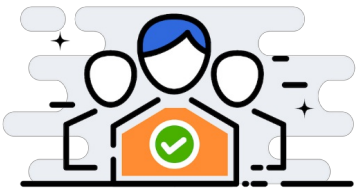
Table of Contents



Abbreviation

EL	Ensemble Learning
IDS	Intrusion Detection System
DP	Data Processing
DM	Data Mining
NIDS	Network Intrusion detection systems

Base Paper Title

Intrusion detection technique in wireless sensor network using grid search random forest with Boruta feature selection algorithm

Modified Title

Concept Analysis and Dynamic Projection Intrusion Cyberattack Detection using Ensemble Learning

Modified Title Explanation

Concept Analysis : Clarify ambiguous concepts in a theory, and to propose a precise operational definition.

Dynamic Projection : Guarantees highly accurate Detection.

Ensemble Learning : Ensemble learning Algorithms helps improve machine learning results by combining several models.

Abstract

Our increasingly connected world continues to face an ever-growing amount of network-based attacks. Intrusion detection systems (IDS) are an essential security technology for detecting these attacks. Although numerous machine learningbased IDS have been proposed for the detection of malicious network traffic the majority have difficulty properly detecting and classifying the more uncommon attack types.

The research in the field of Cyber Security has raised the need to address the issue of cybercrimes that have

caused the req- uisition of the intellectual properties such as break down of computer systems impairment of important data compromising the confidentiality authenticity and integrity of the user. Considering these scenarios it is essential to secure the computer systems and the user using an Intrusion Detection System (IDS). The performance of IDS studied by developing an IDS dataset consisting of network traffic features to learn the attack patterns. Intrusion detection is a classification problem wherein various Ensemble Learning (ML) and Data Mining (DM) techniques applied to classify the network data into normal and attack traffic. Moreover the types of network attacks changed over the years and therefore there is a need to update the datasets used for evaluating IDS.

Introduction

The increasing volume and sophistication of networkbased attacks motivate the development of effective techniques and tools to prevent service disruption unauthorized access and the disclosure of sensitive information . An Intrusion Detection System (IDS) is an important defence tool against sophisticated and increasing network attacks but these systems especially Machine Learning (ML) based systems require large reliable and valid network traffic datasets to be effective. Although the majority of recently available datasets cover a range of network attack types and traffic patterns and include information about the attacking infrastructure modern networks are increasingly diversified such that existing datasets are often not enough to develop effective classification mechanisms. These datasets often suffer from a lack of traffic diversity and volume or fail to cover the full scope of known attack types. To cope up with these new changes we require a more dynamic dataset that will improve the ability of an IDS to detect intrusions. Using deep learning techniques such as Generative Adversarial Networks (GANs) we can fabricate additional data using existing datasets to increase the classification accuracy of an IDS especially for rare attack categories.

Two methods of IDS are Signature-based Intrusion Detection Systems (SNIDS) and Anomaly-based Intrusion Detection Systems (ANIDS). The SNIDS approach is effective for known threats as it looks for specific patterns (or signatures) such as byte sequences in network traffic or known malicious instructions sequences used by malware . Conversely the ANIDS approach uses ML algorithms to analyze and monitor the network traffic in order to detect any suspicious activity thus being an effective method for catching unknown attacks .

The emergence of deep learning and its integration with Reinforcement Learning (RL) has created a class of Deep Reinforcement Learning (DRL) methods that are able to detect the most recent and sophisticated types of network attacks. DRL combines artificial neural networks with a framework of RL that helps software agents (or learning entities) learn how to reach their goals. DRL combines function approximation and target optimization mapping states and actions to the rewards they lead to . This results in a policy that our learning agents can follow to make the best decisions given the current state. To detect network attacks DRL is used to train an agent such that given a state represented as a collection of feature values will take the best action (which in our case acts as a classification of attack type) in order to recognize an attack.

Each network is different in that its behaviors and patterns evolve gradually. Naturally vulnerabilities also evolve. The performance of IDS classification accuracy suffers as existing datasets gradually become out of date invalid and unreliable. Moreover reliable data cannot often be shared due to privacy concerns. Existing publicly available datasets do not include all of the existing network attack types let alone the unknown vulnerabilities and attacks.

To resolve this we need more diverse and up-to-date datasets that properly reflect the characteristics of network intrusions in order to increase the performance of the IDS.

