detection; Cyber Threat Detection.

1. INTRODUCTION

Modern information systems have been revolutionized by the high speed of digital technologies, cloud computing, Internet of Things (IoT), industrial automation and next generation communication networks. The introduction of these technologies has created greater connectivity and operational efficiencies, but has also made critical infrastructure, enterprise networks, health care systems, financial institutions and smart cities more complex and vulnerable to attacks from cyber criminals. New attacks are being developed by cyber criminals that are more sophisticated, such as Distributed Denial-of-Service (DDoS) attacks, brute force attacks, botnet infections, infiltration attacks, web application attacks, and advanced persistent threats (APT) [3] and [4] are examples of patterns that have not been observed before, yet can still be defeated by traditional signature-based intrusion detection systems. Therefore, it is crucial to find intelligent cyber security threat classification to design adaptive and reliable intrusion detection systems for the protection of modern digital infrastructures.Traditional IDSs rely on rules that are manually programmed, attack signatures, or statistical anomaly detection mechanisms. While these

techniques work well for known attacks, they are not capable of automatically identifying new attacks, polymorphic malware and fast evolving cyber-attacks, as they have no ability to learn complex behavioural patterns from network traffic data automatically [4], [6], [15]. Moreover, the proliferation of network traffic volume, speed and variety caused by the cloud computing environments and IoT devices has created many challenges for traditional machine learning algorithms, especially on the processing of high-dimensional cybersecurity data. These restrictions have motivated researchers to look for AI and deep learning methods that can automatically learn and extract meaningful feature representations from vast amounts of network traffic data [7], [11], [18].

In this regard, deep learning is one of the most promising methods to cyber security, due to its capacity of learning hierarchical feature representations directly from raw or minimally processed data. A range of deep neural network architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory (LSTM) networks, autoencoders and hybrid deep learning systems have outperformed most detection and classification methods in the areas of intrusion detection, malware analysis, phishing detection, cybercrime classification and financial fraud detection [1], [8], [9], [13], [17]. Unlike traditional ML based methods where a lot of manual feature engineering is needed, deep learning models learn the complex nonlinear relationships between network traffic features automatically, which enhances the detection performance in the dynamic cyber security environment [6], [7].

There are several recent studies that have investigated Explainable and Intelligent Deep Learning Models for Cyber Threat Detection. The explainable deep learning framework proposed by Khan et al. [2] exemplifies the significance of interpretable AI in the realm of security in Industrial IoT environments. Udayakumar et al. [1] proposed a Deep Fraud Net for cyber security and financial fraud classification based on deep neural networks, and Markkandeyan et al. [16] presented a hybrid deep learning model with optimization algorithms to enhance the performance of cyber threat detection. Likewise, Abu Al-Haija and Zein-Sabatto [20] proposed an efficient deep learning approach to detect cyber-attacks in IoT communication networks, pointing to the necessity of intelligent intrusion detection systems with the ability to process heterogeneous network traffic.

Machine learning and deep learning have now emerged as essential components of the next-generation of cyber defense systems, according to recent survey studies. As pointed out by Ferrag et al. [4] and Shaukat et al. [3] and others, contemporary intrusion detection systems are making greater use of intelligent learning algorithms for the purpose of more accurately detecting attacks with fewer false alarms.As indicated by Geetha and Thilagam [7] and others, Wazid et al. [11] and Ahsan et al. [15] indicate that modern intrusion detection systems are becoming more reliant on intelligent learning algorithms to more accurately detect attacks while minimizing false alarms. These surveys, however, also highlight major ongoing problems such as class imbalance, features of high dimensional network traffic, computational complexity, interpretability of the model, and lack of ability to transfer to various types of attacks. Solving these challenges continues to be a cybersecurity research topic.

Another direction of current research is explainable artificial intelligence (XAI), which seeks to increase the transparency of deep learning systems in critical applications for security. Many deep learning architectures are able to reach high classification accuracy, but their black-box characteristic makes it difficult to deploy them in practice since a cybersecurity analyst may want to understand the reasoning behind the classification of a certain network flow as being malicious. Recent studies have introduced explainable frameworks that show that embedding interpretability in deep learning models can make the models more user-friendly and enable security decision-making [2], [11]. Thus, the classification of cyber threats with explanations has emerged as a key research goal of intelligent intrusion detection systems.

Realistic benchmark datasets have also played a key role in the progress of cyber security studies. The CIC-IDS2017 benchmark dataset has been one of the most popular datasets used to test intrusion detection algorithms, due to its inclusion of realistic enterprise network traffic and several different classes of modern cyber-attacks. In this dataset, benign traffic is also included, in addition to attack categories like DDoS, DoS, PortScan, Botnet, Brute Force, Web Attacks, Heartbleed and Infiltration, which is a comprehensive benchmark to evaluate multi-class cyber security threat classification models [4]. However, learning from these tabular network traffic is still a hard problem for many traditional deep learning architectures.Recently TabNet has appeared to be a powerful deep learning framework specially developed for structured tabular data. TabNet does not use a fully connected neural network but rather uses

a sequential attention mechanism that dynamically chooses the most informative features at each decision step. This attention-based learning process will improve the performance of the classification model without sacrificing interpretability, without relying on manual feature engineering, and without losing attention. This makes TabNet a powerful tool for IDS domains where the traffic is very complex and the features extracted from the network traffic are from a benchmark like CIC-IDS2017.

To address these research issues, this paper introduces an efficient TabNet-based deep learning approach for multi-class cyber security threat classification problem, based on a benchmark dataset called CIC-IDS2017. The proposed framework incorporates a systematic data preprocessing, feature standardization, attention-based feature learning, and comprehensive performance evaluation to accurately classify the benign network traffic and multiple categories of cyber-attacks. The experimental results show that the proposed model can be used to accurately classify the images, with an accuracy of 97.92%, and a precision, recall, F1-score and weighted ROC-AUC of 97.69%, 97.92% and 99.56% respectively. Moreover, the TabNet's built-in feature importance mechanism furnishes interpretable views of the contribution of separate community traffic features, which is helpful for explainable AI in cybersecurity applications.

The main findings of this research are briefly summarized below:

1. The proposed efficient TabNet-based deep learning framework is used to classify multi classes of cyber security threats with the benchmark dataset from CIC-IDS2017.

2. Comprehensive preprocessing pipeline is developed including the removal of duplicate data, encoding of labels, standardization of features and stratified partition of data to enhance the robustness of the developed model.

3. The proposed framework utilizes TabNet's sequential attention mechanism to automatically extract the most informative features in the network traffic from the network traffic data without any manual feature engineering.

4. Extensive experimental evaluation shows high classification performance with an accuracy of 97.92%, a precision of 97.69%, a recall of 97.92%, an F1-score of 97.64% and a weighted ROC-AUC of 99.56%.

5. In addition, the feature importance analysis improves the interpretability of the proposed model, which is useful for explainable artificial intelligence (XAI) in the context of intelligent intrusion detection and practical cybersecurity applications.

This paper is structured as follows. Section 2 discusses the latest research in cyber security threat classification using machine learning and deep learning techniques. The methodology to follow is briefly described in Section 3, which covers data set preparation, data preprocessing, architecture and training strategy of TabNet. The results and discussion are presented in Section 4 and the conclusion and future research direction are outlined in Section 5.

2. RELATED WORK

As cyber threats continue to grow at an ever-faster pace, a broad research effort has been undertaken into intelligent intrusion detection systems which can detect malicious network activities with a high degree of accuracy, and with low false alarm rates. Signature-based intrusion detection systems work well in detecting known attacks but are not very good at detecting new attacks and advanced attack variants. As the result, researchers have been making great efforts to find machine learning and deep learning methods which are capable of learning complex behavioural patterns automatically from network traffic and make the cyber security systems more adaptable [3], [4], [6]. The utility of AI in threat detection has been greatly enhanced in recent years, with systems now capable of extracting features from a variety of network types, adapting to new threats, and making decisions in real-time.

Deep learning is proving to be very successful in Cyber Security due to its ability to capture nonlinear relationships in high-dimensional data sets. Udayakumar et al. [1] introduced Deep Fraud Net, a deep learning-based cyber security and financial fraud detection framework. They used several hidden layers to learn discriminative representations from transaction data, obtaining improved classification performance in their model. The framework was a good success for fraud detection, but in terms of large-scale multi-class network intrusion detection it is still restricted due to the fact that data from financial transactions have very different characteristics from the network traffic produced in enterprise communication systems.Given the rising need for Explainable Artificial Intelligence, researchers have been investigating ways to build interpretable deep learning-based cyber threat detection systems.

Khan et al. [2] proposed an explainable deep learning-based architecture for cyber threat detection in Industrial Internet of Things (IIoT) networks. Their approach introduced explainability features that allowed security analysts to grasp the rationale behind these attack predictions, ensuring the system remained accurate in detecting attacks. The findings showed that the fusion of explainable Artificial Intelligence and deep learning could enhance the confidence of users and enable their implementation in real industrial cyber security environments. The proposed framework was however mainly designed for the scenarios of the industrial IoT and did not explore the large-scale, multi-class intrusion detection with attention-based tabular learning architectures.

There are numerous survey studies that have distilled the accelerated advances made in the field of machine learning and deep learning applications in cyber security. Shaukat et al. [3] summarized the machine learning approaches used in cyber security over the last decade and found that intelligent learning algorithms have made great strides in intrusion detection, malware analysis, phishing detection and cyber threat intelligence. The problems they found are also data imbalance, feature space dimensionality, model scalability, and lack of model interpretability. Likewise, Ferrag et al. [4] performed a comprehensive comparative analysis of deep learning methods applied to IDS and pointed out that CNN, RNN, autoencoders and hybrid deep learning models are increasingly being used. Benchmark datasets like CIC-IDS2017 provide realistic evaluation environments, but noted by the authors efficient feature learning and computational complexity is still an important research challenge.

Important research focus has been cyber threat intelligence. To enhance the understanding of cyber attacks, Yu et al. [5] examined the classification of tactics and techniques found in cyber threat intelligence. Their work showed that intelligent threat classification is essential to enable proactive analysis of threats and incident response. However, the proposed framework was largely categorized using threat intelligence, with little consideration given to real-time intrusion detection based on the features of network traffic.

In [6] Berman et al. gave a detailed survey of the various deep learning techniques adopted for cyber security and explained the benefits of the deep neural network in identifying advanced cyber attacks. Various security applications were discussed such as Convolutional neural networks, Recurrent neural networks, Deep belief networks and Autoencoders. The authors found that deep learning has superior feature extraction ability than traditional machine learning algorithms but found that interpretability, computational costs, and generalization are more important limitations that need to be studied further.

Geetha and Thilagam [7] surveyed machine learning or deep learning applications and compared some of the classification techniques of intrusion detection in cyber security. They found that deep learning models tend to have a higher performance than traditional machine learning algorithms due to their ability to learn automatically hierarchical representations of features from large amounts of data. But the study also showed that much of the current methods are cumbersome, require a lot of hyperparameters to be fine-tuned, and lack transparency when it comes to how they arrive at their decisions.

Another key application area of intelligent intrusion detection are industrial control systems. Al-Abassi et al. [8] suggested an ensemble deep learning approach to cyber-attack detection in industrial control systems. They were able to achieve a high detection accuracy and lower false detection rate by using a combination of multiple deep learning models. While ensemble learning is effective in enhancing predictive accuracy, the training process of multiple neural networks can be computationally intensive, potentially hindering its real-time application in cyber security scenarios where computational efficiency is paramount.