According to the statistics reported for Cyber Security the damages caused by the cyber attacks are expected to reach up to three trillion by with the probability of executing zero-day exploits one-per-day . Moreover the amount of the information stored in private as well as public clouds operated by data-driven companies such as Amazon Web Services Facebook and Twitter will be increased hundred times by . Thus an increase in the data demand for more proficient security systems. The computer systems with loopholes security mechanisms with incompetent security policies and lack of knowledge about the attacks and crimes have increased the targets for zero-day attacks are difficult to trace using the standard security mechanisms such as firewall and anti-virus software . Therefore an Intrusion Detection System (IDS) used to examine the information flowing through the network and to generate an alarm for the probable malicious activities generated by the intruders. An IDS detects the intrusions either by extracting the signatures from the network packets or by analyzing the attacks patterns. An IDS that detects the intrusion by studying the signatures is termed as Signature-based IDS. The Signature-based IDS generates an alert for the matched signature patterns stored in the signature database. In contrast an IDS detecting attacks based on the attack patterns are referred to as Anomaly-based IDS. A comparison table of the different IDS is presented in Table . Regardless of the type of IDS the basic architecture of IDS consists of four steps. The network packets are captured using network sensors or network sniffing tools. The captured data is then filtered and examined. The filtering is performed based on filtering rules and then signature patterns are matched with the already available signature database. An alert is generated by the IDS when a match is found with the stored signature database.

Network security has become one of the most concerning problems for internet users and service providers with drastic increase in the internet usage . A secure network can be defined in terms of its hardware and software protection against various intrusions. A network can be secured by implementing a resilient monitoring analysis and defense mechanisms. Network Intrusion detection systems (NIDS) forms a class of systems which implement these mechanisms in order to defend a network from insider and outsider intrusions. These systems monitor the incoming and outgoing traffic in a network perform time to time analysis and report when some intrusion is detected. NIDS can be broadly categorized into Misuse detection (MD) Anomaly Detection (AD). MD based NIDS use signatures or patterns of already existing attacks to detect intrusions. While AD based NIDS check for strict deviations from the normal profiles of the network traffic and report it as attack.

Objectives

- Autolearning approach to improve its detection capabilities for different types of network intrusions.
- Features extracted from the sequential patterns of the data to detect intrusions.
- Tagging the data instances captured from the network traffic to have a complete understanding of the network interaction.
- Should include all the communication using different protocols whether normal or malicious.
- Should maintain a complete set of well-defined features for classifying the attack

Problem Statement

Complete capture of the network traffic such as communication between host broadcast message domain lookup query the protocol being used.

Existing System

Attacks in wireless sensor networks (WSNs) aim to prevent or eradicate the networks ability to perform its anticipated functions. Intrusion detection is a defense used in wireless sensor networks that can detect unknown attacks. Due to the incredible development in computer-related applications and massive Internet usage it is indispensable to provide host and network security. The development of hacking technology tries to compromise computer security through intrusion. Intrusion detection system (IDS) was employed with the help of machine learning (ML) Algorithms to detect intrusions in the network. Classic ML algorithms like support vector machine (SVM) Knearest neighbour (KNN) and filter-based feature selection often led to poor accuracy and misclassification of intrusions. This article proposes a novel framework for IDS that can be enabled by Boruta feature selection with grid search random forest (BFSGSRF) algorithm to overcome these issues. The performance of BFS-GSRF is compared with ML algorithms like linear discriminant analysis (LDA) and classification and regression tree (CART) etc. The proposed work was implemented and tested on network security laboratory knowledge on discovery dataset (NSL-KDD). The experimental results show that the proposed model BFS-GSRF yields higher accuracy (i.e. %) in detecting attacks and it is superior to LDA CART and other existing algorithms.

For classification random forest with grid search (RFGS) is used. RF is one of the widely used algorithms for classification problems. This algorithm will create several classification trees for predicting the target class. Based on the majority of the vote the final prediction was made. Parameter optimization is used to improve the accuracy of the RF algorithm. The grid search method is used in RF to obtain the classification model with higher accuracy for tuning the parameter. The randomly based search method is more efficient than the gridbased search method for hyperparameter optimization. Two discrete integer parameters such as ntree and mtry are used to tune the parameter. The main objective of the optimization is to minimize the out of bag (OOB) error. After multiple runs optimal parameters value is chosen based on the pair that produces the lowest OOB error.

CART classification is a supervised nonlinear algorithm used for classification and regression. This algorithm constructs the binary decision tree by splitting the attributes which is considered as a node. The whole tree from root to leaf contains a learning sample. For classification the target variable in CART should be categorical whereas the target variable for the regression tree should be continuous. Here the target variable is categorical for performing classification on intrusion detection. The metric Gini index is used to perform the classification task.

Drawbacks of Existing System

- Correctly classified records can not outperform.
- No statistically relevant performing variations among the various algorithms.
- Poor Application Performance
- High complexity, inaccuracy, and inadequacy
- Cannot meet current network business demands
- Computation burden may limit its further application for real scenarios.

Dataset Desc

Dataset URL: https://www.unb.ca/cic/datasets/ids-.html

Dataset Description: Anomaly detection has been the main focus of many researchers due to its potential in detecting novel attacks. However its adoption to real-world applications has been hampered due to system complexity as these systems require a substantial amount of testing evaluation and tuning prior to deployment. Running these systems over real labeled network traces with a comprehensive and extensive set of intrusions and abnormal behavior is the most idealistic methodology for testing and evaluation.

This itself is a significant challenge since the availability of datasets is extremely rare because from one side many such datasets are internal and cannot be shared due to privacy issues and on the other hand the others are heavily anonymized and do not reflect current trends or they lack certain statistical characteristics so a perfect dataset is yet to exist. Thus researchers must resort to datasets that are often suboptimal. As network behaviours and patterns change and intrusions evolve it has very much become necessary to move away from static and one-time datasets towards more dynamically generated datasets which not only reflect the traffic compositions and intrusions of that time but are also modifiable extensible and reproducible.

Proposed System

An intrusion detection dataset can be developed by collecting information from varied sources such as network traffic flows that contains information about the host user behavior and system configurations. This information is required to study the attack patterns and abnormal activity of various network attacks. The network activity is collected through a router or network switch. After collecting the incoming and the outgoing network traffic network flow analysis is performed to study the network traffic. Flow analysis can be described as the process of analyzing the network packet information such as source IP address destination IP address source port number destination port number type of network services to name a few . The network host delivers the system configurations and user information that cannot be extracted from the network flow analysis. For instance information obtained through failed login attempts by observing the intrusion activity.

The evaluation of an IDS model can be performed by implementing Machine Learning (ML) and Data Mining (DM) techniques to classify the network traffic into benign and malicious traffic flow. The ML and DM techniques

implemented on the IDS datasets contains labeled data and network traffic features. These help the classifier to learn different attack patterns to detect a particular attack. The features of the dataset help the classifier to learn the normal traffic patterns as well as attack patterns through which the classifier is able to classify the input data . The dataset used for training the classifier is built by monitoring the network traffic for a particular interval of time. The dataset consists of normal network traffic and anomalous network traffic that helps the classifier to identify the patterns of the data with a sufficient amount of examples. The data collected is divided into a training set and test set for training and testing the classifier respectively. Thus various ML and DM techniques used for developing an IDS.

A single performance metric is not sufficient to measure the efficiency of the algorithm. It is necessary to consider confusion matrix and find the number of false positives and false negatives to derive other performance metrics such as Detection Rate (DR) False Positive Rate (FPR) precision and recall . The accuracy of a particular attack type is also a critical aspect as the classifier may give better accuracy for one attack type but may fail for classifying the other

Advantages of Proposed System

- Implemented parameter optimization, feature optimization, and variability in the size of the dataset.
- Increasing the precision and recall performance.
- Bridges gap and improves the current state of knowledge in field.
- Quick and Efficient to use
- problems are solved on an end-to-end basis
- Ability To Deliver High Quality Results