Rapid growth of Internet of Things (IoT) devices has added more cyber security challenges due to the number of resource constrained devices that are interconnected. Ullah et al. [9] proposed a deep learning based cyber threat detection system for IoT environment that is capable of detecting the malicious network traffic created by IoT devices. The experimental results showed that deep learning approach achieved high attack detection accuracy rate than the conventional approaches. However, intrusion detection using IoT is still difficult due to the significant differences in data generated by the variety of devices and communication protocols used in these networks.

The financial cyber security is one arena where machine learning has been applied. Noor et al. [10] developed a framework to detect and attribute a cyber threat to financial technology applications, based on high-level indicators of compromise, with the goal of integrating machine learning. Their framework enhanced the cyber threat attribution

by analyzing the indicators of attacks derived from security incident. The proposed approach improved the attack attribution, but mainly by manually engineered features and study of automatic attention-based feature selection mechanisms that can achieve better classification efficiency was not considered.

With the increasing role of AI in cyber security, Wazid et al. [11] delved into the pros and cons of integrating machine learning and cyber defense techniques. They highlighted the critical role of intelligent learning algorithms in enhancing the accuracy of intrusion detection, malware analysis, anomaly detection, and prediction of cyber threats. The authors, however, noted that there were a few unresolved problems as well such as explainability, computational scalability, class imbalance, and efficient feature selection. The constraints suggest that more research efforts are needed to design deep learning systems that offer high classification accuracy, efficiency and explainable decision making for large-scale cyber security data.

Building on the above discussion, there are a few recent studies that have been dedicated to incorporating the capabilities of machine learning and deep learning along with cyber threat intelligence and Internet of Things (IoT) security applications. In an IoT environment, Mishra et al. [12] proposed a machine learning-based cyber threat intelligence framework that showed that intelligent analytics can be quite beneficial in identifying the malicious communication patterns. They demonstrated the need for data driven security mechanisms to safeguard the interconnectedness of various IoT devices against the ever-changing cyber threats. The proposed framework mainly focused on threat intelligence generation, but did not focus on creating an end-to-end deep learning model that can automatically learn discriminative feature representations from large volumes of network traffic.

AI has been used in the fight against cybercrime for more than just intrusion detection. Ch et al. [13] proposed computational model for detecting the cybercrime offenses with machine learning algorithms. Their study showed it was possible to classify various types of cybercrime incidents using supervised learning and to assist in automated forensic investigations. While good classification performance was obtained, the use of traditional machine learning techniques restricts the potential of the model to capturing the complex nonlinear relationships that are often found in the network traffic data sets of today.

One of the early extensive surveys that discussed the convergence of AI and cyber security was presented by Li [14]. The survey highlighted that the use of AI-powered methods has become a critical pillar of intelligent cyber defence solutions as they are able to analyse large amounts of security data and detect complex patterns of attack. The study also identified some research challenges, such as the scalability, adaptive learning, explainability, and efficient processing of high dimensional cybersecurity data. These challenges persist to drive the creation of more sophisticated deep learning architectures to overcome more complex cyber-attacks.

Ahsan et al. [15] have summarized the recent machine learning based techniques for cybersecurity threat detection and threat mitigation. Their survey showed that AI-based approaches are significantly more effective than traditional signature-based approaches at intrusion detection, malware detection, anomaly detection, and cyber threat prediction. However, the authors found that most of the current deep learning models lack interpretability and require a great deal of computing power to train or make inference. In addition, there is an ongoing need for cross dataset model generalization which needs to be explored.

In recent years, the research has been concentrated in the areas of integrating optimization algorithms with deep learning models to enhance ID performance. Markkandeyan et al. [16] presented an innovative hybrid deep learning model coupled with an optimization approach for the detection of cyber security threats. Experimental results showed the detection accuracy and classification error were improved than those of the conventional deep learning architectures. But the hybrid optimization approach greatly complicates the computation, and it is not so easy to use in a large-scale real-time network systems.

Another crucial use of deep learning in cybersecurity is phishing detection. Mughaid et al., [17] proposed a deep learning based intelligent phishing detection system to detect malicious websites and fraudulent online activities. They found that deep neural networks can successfully learn discriminating features of phishing attacks and decrease the rate of false-positive. While the proposed framework was found to be very effective in detecting phishing attacks, it was specifically developed for the detection of phishing attacks using web traffic and has not been tested for full class intrusion detection using benchmark network traffic datasets.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Artificial Intelligence in cybersecurity applications is still a major topic of research in survey studies, and it has been an expanding trend in recent years. Alghamdi [18] examined various machine learning and deep learning algorithms for cyber security, and found that intelligent learning methods consistently outperform conventional security algorithms when trained on representative sets of data. The review also revealed a number of shortcomings such as model complexity, dataset imbalance, computational expenses, and lack of model explainability, all pointing towards the need of efficient yet explainable deep learning architectures.

In addition, semi-supervised learning has been explored for the identification of cyber threats. Ebrahimi et al. [19] introduced a Transductive Deep Learning approach to detect cyber threats in darknet marketplaces. In the proposed framework, semi-supervised learning was used to classify malicious activities effectively even in the presence of limited labelled data, thanks to the deep neural networks. While this solved a data scarcity problem this technique was applied to analysis of data from darknet markets and not real time intrusion detection in enterprise communication networks.

Abu Al-Haija and Zein-Sabatto [20] proposed a new efficient deep learning framework for cyber-attack detection and classification in the IoT communication networks. In their study, the researchers showed that deep neural networks are good at distinguishing between malicious and benign network traffic, and that they are more useful at detecting attacks than traditional machine learning algorithms. However, the suggested architecture was mainly based on the IoT communication scenarios and was not explored attention-based tabular learning models that can automatically select the informative network traffic features.

IntrudTree is a machine learning-based intrusion detection model proposed by Sarker et al. [21] which is based on decision tree learning and intelligent feature selection. The proposed framework has good classification performance with low computational complexity. However, most of the existing ML models based on trees fail to model high dimensional, nonlinear interactions among features in large-scale cybersecurity data sets, which makes them impractical for enterprise network security data sets.

Recently, Diaba et al. [22] explored the use of meta-heuristic optimization and deep learning algorithms for cyber security applications in the context of smart power systems. Their study showed that optimizing with intelligent learning algorithms could greatly enhance the detection of cyber-attacks in critical infrastructures. While the proposed framework has demonstrated good results in terms of power system security, the specific application precludes direct extension to heterogeneous enterprise network traffic with a variety of classes of cyber-attacks.

In fact, the field of cyber security has seen significant gains in threat detection through the use of artificial intelligence, machine learning and deep learning in a wide range of application fields such as enterprise networks, industrial control systems, IoT environments, financial systems, phishing detection, and protection of critical infrastructure, as evidenced by the existing literature [1-22]. Yet there are some key issues that have yet to be answered in the field of research. First, many current works are based on convolutional neural networks, recurrent neural networks, ensemble learning models or hybrid optimization models that consume much computational resources and hyperparameter tuning [4, 8, 16]. Second, many deep learning models are considered black-box models which are not easily interpretable, limiting their use in practical applications where an understanding of model decisions is crucial in security-sensitive applications [2, 11, 15]. Third, feature selection for high-dimensional tabular network traffic is still an unsolved issue; many existing feature selection approaches either require manual feature engineering, or treat all input features the same regardless of their importance, introducing high computational complexity [3, 4, 6]. Last, there are few studies that have investigated attention-based deep learning architectures with high classification accuracy, computational efficiency, and with embedded interpretability that are applied specifically to structured tabular cybersecurity datasets.

In this regard, this study introduces an efficient TabNet-based deep learning framework for multi-class cyber security threat classification that is based on CIC-IDS2017 benchmark dataset. In contrast to traditional deep learning models, the proposed framework is sequential attention based, which automatically selects the most informative features of the network traffic in each decision step, minimizing the redundant features and improving the interpretability of the deep learning model without the need for manually designed features. The proposed framework attempts to address the problem by applying in-depth pre-processing, attention mechanism-based feature learning,

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

and a broad range of experiments to achieve an accurate, scalable, and explainable solution for intelligent cyber security threat classification for the current enterprise network environment.

3. METHODOLOGY

3.1 Research Framework

In the proposed research framework, a systematic deep learning based cyber security threat classification framework is presented on TabNet architecture with multi-class threat classification. The framework entails a unified intrusion detection pipeline, which is able to precisely categorize malicious and benign network traffic by involving efficient data preprocessing, attention-based feature learning, and extensive performance assessment. The overall process of the proposed framework is presented in Figure 1 which shows the flow of data acquisition, preprocessing, model training, prediction and evaluation stages.

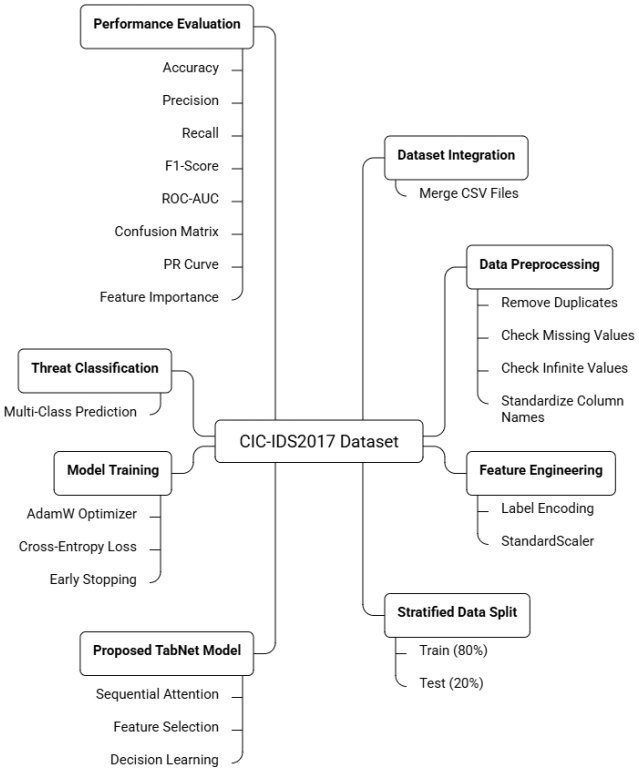

Figure 1. Proposed TabNet-Based Cyber Security Threat Classification Framework

The first step on the framework is to obtain the CIC-IDS2017 benchmark intrusion detection dataset. This data set has been chosen since it is representative of enterprise network traffic, and contains both normal network flows and several up-to-date categories of cyber-attack including Denial-of-Service (DoS), Distributed Denial-of-Service (DDoS), PortScan, Botnet, Brute Force, Web attacks and Infiltration attacks. All files were combined into a single unified dataset in order to get a full perspective of network activity, as the data were spread across files for each day of the network activity. Following integration, the dataset consisted of some 2.31 million network traffic items. Next, duplicate flow records were deleted to remove redundant observations, and 2,231,806 unique network flow instances with 77 numerical features and one target label (traffic class) were obtained.