Hardware & Software Requirements

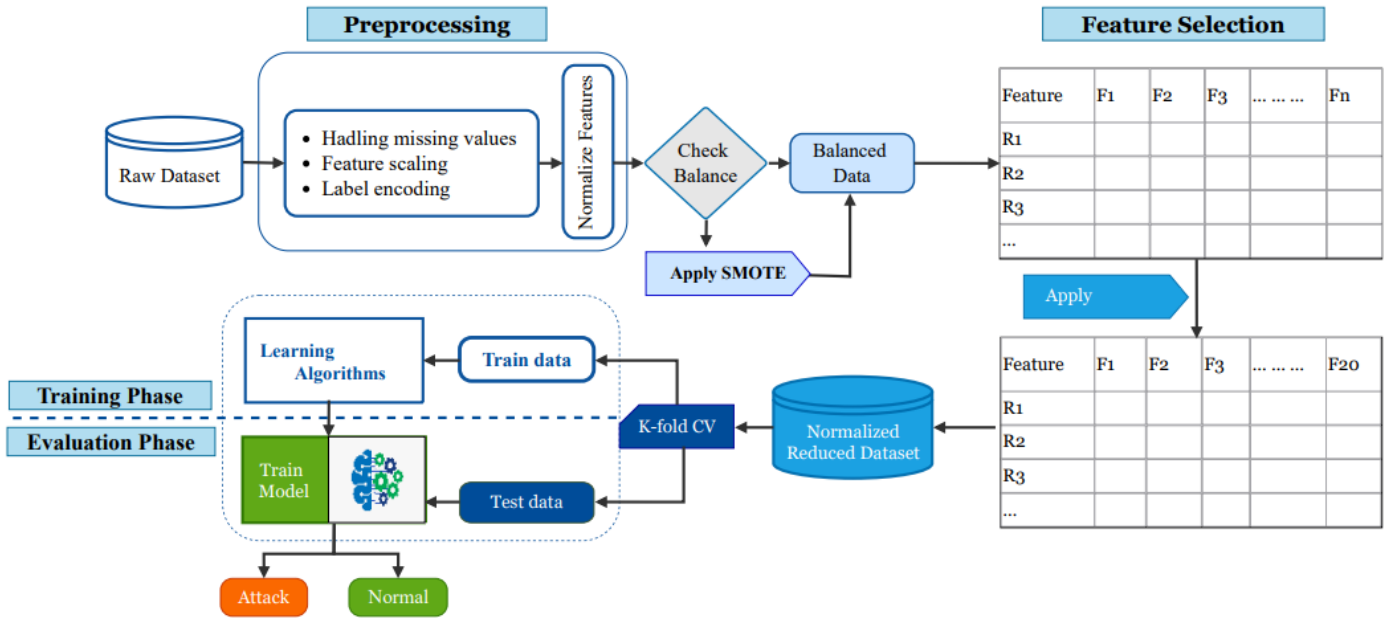
Hardware Requirements

- Processor: Minimum i3 Dual Core
- Ethernet connection (LAN) OR a wireless adapter (Wi-Fi)
- Hard Drive: Minimum 100 GB; Recommended 200 GB or more
- Memory (RAM): Minimum 8 GB; Recommended 32 GB or above

Software Requirements

- Python
- Anaconda
- Jupyter Notebook
- TensorFlow
- Keras

Architecture



Existing Algorithm

- Grid Search Random Forest

Proposed Algorithm

- Ensemble Learning Algorithm

Advantages of Proposed Algorithm

Ensemble Learning Algorithm Advantages Advantages

- Handling multi-dimensional and multi-variety data
- Efficient Handling of Data
- Continuous Improvement

Project Modules

Module 1 : Preprocessing Phase

Different types of feature preprocessing are required for different data types and different machine learning models. Some preprocessing methods are common for all datatypes. Feature scaling is a method used to normalize the range of independent variables or features of data. It is commonly referred to as normalization. Feature scaling impacts non-tree-based models more than tree-based models. Thus, if you want to achieve good results using a non-tree-based model, you should consider normalizing your numerical features.

Module 2 : Learner Based Feature Selection

Feature selection is a method of selecting a subset of the underlying features in order to minimize the feature space to the smallest possible size based on some criteria. Feature extraction is a technique for creating a new set of features that can be utilized alone or in combination. Moreover, it can locate and choose the far more beneficial properties inside data. Its an important stage in the learning workflow since it assists in minimizing the fitting problems, reducing adaptation efficiency on the testing data, reducing training duration, and reducing model interpretability. There are three main kinds of feature selection methods: filter-based, wrapper-based, and embedded feature selection. Build-in feature selection is available in the embedded feature selection method, which helps to build a model without applying any additional feature selection method. To choose features, the filterbased feature technique employs assessment criteria, including information analysis as well as distance assessment. The wrapper-based feature selection approach builds a subset of features in a particular way before evaluating feature selection using the findings of classifiers. Using the embedded feature selection approach, certain properties can be dynamically removed within classifier construction, allowing feature selection and classification to be done simultaneously.

The variable selection, also known as attribute selection or feature selection is the process of choosing the most important features of a given dataset. In a network intrusion detection dataset, there might be several features that do not contribute to the detection of intrusion. So in order to reduce overfitting, improve the accuracy of the model, and reduce the training time, we can carry out feature selection before training the model. In this paper, we use an analyzer function to evaluate the performance of algorithms with different subsets of the dataset to find the best one that results in better accuracy.

Module 3 : Model Training and Evaluation

Hyperparameter tuning involves finding the best set of parameters to give to our algorithm to achieve the best accuracy measures. Generally, there are two techniques that are used for this purpose, namely grid search and random search. In the grid search method, every possible list of values with every combination is evaluated, and in the random search method, random combinations of parameters are tested to find the best possible values for the model. In this work, we utilized the random search method to find the best hyperparameters in single decision tree experiments. In a Random Forest algorithm, parameters such as the number of trees and the depth of the tree can be examined

Evaluation Metrics

We used the accuracy and F1-score (which combines precision and recall) metrics to evaluate the performance our DRL model and other ML algorithms. While the accuracy score only measures the percentage of correctly classified

samples, this selection of performance metrics allows us to also evaluate the percentage of samples that were incorrectly classified. This is especially important for NIDS as the accuracy performance metric is not enough to evaluate imbalanced datasets such as network traffic data which generally include significantly more normal traffic. These performance metrics are derived from the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. Fig. 1 presents this confusion matrix used by our evaluation method.

Accuracy: Accuracy measures the number of correct predictions out of the total predictions made by the model. In this case, accuracy measures the models ability to correctly identify normal and attack traffic records.