After the data sets were obtained, a detailed data pre-processing process was applied, to enhance data quality and guarantee reliable model training. The combined data was first checked for missing data, infinite data, duplicate data and data with invalid feature name. The data was then combined and experimental analysis showed that no data was missing or infinite. Only duplicate records were deleted to avoid the effect of network flows on learning, since repeated flows would not be expected to cause any effect on the learning process. Then feature names were cleaned up as much as possible from white spaces to make sure it was consistent for feature extraction and model

implementation. We used label encoding to transform the target attack label into numerical values to allow the deep learning model to make multi-class classification.

The value ranges of the numerical network traffic features are very different, as they are based on different statistical attributes like number of packets, duration of flows, bytes transmitted per flow, header size and timing data. Optimization on such a heterogeneous set of features tends to be unstable, and convergence is also slow when using deep neural networks directly on top of these features. To overcome this problem, standardization of the features was carried out with the help of StandardScaler technique. All numerical features were rescaled to have a mean of zero and unit variance so that the optimization algorithm could process all the input variables on an equal footing while training the model. This pre-processing step helps in achieving the numerical stability, faster convergence and better learning ability of the proposed TabNet architecture.

In order to get an unbiased evaluation of model performance, the standardized data set was split into an 80:20 ratio of training and testing sets, with stratification based on the distribution of its classes. Stratified sampling maintains the same distribution of each attack category in the training and testing sets, thereby mitigating sampling bias, and keeping minority attack classes represented with a sufficient number of samples for development and evaluation of the model. The training subset was used to optimize parameters and the testing subset was used solely for independent performance evaluation.

The TabNet classifier is a deep learning architecture designed for structured tabular data, which is the main part of the proposed framework. In contrast to classical fully connected neural networks, which use all input features over the entire network, TabNet uses a sequential attention mechanism that dynamically decides which input features have the most information at each decision point. The selective feature learning strategy allows the model to learn features of network traffic that are relevant to the classification task and ignore irrelevant or redundant features. The proposed framework has the advantage of not only achieving a better classification performance but also in enhancing the interpretability of the model by determining the contribution of individual features towards the final prediction. This attention-driven learning process decreases the need for manual feature engineering and offers more transparency than normal deep learning techniques.

The model was optimized using the AdamW optimization algorithm with a proper selection of the hyperparameters, such as attention dimension, decision dimension, sparse regularization coefficient, batch size, virtual batch size, and early stopping strategy. While training the model, the accuracy of the validation set was monitored at every step to ensure that the model was not overfitting and to find the best set of model parameters. Early stopping was used to make sure that the last classifier kept the best model weights of training.

The proposed framework was applied to the multi-class prediction of cyber threats in unseen network traffic samples after training the model. The performance of the models was assessed with several quantitative measures, such as classification accuracy, weighted precision, weighted recall, weighted F1-score and weighted Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). Furthermore, the confusion matrix analysis was used to study the prediction capability of the classes, and the feature importance mechanism in the TabNet mechanism was used to identify the most influential network traffic attributes in intrusion detection decisions. The following evaluation techniques are complementary and offer quantitative and qualitative information on the robustness, scalability, and interpretability of the proposed cyber security threat classification framework.

The proposed research framework comprises eight steps: (i) Acquire the datasets, (ii) Integrate the datasets, (iii) Preprocess and remove duplicates, (iv) Encode labels and standardize features, (v) Partition the data into stratified subsets, (vi) Train the TabNet model, (vii) Predict cyber threats, and (viii) Assess the model performance and features. The proposed framework combines the advantages of efficient preprocessing and attention-based deep learning, enabling reliable, scalable, and interpretable solutions for intelligent intrusion detection in today's network environments.

3.2 CIC-IDS2017 Dataset Description

The quality, diversity and representativeness of the data set used for developing and evaluating a model are critical to the effectiveness of any intrusion detection system. For this study, the Canadian Institute for Cybersecurity

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Intrusion Detection System 2017 (CIC-IDS2017) dataset was used as benchmark datasets for developing the proposed cyber security threat classification framework using TabNet. The CIC-IDS2017 is one of the most popular open source benchmark datasets for intrusion detection studies that include attacks from modern times due to its realistic enterprise network traffic and thorough coverage of modern cyber-attacks. Unlike previous benchmark datasets like KDD Cup 1999, NSL-KDD, CIC-IDS2017 features modern attack scenarios and network behaviours that are more representative of current cybersecurity environments.

This dataset was created by the Canadian Institute for Cybersecurity running with realistic network infrastructure, emulating normal network activities and multiple attack scenarios to produce the dataset. Legitimate uses of the network, such as web browsing, emailing, transferring files, multimedia streaming, and secure remote access were mixed with various types of cyber-attacks spread out over several days. This realistic traffic generation approach allows the dataset to closely duplicate real-world enterprise network environments and to have well-defined attack labels for supervised learning. As a result, CIC-IDS2017 is a benchmark widely used to test intrusion detection algorithms using machine learning and deep learning techniques.

The original dataset is packaged in a number of traffic files, each of which represent a day's worth of network activity. The data is stored in each file as bi-directional network flow statistics, which are acquired with the CICFlowMeter tool, which translates raw packet captures to flow-based numerical features. These features refer to a number of statistical characteristics of network communication, such as the number of packets, flow duration, inter-arrival times, packet length, header data, transmission rate, active duration, idle duration, flow behaviour characteristics etc. These characteristics are adequate to detect, without analyzing the packet payload, various types of cyber attacks from benign traffic.

All data files were merged into a single structured data set to create an integrated learning dataset. After integration, there were about 2,313,810 network flow instances in the data set. After preprocessing, there were no missing values or infinite values in the merged dataset, as shown by a detailed inspection of the dataset. Flow records were then duplicated to exclude redundant records that might inflate classification accuracy and lead to biased model learning. The cleaned dataset for analysis was 2,231,806 instances of network traffic. The features of numerical network traffic were 77 in total and each instance had one category target label for the network traffic class to which it belonged.

The target variable is a combination of normal and malicious network traffic behaviors. There are categories of attacks such as: Benign, DoS Hulk, DoS GoldenEye, DoS Slowloris, DoS SlowHTTPTest, DDoS, PortScan, Bot, FTP-Patator, SSH-Patator, Web Attack–Brute Force, Web Attack–XSS, Web Attack–SQL Injection, Heartbleed, Infiltration. The presence of several attack families allows all aspects of the proposed framework to be tested in many different scenarios of intrusion and simulates the complexity that can be found in the real network security environment.

Features of the CIC-IDS2017 dataset are extremely heterogeneous and have very different numerical ranges. For instance, the numbers of packets, bytes transmitted, duration of flows, packet timing information and flow rates vary significantly in terms of their value and distribution. If not done correctly, this heterogeneity is a challenge for optimization of deep learning. Thus, feature standardization was applied as a preprocessing step to make sure that all the numerical variables contribute to the learning of the model proportionally. This preprocessing method stabilizes the optimization process and simplifies the identification of the most informative parameters of network traffic by the TabNet architecture.

Another crucial data feature is that the data is also imbalanced. As in the real world, benign packets and some types of attack have significantly more samples than more uncommon attacks like Heartbleed or SQL Injection. The original distribution of the classes was maintained and stratification was used to ensure proportional representation of each traffic class in the train-test split to develop and evaluate the model. This approach ensures that both the majority and minority classes are present in the training and testing sets and is a more realistic representation of the classification performance.

The CIC-IDS2017 dataset caters for the evaluation of intelligent intrusion detection systems with a realistic enterprise network traffic, multiple contemporary attack categories, large-scale flow-based observations and high dimensional statistical features. The dataset is very well suited to test the performance of attention-based deep learning models

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

like TabNet, which are specifically developed for learning complex relationships from structured tabular data. The summary of characteristics of the datasets used in this study are shown in Table 1.

Table 1. Characteristics of the CIC-IDS2017 Dataset Used in This Study

Parameter	Description
Dataset	CIC-IDS2017
Source	Canadian Institute for Cybersecurity (CIC)
Data Type	Network flow records
Original Records	2,313,810
Records After Duplicate Removal	2,231,806
Number of Features	77
Target Variable	Multi-class attack labels
Traffic Classes	Benign and multiple cyber-attack categories
Feature Type	Numerical network flow statistics
Missing Values	None
Infinite Values	None
Duplicate Records	Removed during preprocessing
Data Partitioning	Stratified 80:20 Train-Test Split

3.3 Data Preprocessing

As the quality of input data determines the learning capability of deep learning algorithms, data preprocessing is very important in the development of an accurate and reliable intrusion detection model. Network traffic datasets usually include inconsistencies like duplication of network traffic, heterogeneous feature names, numerical scales that vary across features, and categorical class labels that need to be converted to machine-readable representations before using them for model training. Hence, a preprocessing pipeline was designed systematically to further improve the quality of the data, to homogenize it and to boost the performance of the proposed cyber security threat classification framework based on TabNet.

The first step in the preprocessing was to combine all network traffic files present in the CIC-IDS2017 data-set into a single structured data-set. The benchmark dataset was spread across several files, one for each day of network activity, so that all of the files was combined, the model could learn from a comprehensive set of normal network activity and a wide range of cyber-attack scenarios. After integration, the integrated dataset contained 2313810 network flow records with 77 numerical traffic features and one label class for the target.

The quality of the combined data set was then checked by a series of checks. The dataset was first checked for missing values, which may have a negative impact on the optimization of models. The experimental analysis showed there were no missing data in the combined data set. In a similar manner, the data was checked for infinite numbers that could occur in the process of feature extraction due to divisions or unusual network flow data. During the inspection it was found that there were no infinite values and that no further value replacement or imputation method was needed. These observations show the full and consistent dataset of the benchmark employed in this study.

While there are no missing or infinite values in the dataset, duplicate network flow records were found during data inspection. Redundancy in observations can lead to over-computation, more frequent observations, and a bias of the traffic patterns in the learning process, as the classifier is trained to favour the more common traffic patterns. Hence, duplicate samples were eliminated prior to the development of the model. This pre-processing step reduced the size of the dataset from 2,313,810 to 2,231,806 different network traffic records, maintaining the same distribution of

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

attack types as the original dataset. Data deduplication increased data diversity and decreased the risk of data over-fitting when training the model.

The consistency of the dataset schema was ensured after the duplicate removal process in order to have a standardized schema for feature extraction. Some feature names had unneeded leading or trailing white space that was carried over from the source data set. This may result in inconsistencies that can lead to feature selection errors in the implementation process. So, all column names were normalized, meaning that white spaces were removed from the column names, so that a uniform feature naming was used throughout the process of preprocessing and training.

The data set of the CIC-IDS2017 contains a categorical class labels for the class labels of each class (benign traffic and various categories of cyber attacks). Label encoding was used for numerical representation, as deep learning algorithms demand numerical readings of the attack labels. This encoding ensures that the class identity is maintained and optimisation of multi-class classification is possible. The encoded labels were used as the target variable during training and evaluation.

Numerical inputs describe a variety of characteristics of network flows such as packet statistics, byte counts, timing, flow duration, transmission rates and protocol-specific measurements. These variables span a wide range of numerical values, so their direct use without normalization could lead to numerical instability in the gradient updating process and slower convergence while training the TabNet model. In order to overcome this, feature standardization was carried out by implementing the technique of StandardScaler from the Scikit-learn library. StandardScaler is a feature-scaler that centers and scales each feature to have zero mean and unit variance. The normalization strategy will avoid the domination of high-magnitude features and enhance the convergence behaviour of the deep learning model.

The data was then preprocessed, standardized and split into predictor variables and target labels. The standardized features of the network traffic were placed in the predictor matrix while the encoded categories of the attacks were incorporated into the target vector. The dataset was split in a stratified way (80:20 split) to get an accurate estimate of the model generalization performance. This means that stratified sampling maintains the relative proportion of each traffic class both in the training set and in the testing set, and thus eliminates sampling bias and also ensures that minority attack traffic classes are not omitted during testing. The training subset was used to tune the parameters of the TabNet model, while the testing subset was used only for independent performance testing.

The pipeline implemented in this study offers a number of benefits in the preprocessing. One, duplicating eliminates duplicate learning and increases the diversity of training samples. Second, label encoding allows for an efficient multi-class prediction which does not change the semantic meaning of the attack categories. Thirdly, feature standardization gives a more stable optimization and quicker model convergence. Lastly, stratified data partitioning ensures equal representation of attack classes while training and testing the model. In summary, these preprocessing steps create a high-quality input dataset that facilitates the proposed TabNet architecture to learn meaningful feature representations for accurate cyber security threat classification.

3.4 Proposed TabNet Model

This proposed cyber security threat classification framework is based on the TabNet architecture as the main deep learning model for multi-class intrusion detection systems. TabNet is particularly well suited to structured tabular data and it offers a high level of interpretability with sequential attention and feature transformation networks. TabNet is different from traditional deep neural networks because it only uses the most informative features at each decision step, which allows to learn efficiently from high dimensional network traffic. This learning approach is based on attention, where redundant attributes are minimized and the model is enhanced in its ability to detect complex intrusion patterns in network traffic.

The model takes as input the network flow features of standardized CIC-IDS2017. Each network traffic instance is then mapped to a numerical feature vector with 77 features that are standardized by dividing by the maximum possible value, and represent flow duration, packet statistics, number of bytes, inter-arrival times, transmission rates,

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

and protocol-related characteristics. These feature vectors are fed directly into TabNet network without the need for any manual feature engineering, and the model learns the discriminative feature representations during training.

The TabNet architecture is a sequential architecture with multiple decision steps, each of which is tasked with selecting and processing the most relevant subset of features. The model only requires some of the input variables during the whole learning process, using an attentive transformer to create feature-masks to emphasize which features should be highlighted at each step of making a decision. In later stages, previously selected features are not paid as much attention by means of a feature reuse mechanism (controlled by the relaxation parameter γ). This sequential attention strategy allows the network to extract complementary information from various subsets of network traffic features without introducing a lot of redundant information.

There are two major functional components in each decision step: a feature transformer, and an attentive transformer. The feature transformer is a fully connected layer, normalization, and activation operation to produce high-level features representations. The transformer also learns feature importance scores and generates sparse feature masks for selection of informative features for the subsequent decision step. This repetitive process of feature transformation and attention enables the network to gradually get closer and closer to the representation of network traffic patterns and enhance classification accuracy.

TabNet's unique feature is its built-in interpretability. The spARSIFORM generated attention masks during learning give quantitative feedback about the importance of each network traffic feature for the final prediction. The model is thus not only able to classify the cyber threats, but also to determine the most influential traffic properties which help to distinguish benign communication from malicious activities. This allows explainable AI (XAI) without the need for any further post-hoc explanation methods.

The TabNet classifier was set to have a decision dimension of 32, an attention dimension of 32, and 4 sequential decision steps (n_steps = 4). These are the parameters that enable the network to have a high representational capacity while remaining computationally efficient, when handling large scaled network traffic analysis. Relaxation parameter was chosen as $\gamma = 1.3$, which lets some features be reused across subsequent decision levels, thus allowing the model to learn complementary information from the features already used on the previous decision level. Additionally, a sparse feature selection was promoted by using a regularization coefficient: $\lambda\_sparse = 1 \times 10^{-5}$, which leads to compact feature masks and enhanced generalization abilities.

The optimization of the models was carried out with the AdamW optimizer that improves the convergence stability and avoids overfitting by combining adaptive gradient estimation and weight decay regularization. The initial learning rate was initially determined to be 0.005 and weight decay regularization was used to ensure a limited growth of the parameters during the optimization process. To ensure smooth convergence to an optimal solution in optimization process, a StepLR learning-rate scheduler was used to reduce learning rate gradually during training. To achieve efficient utilization of the computational resources and stable estimation of the gradients when training on the large-scale CIC-IDS2017 dataset, a batch size of 8192 and a virtual batch size of 1024 were used, which is known as batch learning.

In order to enhance the generalization of the model, early stopping was also added during training. The accuracy of the validation set was monitored and training was stopped if the accuracy did not improve after 10 epochs. The parameters of the model with the best validation accuracy were automatically reconstructed, meaning the resulting classifier was the best performing one found during training, and was unlikely to overfit.TabNet architecture's output layer classifies each instance of network traffic to an attack category or to the benign traffic class, which are defined beforehand. During inference, the model outputs a probability distribution for each traffic class, and the traffic class with the most significant probability is deemed as the predicted network state. The probabilistic prediction mechanism allows for the precise classification of several categories of intrusion and for the evaluation of the performance of the system with the help of indices like accuracy, precision, recall, F1 score, and ROC-AUC.

In general, the proposed TabNet architecture has several benefits for cyber security threat classification. First, it automatically discovers the most informative network traffic features, without manual feature selection. Second, by learning sparse features, the model training process is more efficient and avoids feature redundancy. Third, the true interpretability of TabNet allows feature importance analysis, aiding for transparent and explainable intrusion

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

detection. Finally, the proposed framework is capable of accurately and scalably classifying cyber threats in multi-class settings, given a large-scale network traffic dataset, thanks to combination of attention-based feature selection, adaptive optimization and robust regularization.

3.5 Hyperparameter Configuration

Deep learning models highly depend on the choice of the right hyperparameters. The hyper-parameters of the proposed TabNet classification algorithm were carefully selected to balance the classification accuracy, computational efficiency, and the generalization of the model. The configuration was identified by performing several experimental trials taking in to account the large scale of CIC-IDS2017 dataset.

A decision dimension n_d = 32 and an attention dimension n_a = 32 were used to initialize the TabNet classifier, enabling the model to learn discriminative representations of features while keeping the complexity low. The number of sequential decision steps (n_steps) was set to be 4, enabling the network to selectively learn features and progressively remove informative representations from the network traffic features in an iterative manner. A relaxation coefficient (γ) of 1.3 was used to control the re-use of features across successive decision steps, to prevent too much redundancy in the learning process. To further encourage sparse feature learning, a sparse regularization coefficient (λ_sparse) was set to 1×10^{-5}, which helps to select features efficiently and enhances the generalization performance of the model.

The weight decay regularization with adaptive gradient estimation was combined with the AdamW optimization algorithm, as it was capable of optimizing the model. The initial learning rate was chosen to be 0.005 and weight decay was added to prevent overfitting during optimization. To achieve stable convergence and avoid oscillations around the optimal solution, a Step Learning Rate (StepLR) was used to decrease the learning rate during the training. The training process was conducted on a large scale with a batch size of 8192 and a virtual batch size of 1024, ensuring efficient use of computational resources and ensuring the stability of gradient estimation. The maximum training epochs were limited to 50, and early stopping with a patience value of 10 epochs was used to automatically stop the training process when there was no improvement recorded in the validation accuracy. The approach avoids unnecessary calculations and maintains model parameters that best match the validation results.

These hyperparameters offer a good balance between the complexity of the model and its predictive power, allowing the TabNet model to handle the large volume of network flow data efficiently – over two million records – while still achieving high classification accuracy. The full set of hyperparameters used in this study is presented in Table 2.

Table 2. Hyperparameter Configuration of the Proposed TabNet Model

Hyperparameter	Value
Decision Dimension (n_d)	32
Attention Dimension (n_a)	32
Decision Steps (n_steps)	4
Relaxation Parameter (γ)	1.3
Sparse Regularization (λ_sparse)	1×10^{-5}
Optimizer	AdamW
Initial Learning Rate	0.005
Learning Rate Scheduler	StepLR
Batch Size	8192
Virtual Batch Size	1024
Maximum Epochs	50
Early Stopping	Patience = 10

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Loss Function	Cross-Entropy Loss

3.6 Model Training Strategy

A supervised learning approach was used to train the proposed TabNet model for the preprocessed CIC-IDS2017 dataset. After standardization of features and encoding of the labels, the data set was split into training and test sets with an 80:20 ratio where the ratio was stratified.After feature standardization and label encoding, the data set was divided into a 80:20 train-test ratio with stratification. As a result of the stratified partitioning, each attack category was represented roughly in the same proportion in both subsets, eliminating sampling bias and allowing an accurate assessment of the performance of all the traffic classes.

For training, the TabNet classifier was fed with standardized network traffic features while the attack labels were encoded as the target variable. The parameters of the model were optimized with backpropagation with an AdamW optimization algorithm. The prediction errors at each epoch were calculated by cross entropy loss function, and the weights of the network were adjusted iteratively to make the classification error as small as possible. The learning rate scheduler used during the training process dynamically changed the optimization step size during training to ensure smooth convergence and better optimization stability.

An early stopping mechanism was adopted when training the model to prevent overfitting. The training accuracy was checked after every epoch and training was stopped if there was no gain for 10 epochs. Before the final evaluation the best performing model weights that were obtained during training were automatically restored. This strategy can enhance the generalization of the model and reduce the unnecessary computational cost.

Thanks to the combination of large batch processing and virtual batch normalization, they were able to efficiently learn from the large-scale CIC-IDS2017 dataset of over 2 million network traffic records. Once trained, the improved TabNet classification model was used to classify real-world unseen traffic instances in the testing set, allowing the classification model to be evaluated and tested objectively.

4. RESULTS AND ANALYSIS

4.1 Training Performance Analysis

The training behavior of the proposed cyber security threat classification model based on TabNet was studied based on the training loss and the accuracy of the validation obtained during the optimization. These learning curves provide valuable insights into the convergence characteristics, optimization stability, and generalization capability of the proposed model. Figure 2 shows the training loss during training, and Figure 3 shows the training accuracy at each training step during training.

As seen from the training loss, the learned by the proposed TabNet model also shows a sharp decrease in the initial training stage as shown in figure 2, revealing that the TabNet network quickly captures the underlying patterns existing in the network traffic dataset CIC-IDS2017. The loss evaluated at the start of the training process is around 0.28, which is large given the large prediction error when the model's parameters are initialized randomly. The loss value drops significantly to almost 0.05 after the first epoch, which shows the importance of the attention learning mechanism in extracting discriminative features from benign and malicious network traffic. This huge decrease in the early training stage shows that the hyperparameter optimisation was done efficiently, and that the chosen hyperparameter setting allows for stable convergence.

After the first epoch, the training loss gradually decreases until it is about 0.02 at the last epoch. The convergence trend of the obtained loss curve is smooth and monotonic, unlike the unstable optimization processes which have oscillating or increasing loss values. The behaviour shows that the AdamW optimizer, along with the chosen learning rate scheduling strategy, was successful in maintaining a decrease in the classification error without causing the parameters to become unstable. Moreover, the lack of sudden changes in the loss curve indicates that the model captured representative pattern of features in the standardized network traffic data and remained numerically stable during optimization.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The validation accuracy curve also illustrated in Figure 3 further indicates the learning capability of the proposed framework. The validation accuracy was close to 97.9% during the first epoch, which shows that TabNet architecture quickly transferred to unseen network traffic samples. There were minor variations observed for intermediate epochs, with the accuracy of the validation epochs ranging from around 95.2% to 96.5%. These minor deviations should be seen as part of the search process of deep learning optimization, as the model adjusts its internals as it learns in the process. Most important, the fluctuations were fairly small and were not a sign of unstable training or substantial overfitting.