Precision: Precision measures the number of correct positive predictions out of the total number of positive predictions. In this case, precision measures the models degree of correctness in predicting attack records over the total number of attacks predicted.

Recall: Recall measures the number of correct positive predictions out of the total number of positive instances in the dataset. In this case, recall measures the models ability to correctly identify attack traffic records. From this definition, recall is also referred to as the true positive rate, detection rate, or sensitivity.

F1-Score: F1-score is the harmonic mean of the precision and recall values, essentially a combined measure of the two performance metrics. F1-score quantifies how discriminative the model and acts as a good indicator of performance since a decrease in either precision or recall results in a significant decrease in the F1-score. In addition, for multiclass classification we present both the unweighted and weighted F1-scores. The weighted F1-score accounts for label imbalance by considering the number of instances of each label when calculating the average F1-score.

Literature Survey

Literature Survey 1	
Title	Wireless Network Intrusion Detection Based on Improved Convolutional Neural Network
Authors	Hongyu Yang and Fengyan Wang
Published Year	2019
Efficiency	Easy scalability Simple to understand and interpret Simplicity and Explainability.
Drawbacks	Additional configuration is required High complexity of installing and maintaining It is not an easy-to-use method
Description	The diversification of wireless network traffic attack characteristics has led to the problems what traditional intrusion detection technology with high false positive rate, low detection efficiency, and poor generalization ability. In order to enhance the security and improve the detection ability of malicious intrusion behavior in a wireless network, this paper proposes a wireless network intrusion detection method based on improved convolutional neural network (ICNN). First, the network traffic data is characterized and preprocessed, then modeled the network intrusion traffic data by ICNN. The low-level intrusion traffic data is abstractly represented as advanced features by CNN, which extracted autonomously the sample features, and optimizing network parameters by stochastic gradient descent algorithm to converge the model. Finally, we conducted a sample test to detect the intrusion behavior of the network. The simulation results show that the method proposed in our paper has higher detection accuracy and true positive rate together with a lower false positive rate. The test results on the test set KDDTest + in our paper show that compared with the traditional models, the detection accuracy is 8.82% and 0.51% higher than that of LeNet-5 and DBN, respectively, and the recall rate is 4.24% and 1.16% higher than that of LeNet-5 and RNN, respectively, while the false positive rate is lower than the other three types of models. It also has a big advantage compared to the IDABCNN and NIDMBCNN methods. Aiming at problem of wireless network intrusion detection technology based on deep learning method that has low detection efciency and is prone to face over-tting and gen- eralization issues in the model training process. This paper proposes a wireless network intrusion detection based on improved convolutional neural network. The classication training and test experiments are carried out in IBWNIDM using the pre-processed training set and test set data. The experimental results show that the accuracy and true positive rate of intrusion detection of IBWNIDM are higher and the false positive rate is lower. Overall, IBWNIDM leverages the advantage of the deep learning model for feature extraction of sample data.

Literature Survey 2	
Title	Generalized Intrusion Detection Mechanism for Empowered Intruders in Wireless Sensor Networks
Authors	Wenming Wang , Haiping Huang , Qi Li , Fan He and Chao Sha
Published Year	2020
Efficiency	Relatively simple and computationally inexpensive method It is a fast and easy procedure to perform It provides easy information processing and cost reduction as well.
Drawbacks	Unsuitable for large scale scenarios. Maximizes the complexity of the problem High level of communication and computation overheads
Description	Intrusion detection as one of the most important approaches to guarantee wireless sensing network security has been studied adequately in previous work. However, with the development of electronic anti-reconnaissance technology, the intruder may obtain the location information of detection nodes and perform path planning to avoid being detected. Such intruder is defined as an empowered intruder who will bring new challenges for traditional intrusion detection methods. Moreover, some subareas may have coverage holes due to random initial deployment of detection nodes, the desired effect of detection cannot be achieved. To address these issues, we propose a vehicle collaboration sensing network model, where mobile sensing vehicles and static sensor nodes cooperate to provide intrusion detection against empowered intruders. Our proposal (named as IDEI) consists of a target pursuit algorithm of mobile sensing vehicles and a sleep-scheduling strategy of static nodes. Mobile sensing vehicles will track the empowered intruder and fill up the coverage breaches, while static nodes follow a sleep-scheduling mechanism and will be awakened by detection nodes nearby when the intruder is detected. Simulation experiments are conducted to compare our proposal with existing methods such as KMsn and MTTA in terms of intrusion detection performance, energy consumption and moving distance of sensor nodes. The parameter sensitivity of IDEI is also studied with extensive simulations. The theoretical analysis and simulation results indicate that our proposal can achieve better efficiency and availability. In this paper, we rst put forward the model of empow- ered intruder. Compared with naive intruders, the empowered intruder can locate detection nodes nearby and escape from them to reduce the probability of being detected. Aiming at the challenge brought by the empowered intruder, a dis- tributed intrusion detection scheme IDEI based on vehicle collaboration sensing network is proposed. Mobile sens- ing vehicles are utilized to track the empowered intruder to achieve high-quality monitoring and a sleep-scheduling mechanism is designed for static sensors to reduce energy consumption. In addition, there is a mobile sensing vehicle that acts as an edge computing node in each monitoring area to satisfy the requirements of low latency and high-quality service.

Literature Survey 3	
Title	A Linear Systems Perspective on Intrusion Detection for Routing in Reconfigurable Wireless Networks
Authors	Jaime Zuniga-Mejia , Rafaela Villalpando-Hernandez , Cesar Vargas-Rosales and Andreas Spanias
Published Year	2019
Efficiency	Enhance correlation strength with finer and more compact information Simple to understand and interpret Performs better on various circumstances and environment
Drawbacks	High level of communication and computation overheads Difficult and Less Commonly used Cannot meet current network business demands
Description	Reconfigurable wireless networks, such as ad hoc or wireless sensor networks, do not rely on fixed infrastructure. Nodes must cooperate in the multi-hop routing process. This dynamic and open nature make reconfigurable networks vulnerable to routing attacks that could degrade significantly network performance. Intrusion detection systems consist of a set of techniques designed to identify hostile behavior. In this paper, there are several approaches for intrusion detection in reconfigurable network routing such as collaborative, statistical, or machine learning-based techniques. In this paper, we introduce a new approach to intrusion detection for reconfigurable network routing based on linear systems theory. Using this approach, we can discriminate routing attacks by considering the systems z-plane poles. The z-plane can be thought of as a two dimensional feature space that arises naturally. It is independent of the number of network attack detection metrics and does not require extra dimensionality reduction. Two different host-based intrusion detection techniques, inspired by this new linear systems perspective, are presented and analyzed through a case study. The case study considers the effects of attack severity and node mobility to the attack detection performance. High attack detection accuracy was obtained without increasing packet overhead for both techniques by analyzing locally available information. In this work, we proposed two different IDS for routing in RWN based on the same perspective of considering a network node as a linear system. This new perspective allows us to gain some intuitive understanding of the problem. Addi- tionally, by using the system poles on the z-plane as the feature space for attack detection, we can represent all the relevant information in two dimensions. This two dimen- sional feature space is guaranteed to be independent of the number of input and output signals considered as relevant network metrics for a given attack detection. Good detection accuracy was obtained for both attack detection techniques. For more elaborate scenarios than the simple case presented in Section IV we need to consider additional inputs. The root locus approach is more robust to mobility and has a lower computational cost when compared to the black box method and hence more feasible for low power devices.

Literature Survey 4	
Title	Real-Time Intrusion Detection in Wireless Network: A Deep Learning-Based Intelligent Mechanism
Authors	Liqun Yang , Jianqiang Li , Liang Yin , Zhonghao Sun , Yufei Zhao and Zhoujun Li
Published Year	2020
Efficiency	Can improve the worst-case performance Corresponding time cost is greatly reduced. Boost the Performance
Drawbacks	Prone to Errors High complexity of installing and maintaining Heavyweight
Description	With the development of the wireless network techniques, the number of cyber-attack increases significantly, which has seriously threat the security of Wireless Local Area Network (WLAN). The traditional intrusion detection technology is a prevalent area of study for numerous years, but it may not have a good detection performance in a real-time way. Therefore, it is urgent to design a detection mechanism to detect the attacks timely. In this paper, we exploit a CDBN (Conditional Deep Belief Network)-based intrusion detection mechanism to recognize the attack features and perform the wireless network intrusion detection in real time. To avoid the impact of the imbalanced dataset and the data redundancy on the detection accuracy, a window-based instance selection algorithm SamSelect is adopted to undersample the majority class data samples, and a Stacked Contractive Auto-Encoder (SCAE) algorithm is proposed to reduce the dimension of the data samples. By doing so, our proposed mechanism can effectively detect the potential attack and achieve high accuracy. The experiment results show that CDBN can be effectively combined with SamSelect and SCAE, and the proposed mechanism has a high detection speed and accuracy, with the average detection time 1.14 ms and the detection accuracy 0.974. In this paper, we propose an improved Deep Belief Network based scheme for detecting wireless network intrusion. Our proposed scheme employs Conditional Deep Belief Network (CDBN) to efciently learn the temporal behav- ior features between the experimental data. We adopt a window-based under-sampling algorithm SamSelect to balance the numbers of the normal samples and that of the attack samples in the AWID training dataset. We use Stacked Contractive Auto-encoder (SCAE) algorithm to eliminate the redundancy of experimental data. In the simulations, we illustrate our work by four cases, and the rst two cases show that SCAE is feasible to reduce the dimensionality with the average reconstruction error 0.058. The detection accuracy increases as the number of the hidden layers of CDBN increases, and the highest detection accuracy is 0.974 when the number of hidden layers is 6.