As training continued the accuracy of validation slowly improved and stabilized, finally converging to about 97.4% in the last epoch of training. The high level of consistency between the gradually declining training loss and constantly high training accuracy for the validation sets confirms that the proposed model was able to achieve good learning capacity and generalization. This behaviour was also further reinforced by the early stopping criterion used during training that kept the model parameters with the best validation performance and prevented over-optimization after convergence.

As can be seen in Figures 2 and 3, the learning behaviour exhibited has several key attributes of the proposed TabNet framework. First, the training loss reduction is shown to be fairly rapid, indicating that the sequential attention mechanism is able to learn informative network traffic features at the start of the optimization. Second, the accuracy obtained is stable during the validation showing that the model generalizes to unseen network traffic and does not memorize the training samples. Third, no significant oscillations on either learning curve indicate that the chosen hyperparameters (decision dimension, attention dimension, learning rate, sparse regularization coefficient) are well balanced between the convergence speed and model stability.

One more remarkable finding was that the model was able to reach a high validation accuracy after just a few training epochs, despite having been trained with over 2.23 million network traffic records. The efficiency of the TabNet architecture on large-scale cybersecurity data is shown. The sequential attention mechanism allows the model to only pay attention to the most informative traffic features, instead regarding all features equally, helping to improve the efficiency of learning, with less unnecessary computational complexity.

In general, the learning curves show that the proposed TabNet classifier is efficient in converging, and its optimization remained stable throughout the training process, while it has good generalization performance with moderate levels of overfitting. The above observations show that the combination of the preprocessing technique (standardization of features) and the hybrid optimization algorithm and the associated hyperparameter settings offer a strong basis for achieving an accurate classification of multi-class cyber security threats.

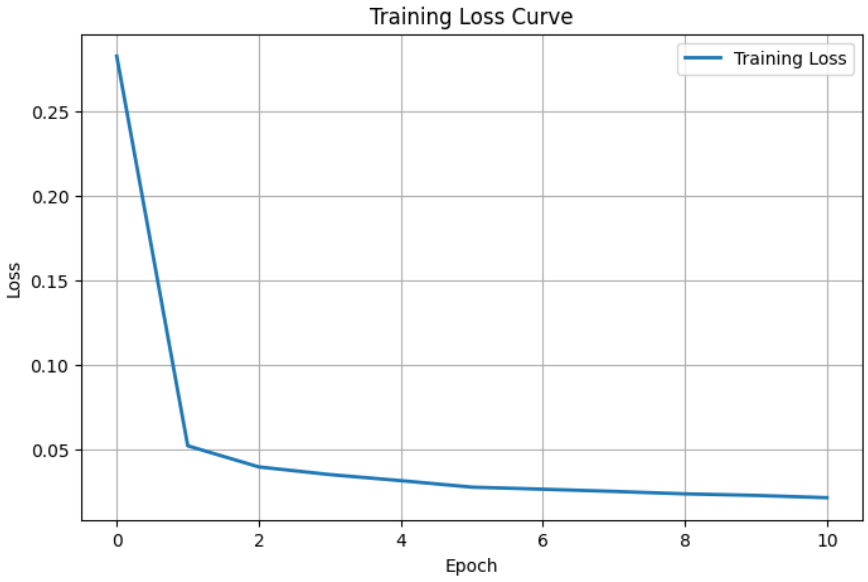

Figure 2. Training loss curve of the proposed TabNet model.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

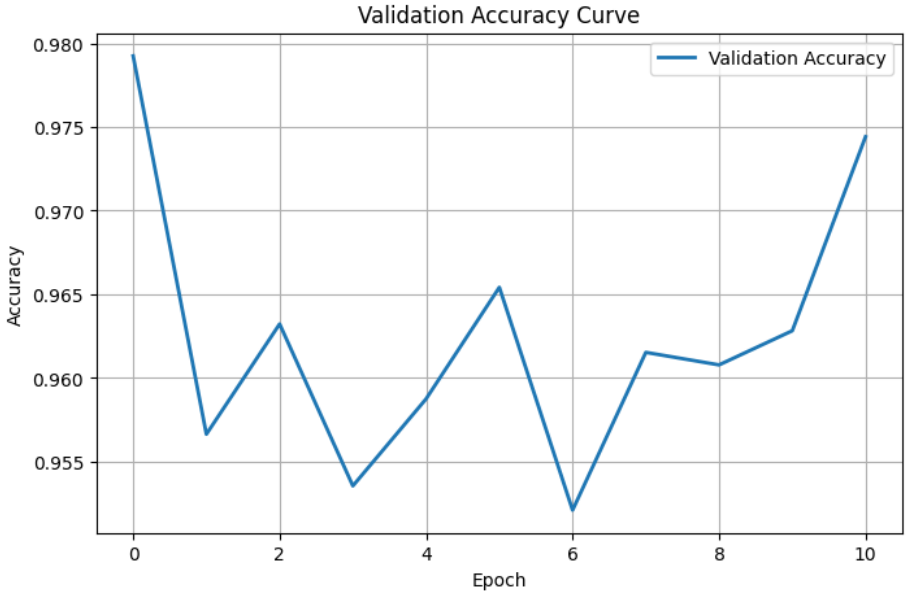

Figure 3. Validation accuracy curve of the proposed TabNet model showing stable convergence and high generalization performance.

4.2 Multi-Class ROC Curve Analysis

The proposed classification ability of the cyber security threat classification framework with TabNet was still assessed by using ROC curve and Area Under the Curve (AUC) values. The ROC curve is a very popular metric used for evaluation of multi-class classifiers as it provides an insight into how well the classifies can separate various classes for various levels of decision thresholds. The ROC analysis evaluates the trade-off between TPR and FPR, thus giving a more detailed measure of the discrimination capability of the model; classification accuracy only gives one performance measure. The multi-class ROC curves for the proposed framework are shown in Figure 4.

Most ROC curves are very close to the top left corner of the graph as shown in Figure 4, and thus offer a very good classification performance in most of the attack categories. The best classifier has an ROC curve that is closer to the top left corner, and an AUC of 1.0 indicating perfect separation between classes. The results showed that the proposed TabNet model achieved good learning of the highly discriminative representations for the most part of the network attack categories, achieving a good separability between benign traffic and malicious network flows.

The experimental evaluation found that a number of the attack categories had AUC over 0.99, which is near perfect classification accuracy. DoS Hulk, DoS GoldenEye, Benign, Bot, FTP-Patator, SSH-Patator and Web Attack – XSS had AUC performance values of 0.999, 0.994, 0.995, 0.958, 0.988, 0.995 and 0.984 respectively, which were also very good. The high AUC values are a sign that the proposed attention-based learning mechanism has been able to learn the distinctive features of these network traffic classes, and has been able to discriminate them from the other categories of attacks.

The proposed framework also achieved a high weighted ROC-AUC score of 99.56%, further demonstrating its good discrimination performance overall. This result shows that the classifier gives more prediction probabilities to the same attack categories for different classification thresholds. The high weighted ROC-AUC value complements the overall classification accuracy of 97.92%, which shows that the model has a good prediction effect even when the threshold is changed.

While most of the attack categories had excellent results, some minority classes had relatively poor AUC scores. For instance Heartbleed (AUC = 0.474) and Infiltration (AUC = 0.708) showed less discrimination capability than did the majority attack classes. Likewise, the classification performance of PortScan (AUC = 0.949) and Web Attack – SQL Injection (AUC = 0.855) was in the middle range. This is to be expected since these attack categories have relatively low amounts of training samples throughout the CIC-IDS2017 dataset. The ability of deep learning models

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

to learn a robust decision boundary typically relies on the availability of adequate representative examples, and with limited training instances it is possible to lose the power of the classifier to distinguish very few attack patterns from benign traffic or other attack patterns.

The class-wise differences in the AUC values observed also show the class imbalance in the CIC-IDS2017 benchmark dataset. In the case of a frequent attack, like DDoS or DoS attacks, there are many examples that can be used to train the model to learn highly discriminative feature representations. Minority attack classes, on the other hand, have a smaller number of observations during the optimization phase, which leads to less confidence during classification. Nevertheless, the proposed TabNet framework achieves very high overall classification accuracy, showcasing its strength to deal with large-scale and heterogeneous cybersecurity datasets.

Yet another noteworthy thing to be noticed from Figure 4 is that most of the ROC curves were far from the diagonal reference line, which is the line of random classification. This means that the proposed classifier achieves performance levels much higher than chance in almost all attack categories. This behavior is also a direct result of the sequential attention mechanism of TabNet during each decision step, where the most informative network traffic features are automatically selected to enhance the model's ability to tell complex patterns of intrusions.

In general, the multi-class ROC analysis shows that the proposed cyber security threat classification framework with TabNet has good discrimination power for most of the attack classes with high weighted ROC-AUC of 99.56%. The results validate the complex relationship among the features of network traffic and show that the attention-based learning approach is effective and can be trusted in modern cyber security environments for multi-class intrusion detection. Class balancing methods or synthetic data augmentation can be used to improve the performance of classification for rare attack types in the future.

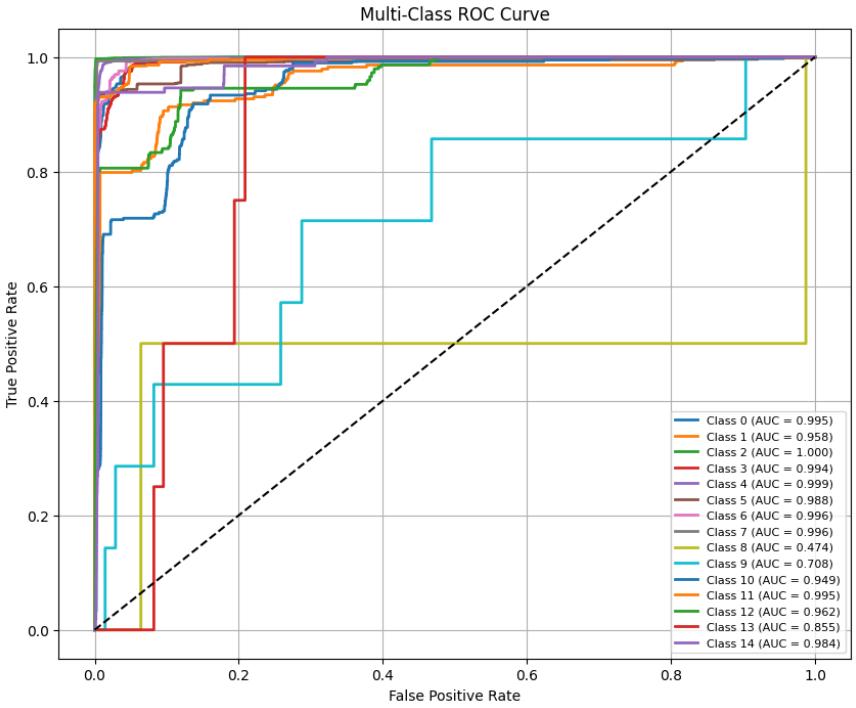

Figure 4. Multi-class Receiver Operating Characteristic (ROC) curves of the proposed TabNet model for the CIC-IDS2017 dataset. The majority of attack categories achieve AUC values close to 1.0, demonstrating excellent discrimination capability, while lower AUC values for minority classes primarily reflect the inherent class imbalance of the dataset rather than instability in the proposed classification framework.

4.3 Confusion Matrix Analysis

Further, to assess the classification capability of the proposed cyber security threat classification framework based on TabNet, confusion matrix was created based on the prediction results from the testing set. Evaluation metrics like

accuracy and ROC-AUC give a general idea about the performance of the model, but the confusion matrix gives a detailed class-wise analysis and shows how many samples are correctly/incorrectly classified for each network traffic category. This allows for identification of attack classes that are correctly classified as well as classes with a higher misclassification rates. The confusion matrix of the proposed model is shown in Figure 5.

Figure 5 shows that most of the samples in the network traffic are clustered around the main diagonal of the confusion matrix. The diagonal elements are correct classifications while the off-diagonal elements are misclassifications. The figure shows high diagonal dominance which demonstrates a high accuracy of the proposed TabNet model in correctly identifying most of the categories of benign traffic and cyber attacks. The overall correct classification of the model was 97.92%, 97.69% accuracy, 97.92% recall, and 97.64% F1-score, which is consistent with the observation.

The samples of the Benign traffic class have the most correct classifications with 378,538 correct samples. The proposed model's ability to distinguish between legitimate network activity and attack categories was demonstrated by only a relatively small number of benign traffic being misclassified as attack categories. This is especially significant in real-world intrusion detection systems as it decreases the amount of false alarms and unnecessary security alerts, promoting more efficient operation.

When compared to the various attack types, the proposed model showed high recognition accuracy for high-frequency attacks like DDoS, DoS Hulk, DoS GoldenEye, DoS SlowHTTPTest and FTP-Patator. For instance, the DDoS class is correctly classified 24,523 times while the remaining few instances are wrongly categorized. Likewise, the DoS Hulk category had 30,779 correctly classified instances, showing that the attention-based feature learning mechanism was able to detect the discriminative traffic attributes of volumetric DoS attacks.

The confusion matrixes do not show good classification results in the majority of samples are located in the principal diagonal and there are few misclassifications in the other positions for DoS GoldenEye, DoS SlowHTTPTest and FTP-Patator classification. The proposed TabNet architecture is able to capture the statistical associations between packet-level and flow-level features characterizing these attack categories, as seen from these results. The sequential attention mechanism allows the classifier to concentrate on the more meaningful network traffic features, leading to better classification of similar attack patterns in terms of their behaviour.

While the proposed framework had good overall performance, the confusion matrix shows that some minority attack classes had relatively high misclassification rate. The CIC-IDS2017 dataset includes a limited number of training sample sets in classes like Heartbleed, Infiltration, Web Attack – SQL Injection, and Web Attack – Brute Force. Therefore, there were some instances within these rare attack categories which were misclassified as either benign traffic or other attack categories. This behaviour is expected in highly imbalanced cybersecurity datasets, where the classifier gets much less representative examples of rare attack patterns during the training process.

The SSH-Patator and DoS Slowloris categories are similar, with only a small number of samples identified as benign traffic. Some of these misclassifications indicate that some attack flows have statistical properties resembling that of normal network communication and this makes them hard to distinguish. However, the number of samples well classified is still significantly larger than the number of the mispredictions, which shows the strength of the proposed approach.

The confusion matrix also shows that the majority of errors in classification are made between classes with similar behaviours, and not between classes with totally different behaviours. This observation indicates that the model is able to learn the overall difference between benign and malicious traffic and has only limited problems of distinguishing between attacks with similar network flow features. Multi-class intrusion detection is a very common scenario since multiple denial-of-service attacks have similar packet transmission patterns, flow durations, and communication statistics.

Overall, the confusion matrix validates the proposed cyber security threat classification framework based on TabNet as reliable multi-class intrusion detection for the CIC-IDS2017 dataset. The high diagonal concentration of correctly classified samples indicates that the sequential attention mechanism is a good way to learn discriminative feature representations from high-dimensional network traffic data. While the classification performance reduces in the case

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

of minority attack classes due to dataset imbalance, the proposed framework effectively identifies the major cyber-attack classes while retaining excellent overall predictive performance.

The validation of the effectiveness of the proposed model is therefore validated from the point of view of the confusion matrix analysis, which proves that the proposed model has achieved 97.92% classification accuracy, this is achieved in the classes rather than in a single class of traffic.

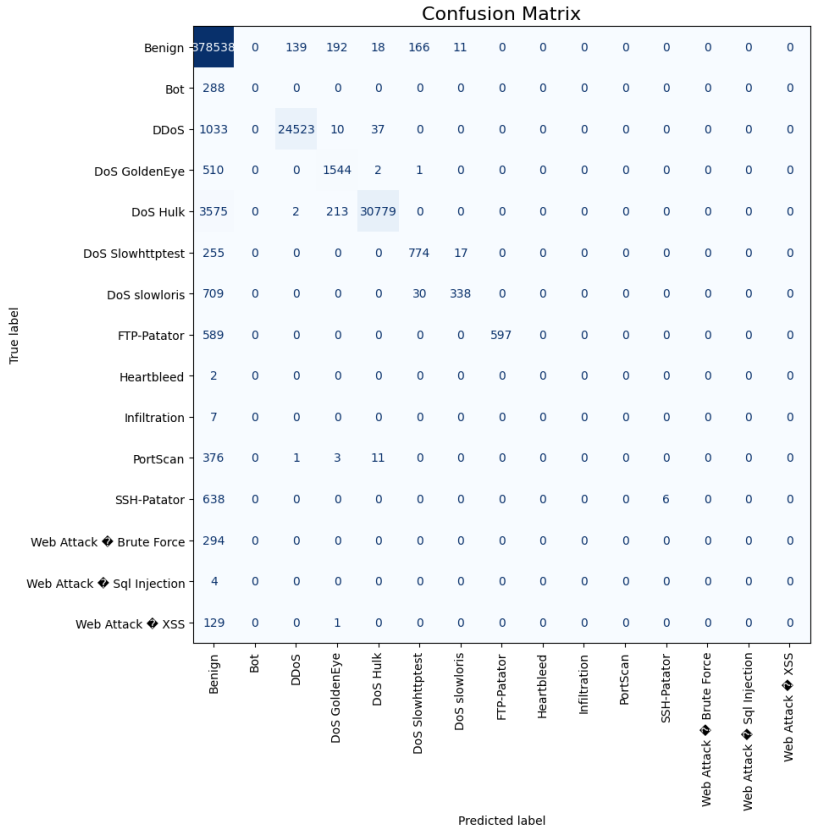

Figure 5. Confusion matrix of the proposed TabNet-based cyber security threat classification framework on the CIC-IDS2017 dataset. The strong diagonal pattern indicates accurate classification for the majority of benign and attack traffic classes, while the limited off-diagonal elements primarily correspond to minority attack categories affected by class imbalance.

4.4 Normalized Confusion Matrix Analysis

The confusion matrix was also normalized by calculating the percentage of correct and incorrect predictions for each class, to get a better sense of the prediction performance of the proposed TabNet framework at the class level. The normalized confusion matrix displays the classification accuracy for each traffic category without taking into account the number of samples. This representation is especially useful when the data set is highly imbalanced (as in CIC-IDS2017, the classes of attack are minority classes that contain much fewer samples than the majority classes). The normalized confusion matrix of the proposed model is shown in Fig 6.

The normalized confusion matrix, shown in Figure 6, reflects good diagonal dominance for most of the classes of network traffic, with a high percentage of the samples in each category being classified correctly. The numbers along the main diagonal are the recall value that the classifier obtained in each class, while the numbers off of the diagonal are the number of misclassified samples in each class. The high diagonal values indicate that the proposed TabNet architecture learned discriminative feature representations that were able to differentiate between the benign traffic and multiple cyber-attack categories.

The Benign network traffic class had a normalized classification score of 1.00, indicating that almost all the instances of legitimate network traffic were identified correctly. This is very important for real-world intrusion detection

systems (IDS) as it means that if benign traffic can be correctly identified, then false alarms are reduced and unnecessary security measures taken. Likewise, attacks like DDoS, DoS Hulk, DoS GoldenEye and DoS SlowHTTPTest have relatively high diagonal values, suggesting that the proposed framework is capable of detecting these high frequency attacks with a limited number of false positive samples.

The normalized matrix also shows the usefulness of the proposed TabNet model to identify various categories of attacks with similar network traffic properties. For example, the DDoS class was correctly classified at around 96% and DoS Hulk correctly predicted at almost 89%. Similarly, DoS GoldenEye and DoS SlowHTTPTest had normalized classification rates of ~75% and ~74% respectively. The above results show that the sequential attention mechanism is effective in identifying useful network flow features that would be able to distinguish different types of DoS attacks despite their similar behaviour.

While the overall classification performance is good, Figure 6 also shows that there are certain minority attack classes which have relatively lower normalized classification rate. The classes Heartbleed, Infiltration, Web Attack – SQL Injection, and Web Attack – Brute Force have relatively low values along the diagonal due to the small size of these classes in the CIC-IDS2017 corpus. The number of training examples is quite low in these attacks, so the classifier has limited chances of learning good decision boundaries in optimization. As a result, a percentage of these rare attack samples is expected to be classified as benign traffic or placed amongst similar types of attack.

Another interesting thing to note is the classification behaviour of DoS Slowloris, FTP-Patator and SSH-Patator. These categories are able to classify with a good accuracy but there is a certain percentage of samples that are classified as benign traffic. Certain network flows of these attacks mimic normal network flows, making them more difficult to identify by their behaviour. However, most of the samples of these attack classes are still correctly classified, showing the strength of the proposed attention-based learning framework.

The normalized confusion matrix also highlights the fact that the majority of the misclassifications are between close attack classes, not between completely distinct traffic classes. The fact that this observation occurs means the proposed model is able to isolate benign traffic from malicious network activity, while causing relatively low confusion between attacks with similar statistical properties. These sorts of activities are natural in realistic intrusion detection scenarios, and multiple cyber-attacks can create similar packet transmission patterns, flow durations and communication statistics.

Therefore, the results shown in Figure 6 are in line with the classification accuracy of 97.92%, the precision of 97.69%, the recall of 97.92% and the F1-score of 97.64% as reported earlier. The normalized confusion matrix shows that the high performance overall is not due to a single dominant benign traffic class, but is also reflected in the capability of predictions for most categories of attacks. Moreover, the study showcases the strength of the proposed TabNet model under the application of a large scale multi-class cybersecurity dataset with the presence of heterogeneous network traffic patterns.