Literature Survey 5	
Title	Deep Belief Network Integrating Improved Kernel-Based Extreme Learning Machine for Network Intrusion Detection
Authors	Zhendong Wang , Yong Zeng , Yaodi Liu and Dahai Li
Published Year	2021
Efficiency	Trustworthy and reliable, which refers to obtain explainability. It provides easy information processing and cost reduction as well. Proving High Robustness and imperceptibility
Drawbacks	Unsuitable for large scale scenarios. Difficult to be used in large-scale parallel computing. Complexity of its Real Time Implementation
Description	Deep learning has become a research hotspot in the field of network intrusion detection. In order to further improve the detection accuracy and performance, we proposed an intrusion detection model based on improved deep belief network (DBN). Traditional neural network training methods, like Back Propagation (BP), start to train a model with preset parameters such as the randomly initialized weights and thresholds, which may bring some issues, e.g., attracting the model to the local optimal solutions, or requiring a long training period. We use the Kernel-based Extreme Learning Machine (KELM) with the supervised learning ability to replace the BP algorithm in DBN in a bid to ameliorate the situation. Considering the problem of poor classification performance usually caused by randomly initializing kernel parameters with KELM, an enhanced grey wolf optimizer (EGWO) is designed to optimize the parameters of KELM. In order to improve the search ability and optimization ability of the traditional grey wolf optimizer algorithm, a novel optimization strategy combining the inner and outer hunting is introduced. Experiments on KDDCup99, NSL-KDD, UNSW-NB15 and CICIDS2017 datasets show that the proposed DBN-EGWO-KELM algorithm has greater advantages in terms of its accuracy, precision, true positive rate, false positive rate and other evaluation indices compared with BP, RBF, SVM, KELM, LIBSVM, CNN, DBN-KELM and other intrusion detection models, and can effectively meet the requirements of intrusion detection of complex networks. We propose an intrusion detection method based on an improved deep belief network. A novel kernel extreme learning machine classication model is designed using enhanced grey wolf optimizer optimization, which extracts Z. Wang et al.: DBN Integrating Improved KELM for Network ID Z. Wang et al.: DBN Integrating Improved KELM for Network ID data features by employing the dimensionality reduction abil- ity of DBN for complex high-dimensional network intrusion data features. The combination with the enhanced grey wolf optimizer is viable to optimize the kernel extreme learning machine classication model for the purpose of improving the performance of KELM.

Literature Survey 6	
Title	Network Intrusion Detection Based on Extended RBF Neural Network With Offline Reinforcement Learning
Authors	Manuel Lopez-Martin , Antonio Sanchez-Esguevillas , Juan Ignacio Arribas and Belen Carro
Published Year	2021
Efficiency	👍 Lowering the Complexity Threshold 👍 Tolerates Variations 👍 Provides the integrity and nontransferablity.
Drawbacks	👎 High level of communication and computation overheads 👎 Big payloads 👎 Narrowly specialized knowledge
Description	Network intrusion detection focuses on classifying network traffic as either normal or attack carrier. The classification is based on information extracted from the network flow packets. This is a complex classification problem with unbalanced datasets and noisy data. This work extends the classic radial basis function (RBF) neural network by including it as a policy network in an offline reinforcement learning algorithm. With this approach, all parameters of the radial basis functions (along with the network weights) are learned end-to-end by gradient descent without external optimization. We further explore how additional dense hidden-layers, and the number of radial basis kernels influence the results. This novel approach is applied to five prominent intrusion detection datasets (NSL-KDD, UNSW-NB15, AWID, CICIDS2017 and CICDDOS2019) achieving better performance metrics than alternative state-of-the-art models. Each dataset provides different restrictions and challenges allowing a better validation of results. Analysis of the results shows that the proposed architectures are excellent candidates for designing classifiers with the constraints imposed by network intrusion detection. We discuss the importance of dataset imbalance and how the proposed methods may be critically important for unbalanced datasets. Network intrusion detection is an increasingly important problem in modern data networking, and it is an active research eld in which many types of machine learning and deep learning models have been applied. We propose novel extensions to the RBFNN model. These extensions are based on an end-to-end training scheme using gradient descent for all the parameters of the network: the network weights, and the centers and dispersion parameters of the radial basis functions. This end-to-end training scheme allows us to pro- pose several alternative loss functions, different from the cross-entropy generally used for classication. It also allows a complete RBFNN network to be included as the policy network of an ofine reinforcement learning model where the loss function is replaced by a reward function that is not necessarily differentiable. This approach also offers the opportunity to include additional dense hidden layers after the initial RBF layer.