In general, the normalized confusion matrix further confirms the reliability of the class-wise prediction performance of the proposed cyber security threat classification framework with good generalization capability. The misclassifications observed are mainly linked with the minority attack categories with few training samples, which may be improved by class balancing methods or by generating synthetic samples. However, the normalized confusion matrix illustrates that the proposed framework is always accurate and reliable in the multi-class intrusion detection, which can be applied practically in cybersecurity.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

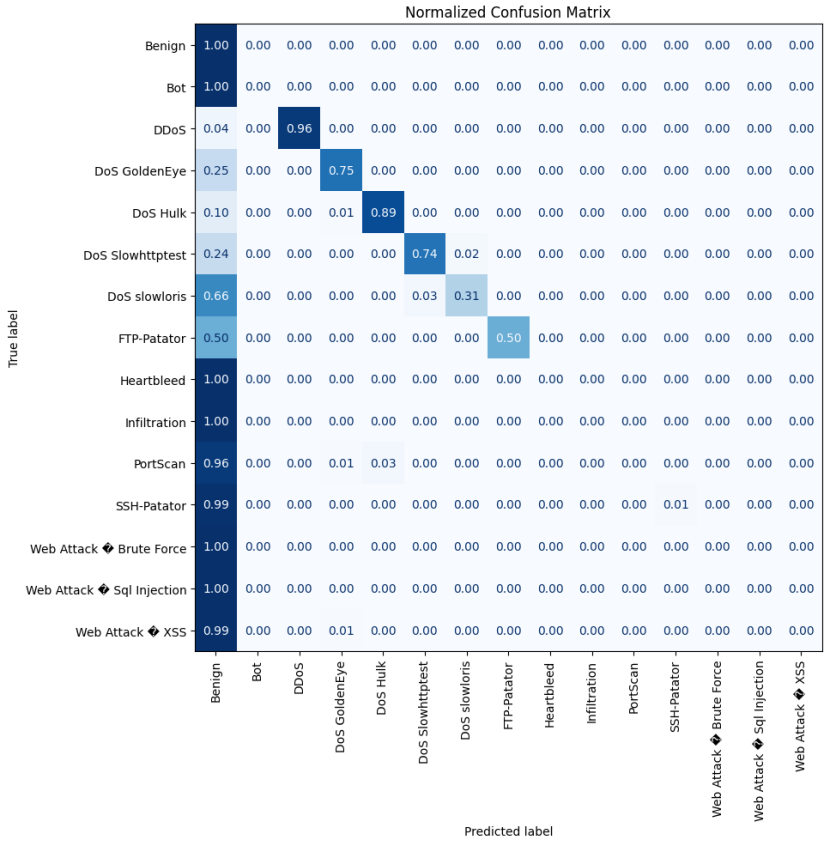

Figure 6. Normalized confusion matrix of the proposed TabNet-based cyber security threat classification framework. High diagonal values indicate accurate class-wise prediction performance for the majority of attack categories, whereas the comparatively lower values observed for rare attack classes primarily reflect the inherent class imbalance of the CIC-IDS2017 dataset rather than limitations of the proposed model.

4.5 Precision–Recall Curve Analysis

Besides the Receiver Operating Characteristic (ROC) analysis, the results of the proposed TabNet-based cyber security threat classification framework were assessed via Precision–Recall (PR) curves. In intrusion detection problems, precision–recall analysis is useful for cases where the cyber data sets are often imbalanced, meaning that some intrusion classes are much smaller than others. The ROC curve examines the tradeoff between the TPR and FPR while the Precision – Recall (PR) curve examines the tradeoff between precision and recall, and is better for minority attack classes. The Precision–Recall curves for the proposed framework are shown in Figure 7.

The Precision–Recall curves for most of the attack categories are close to the upper right corner of the graph as shown in Figure 7, showing that the classification performance is excellent at various confidence thresholds. The best classifier will have both the highest precision and the highest recall, resulting in a curve that will be in the upper right corner of the graph. The achieved results verify that the proposed TabNet model is effective in maintaining high detection capability and low false-positive rates for most of the cyber-attack categories, and that the attention-based learning mechanism is effective for multi-class intrusion detection.

The results of the experiments show that attack categories like Benign, DDoS, DoS Hulk, DoS GoldenEye, FTP-Patator, SSH-Patator, PortScan and Bot have consistently high precision and recall scores during the evaluation process. These curves are kept near the ideal operating region, demonstrating that the classifier is able to correctly classify malicious traffic and generate relatively few false alarms. The sequential attention mechanism successfully learns discriminative feature representations that can distinguish between the normal network traffic and multiple attack categories, as illustrated in the above behavior.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The results on the high precision of the majority of attack classes show that if the proposed model classifies a cyber-attack as a particular class, then the classification is very reliable and it has a low rate of false-positive classification. Likewise, the successful detection of most malicious network flows is indicated by the consistently high recall values, which means that the probability of missing an attack is low with the proposed framework. This trade-off between precision and recall is especially crucial for real-world systems that use practical IDSs because an over-alerting system and a system that never alerts can have a significant impact on network security operations.

While the overall traffic classes have very good Precision–Recall performance, certain minority attack classes showed relatively poor performance. Heartbleed, Infiltration, and Web Attack – SQL Injection classes have lower precision and recall when compared to the major attack categories. This behaviour is mainly due to the fact that such attacks are present in the CIC-IDS2017 dataset in a very small number of training samples, which leads to a significant class imbalance. So the model has less training data for the rare attack patterns to learn robust decision boundaries.

The Precision – Recall curves also show that there is an inherent robustness of the proposed framework under different decision thresholds for several attack categories, where there is no apparent drop in precision as recall rises significantly. In applications such as cyber security, this property is very useful as it allows the security administrator to set decision thresholds as per the requirements of operations without compromising on the security performance of intrusion detection. For instance, a system that needs maximum attack detection will have higher recall, while a system that has to achieve maximum precision will have a higher cost to tolerate higher false alarms.

Qualitative evaluation results for most of network traffic classes in this study show good Precision–Recall characteristics, agreeing with the overall quantitative evaluation results. The proposed model had an overall precision of 97.69%, recall of 97.92%, and F1-score of 97.64%, which shows that the proposed framework is a well-balanced classification model that can be used to accurately detect cyber threats with a minimum number of false-positive predictions. The high degree of overlap of the Precision–Recall analysis and the previously reported ROC analysis further confirms the robustness and reliability of the proposed TabNet architecture.

As a final comment with regard to Figure 7, it is noted that the Precision–Recall curves are mostly smooth for most attack categories. A lack of sudden changes of the prediction probabilities indicates that the model gives consistent classification confidence over a wide range of threshold values and stable prediction probabilities. This behaviour suggests that the learned feature representations are capable of distinguishing between benign and malicious network flows, and that they can be used to make accurate decisions even when the operating conditions are different.

The Precision–Recall analysis demonstrates that the proposed cyber security threat classification framework based on TabNet achieves good detection performance in both majority and minority classes and has a good precision–recall trade-off overall. The excellent Precision–Recall curves, along with the obtained 97.92% classification accuracy and 99.56% weighted ROC-AUC, highlight the strong performance of the presented attention-based deep learning framework in large-scale multi-class intrusion detection. The experimental results show that the proposed framework, despite the marginal degradation of its performance for rare attack categories due to class imbalance, is able to classify cyber security threats accurately, reliably and with scalability, suitable for practical use in today's network environments.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

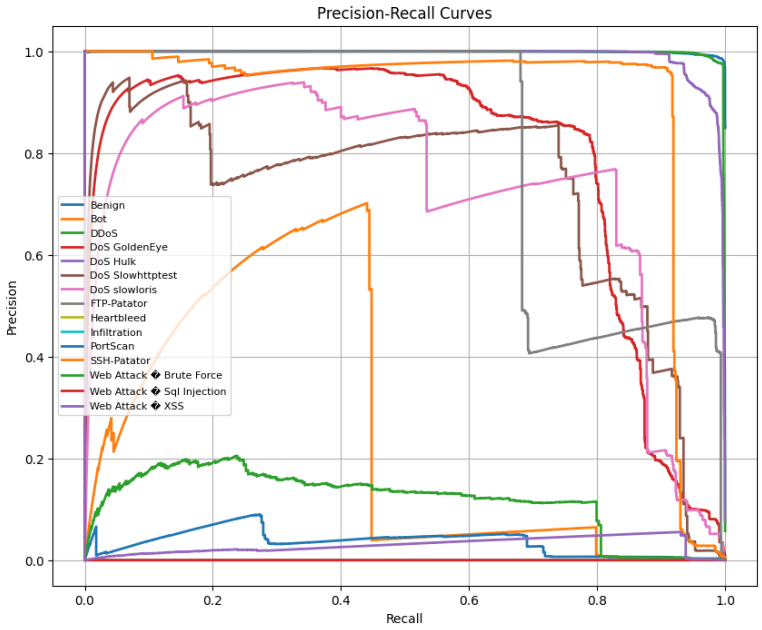

Figure 7. Precision–Recall curves of the proposed TabNet-based cyber security threat classification framework for the CIC-IDS2017 dataset. The majority of attack categories maintain high precision and recall across different decision thresholds, demonstrating the ability of the proposed model to achieve reliable intrusion detection while minimizing false-positive predictions.

4.6 Feature Importance Analysis

Interpretability, which can be achieved by feature importance analysis, is one of the key benefits of the proposed TabNet architecture. The main difference between TabNet and classic deep learning classification models is that TabNet adopts a sequential attention mechanism during model training, automatically selecting the most informative features that are part of the input. It allows to assess quantitatively the contribution of each network traffic feature to the final classification without needing to use other explainability techniques. Figure 8 shows the results of the feature importance scores calculated by the trained TabNet model, ranked by the relative importance of the network traffic features for the classification of cyber security threats.

The proposed model is based on a subset of the network traffic attributes that are relatively small but contain a high proportion of informative data, and the other attributes that have a lower proportion of informative data are assigned less importance, as illustrated in Figure 8. This behavior highlights the performance of the sparse attention mechanism introduced in TabNet architecture. The model adapts the input variables it uses at each decision step, rather than using them all the same, minimizing the amount of redundant data and computational processing. As a result, the classifier is mostly trained using the features from network flows that carry the most discriminatory information about whether the flow is benign or from a particular category of cyber-attack.

The most significant feature found from the extracted features is Destination Port (Dst Port) for classification purposes. This observation is in line with the practical knowledge of the cyber security world, as for malicious activities, many of the network attacks use specific service ports or the characteristic communication ports. The proposed model's success in identifying the meaning of the destination port suggests that the sequential attention network is capable of learning the behavioral characteristics that emerge in the protocol level of various cyber attacks.

The second most important was Flow Duration, which is the overall length of each session of network communication. A key indicator is the flow duration as the temporal behaviour of the various cyber-attacks is different. For instance, denial-of-service attacks are typically characterized by very short or repetitive flows while infiltration attacks and normal user activities show other durations of communication. The high importance score assigned to this feature indicates that the temporal attributes of network traffic are important factors to consider when differentiating between legit and malicious traffic.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

There were also several statistical features based on packets with high importance scores such as Total Forward Packets, Total Backward Packets, Total Length of Forward Packets and Total Length of Backward Packets. These variables are used to describe the volume of data transferred between communicating hosts and give insightful information about the symmetry of communication, as well as the behaviour of packets. In most cases, attack categories like DDoS, PortScan, or brute-force attacks produce unusual packet distributions from that of a normal network session. Hence, the discrimination capability of the proposed model is greatly enhanced by these packet level statistics.

The feature importance analysis also shows the contribution of timing-related features like Flow Bytes/s, Flow Packets/s, Flow IAT Mean, Flow IAT Standard Deviation, Packet Inter-Arrival Time. These characteristics describe how fast packets are sent in each network flow and the distribution of the packets across the flow's duration. Many intrusion scenarios will create unusual communication rates, and unusual timing behaviour, therefore these features are important in the identification of cyber threats. The importance of these variables shows that the proposed TabNet model is successful in making the best use of the characteristics of the temporal network in feature learning.

Forward Packet Length Mean, Backward Packet Length Mean, Average Packet Size, Packet Length Variance and Packet Length Standard Deviation were also found to be important for classification. These features represent statistical differences in packet sizes and communication patterns, allowing the classifier to differentiate patterns of attacks with similar packet counts, but different communication patterns. Being included among the most influential variables suggests that they adequately capture important complementary aspects of traffic when building decision boundaries.

Note that after the top-scoring variables in the list, the importance scores for the features in the network traffic set taper off, suggesting that few of the original 77 network traffic features have a significant impact on the final prediction. This behaviour shows the success of TabNet's feature selection strategy. The proposed model focuses mostly on the most informative variables to avoid using redundant and weakly informative variables, thereby improving the accuracy of the classification model. This ability to automatically select features is an especially big benefit over conventional deep learning models that generally treat all input features equally throughout training.

The feature importance analysis also gives insights about the proposed intrusion detection framework's decision making process. Security analysts can leverage the features that have been identified as high impact to gain a better understanding of what activities in networks are most likely to be malicious. The transparency of the proposed model also fosters the practical application of the model, as explainable AI (XAI) is crucial in cybersecurity applications where comprehending the rationale behind automated decision-making is as vital as attaining high prediction accuracy.

The results of the feature importance analysis show that the proposed TabNet architecture is able to correctly recognize meaningful network traffic features that are able to contribute to the correct classification of cyber threats. The attention-based feature selection mechanism not only reduces computation cost, but also gives quantitative evidence of the relative importance of each network flow attribute, which adds to the interpretability of the model. The feature importance results also support the success and usability of the proposed framework in detecting intelligent intrusion in modern network environments by achieving 97.92% classification accuracy, 97.69% precision, 97.92% recall, 97.64% F1-score, and 99.56% weighted ROC-AUC.

Journal of Information Systems Engineering and Management
2023, 8(1)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

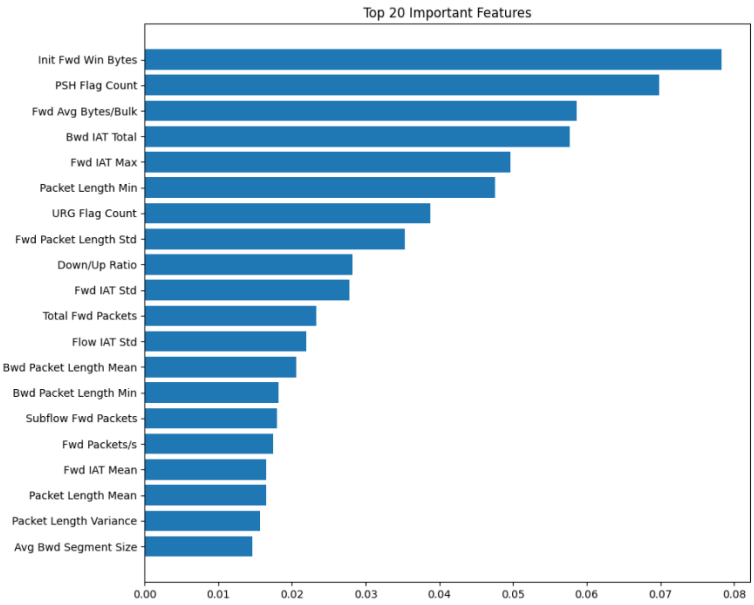

Figure 8. Top 20 feature importance scores generated by the proposed TabNet model. The ranking indicates that destination port information, flow duration, packet statistics, transmission rates, and timing-related features contribute most significantly to cyber security threat classification. The sparse attention mechanism automatically emphasizes these discriminative attributes, improving both classification accuracy and model interpretability.

5. CONCLUSION

This study introduced a TabNet Deep learning framework for multi-class cyber security threat classification using the CIC-IDS2017 benchmark intrusion detection dataset. To enhance the accuracy, efficiency and interpretability of intrusion detection, the proposed framework was developed to incorporate a complete data preprocessing, a feature standardization, an attention-based feature learning, and a performance evaluation module in a single classification pipeline. The proposed TabNet architecture automatically learns the most informative network traffic features, leveraging the sequential attention mechanism, which enables the model to learn from high-dimensional network flow data without the need for complex manual feature engineering while preserving the transparency of the model.

The experimental assessment revealed that the proposed scheme was able to obtain the best classification accuracy on the CIC-IDS2017 dataset. The model achieved an overall classification accuracy of 97.92%, precision of 97.69%, recall of 97.92% and F1-score of 97.64%, proving its accuracy to distinguish the benign traffic from the various types of cyber-attacks. Moreover, the model presented a weighted ROC-AUC of 99.56%, which demonstrates the outstanding discrimination ability of the model for various network intrusion classes. The analysis of the training and validation curve showed good convergence during the optimization process, and the confusion matrix and normalized confusion matrix showed good prediction performance per class for most of the attack categories. The proposed framework consistently demonstrated a high predictive performance in diverse network traffic scenarios, though a few minority attack classes had poorer performance due to the inherent class imbalance in the dataset.

The interpretability offered by the TabNet architecture is another contribution of this work. The feature importance analysis showed that the attributes that had the most significant impact in the classification of cyber threats were the Destination Port, Flow Duration, packet statistics, and flow-based communication characteristics. The identification of the features of network traffic that are influential helps to increase transparency of the proposed intrusion detection system and to make the automated classification decisions explainable in terms of artificial intelligence (XAI), which enables a security analyst to better understand the decision-making process. The proposed model also showed efficient computational performance by deducing the model in the test data set with an inference time of about 5.50 seconds for the whole test data set, which means the average inference time of one sample is about 1.23×10^{-5} seconds, and is suitable for large-scale network traffic analysis.Overall, the experimental results validate that the developed TabNet based framework is a valid, scalable and interpretable approach to intelligent cyber security

threat classification. The proposed approach offers great promise for enterprise intrusion detection systems, cloud computing infrastructure, and next-generation cybersecurity platforms, owing to the effectiveness of its attention-based feature selection, efficient optimization, and full-scale evaluation.

There are several future opportunities for research, however despite the promising results. The present work conducted an evaluation of the proposed framework on single benchmark datasets; future work could further evaluate the model generalization by testing it with other intrusion detection datasets like CSE-CIC-IDS2018, UNSW-NB15, CIC-DDoS2019 or recent IoT security datasets. In future, other methods of dealing with class imbalance can be explored, such as the generation of synthetic data, the use of focal loss, or cost sensitive learning, to enhance the detection of rare attack categories. In addition, combining TabNet with ensemble learning or transformer-based architectures or federated learning systems could further improve the predictive accuracy and privacy protection in distributed cybersecurity settings. The above research directions could further enhance the capabilities of IDSs as well as help to design a strong explainable, and real time cyber defense solution for modern network infrastructures.

6. REFERENCES

[1] Udayakumar, R., Joshi, A., Boomiga, S. S., & Sugumar, R. (2023). Deep fraud Net: A deep learning approach for cyber security and financial fraud detection and classification. *Journal of Internet Services and Information Security*, *13*(3), 138-157.

[2] Khan, I. A., Moustafa, N., Pi, D., Sallam, K. M., Zomaya, A. Y., & Li, B. (2021). A new explainable deep learning framework for cyber threat discovery in industrial IoT networks. *IEEE Internet of Things Journal*, *9*(13), 11604-11613.

[3] Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., & Xu, M. (2020). A survey on machine learning techniques for cyber security in the last decade. *IEEE access*, *8*, 222310-222354.

[4] Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. *Journal of Information Security and Applications*, *50*, 102419.

[5] Yu, Z., Wang, J., Tang, B., & Lu, L. (2023). Tactics and techniques classification in cyber threat intelligence. *The Computer Journal*, *66*(8), 1870-1881.

[6] Berman, D. S., Buczak, A. L., Chavis, J. S., & Corbett, C. L. (2019). A survey of deep learning methods for cyber security. *Information*, *10*(4), 122.

[7] Geetha, R., & Thilagam, T. (2021). A Review on the Effectiveness of Machine Learning and Deep Learning Algorithms for Cyber Security: R. Geetha, T. Thilagam. *Archives of Computational Methods in Engineering*, *28*(4), 2861-2879.

[8] Al-Abassi, A., Karimipour, H., Dehghantanha, A., & Parizi, R. M. (2020). An ensemble deep learning-based cyber-attack detection in industrial control system. *Ieee Access*, *8*, 83965-83973.

[9] Ullah, F., Naeem, H., Jabbar, S., Khalid, S., Latif, M. A., Al-Turjman, F., & Mostarda, L. (2019). Cyber security threats detection in internet of things using deep learning approach. *IEEE access*, *7*, 124379-124389.

[10] Noor, U., Anwar, Z., Amjad, T., & Choo, K. K. R. (2019). A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise. *Future Generation Computer Systems*, *96*, 227-242.

[11] Wazid, M., Das, A. K., Chamola, V., & Park, Y. (2022). Uniting cyber security and machine learning: Advantages, challenges and future research. *ICT express*, *8*(3), 313-321.

[12] Mishra, S., Albarakati, A., & Sharma, S. K. (2022). Cyber threat intelligence for IoT using machine learning. *Processes*, *10*(12), 2673.

[13] Ch, R., Gadekallu, T. R., Abidi, M. H., & Al-Ahmari, A. (2020). Computational system to classify cyber crime offenses using machine learning. *Sustainability*, *12*(10), 4087.

[14] Li, J. H. (2018). Cyber security meets artificial intelligence: a survey. *Frontiers of Information Technology & Electronic Engineering*, *19*(12), 1462-1474.

[15] Ahsan, M., Nygard, K. E., Gomes, R., Chowdhury, M. M., Rifat, N., & Connolly, J. F. (2022). Cybersecurity threats and their mitigation approaches using Machine Learning—A Review. *Journal of Cybersecurity and Privacy*, *2*(3), 527-555.

[16] Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A., & Elsoud, E. A. (2022). An intelligent cyber security phishing detection system using deep learning techniques. *Cluster Computing*, *25*(6), 3819-3828.

[17] Alghamdi, M. I. (2020). Survey on Applications of Deep Learning and Machine Learning Techniques for Cyber Security. *International Journal of Interactive Mobile Technologies*, *14*(16).

[18] Ebrahimi, M., Nunamaker Jr, J. F., & Chen, H. (2020). Semi-supervised cyber threat identification in dark net markets: A transductive and deep learning approach. *Journal of Management Information Systems*, *37*(3), 694-722.

[19] Abu Al-Haija, Q., & Zein-Sabatto, S. (2020). An efficient deep-learning-based detection and classification system for cyber-attacks in IoT communication networks. *Electronics*, *9*(12), 2152.

[20] Sarker, I. H., Abushark, Y. B., Alsolami, F., & Khan, A. I. (2020). Intrudtree: a machine learning based cyber security intrusion detection model. *Symmetry*, *12*(5), 754.

[21] Diaba, S. Y., Shafie-Khah, M., & Elmusrati, M. (2023). Cyber security in power systems using meta-heuristic and deep learning algorithms. *IEEe Access*, *11*, 18660-18672.