Literature Survey 7	
Title	Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network
Authors	Kaiyuan Jiang , Wenya Wang , Aili Wang and Haibin Wu
Published Year	2020
Efficiency	Achieve a well-balanced tradeoff among various parameters. Built-in error handling Easy scalability
Drawbacks	Narrowly specialized knowledge Maximizes the complexity of the problem High level of communication and computation overheads
Description	Intrusion detection system (IDS) plays an important role in network security by discovering and preventing malicious activities. Due to the complex and time-varying network environment, the network intrusion samples are submerged into a large number of normal samples, which leads to insufficient samples for model training and detection results with a high false detection rate. According to the problem of data imbalance, we propose a network intrusion detection algorithm combined hybrid sampling with deep hierarchical network. Firstly, we use the one-side selection (OSS) to reduce the noise samples in majority category, and then increase the minority samples by Synthetic Minority Over-sampling Technique (SMOTE). In this way, a balanced dataset can be established to make the model fully learn the features of minority samples and greatly reduce the model training time. Secondly, we use convolution neural network (CNN) to extract spatial features and Bi-directional long short-term memory (BiLSTM) to extract temporal features, which forms a deep hierarchical network model. The proposed network intrusion detection algorithm was verified by experiments on the NSL-KDD and UNSW-NB15 dataset, and the classification accuracy can achieve 83.58% and 77.16%, respectively. In this paper, a novel method for intrusion detection system based on the combination of hybrid sampling and deep hier- archical network has been proposed and discussed. Firstly, we combine OSS and SMOTE to construct a balanced dataset for model training. It can reduce the training time of the model and solves the common problems to some extent of inadequate training from unbalanced samples. In addition, a network data preprocessing method is established for com- proposed deep hierarchical network model. Then, classify the input data through the hierarchical network model con- structed by CNN and BiLSTM. The model extracts feature automatically through repeated multi-level learning by taking advantage of the outstanding features of deep learning. Two intrusion datasets (NSL-KDD and UNSW-NB15) have been employed to evaluate the performance of the pro- posed approach. Based on the statistical signicance tests, it could be concluded that the proposed approach outperforms other classiers, such as Random Forest, LeNet, AlexNet, CNN and BiLSTM. The proposed method yields a superior K. Jiang et al.

Literature Survey 8	
Title	Intelligent Processing of Intrusion Detection Data
Authors	Tao Duan , Youhui Tian , Hanrui Zhang , Yaozong Liu , Qianmu Li , Jian Jiang and Zongsheng Shi
Published Year	2020
Efficiency	It is a fast and easy procedure to perform Powerful Reprsentation Capability Minimal time cost and memory usage.
Drawbacks	Poor Application Performance Difficult to be used in large-scale parallel computing. Heavyweight
Description	Intrusion detection technology, as an active and effective dynamic network defense technology, has rapidly become a hot research topic in the field of network security since it was proposed. However, current intrusion detection still faces some problems and challenges that affect its detection performance. Especially with the rapid development of the current network, the volume and dimension of network data are increasing day by day, and the network is full of a large number of unlabeled data, which brings great pressure on the data processing methods of IDS. In view of the tremendous pressure of intrusion detection brought by the current complex and high-dimensional network environment, this paper provides a feasible solution. Firstly, this paper briefly outlines the necessity of feature learning, the shortcomings of traditional feature learning methods and the new breakthroughs brought by deep belief network in feature learning, and focuses on the principle and working mechanism of deep belief network and Principal Component Analysis (PCA). Then, it constructs the intrusion detection model based on PCA-BP and DBN respectively. And through the experimental evaluation of the two detection models, a comparative experiment between deep belief network and principal component analysis is constructed. The experimental results show that deep belief network has unique advantages and good performance in feature learning. Therefore, deep belief network can be applied in the field of intrusion detection to extract effective features from the current high-dimensional and redundant network data, thereby improving the detection performance of IDS and its adaptability to the current complex and high-dimensional network environment. Due to the high dimensionality and redundancy of cur- rent network data, feature learning has become a necessary process for current intrusion detection data processing mod- els. However, the traditional feature learning methods have T. Duan et al.: Intelligent Processing of Intrusion Detection Data certain shortcomings. And the advent of deep learning has brought new directions to feature learning. In order to verify the unique advantages of deep learning-related technologies in feature learning, this paper constructs a PCA-BP-based intrusion detection data processing model and a DBN-based intrusion detection data processing model.

Literature Survey 9	
Title	Intrusion Detection for IoT Based on Improved Genetic Algorithm and Deep Belief Network
Authors	Ying Zhang , Peisong Li and Xinheng Wang
Published Year	2019
Efficiency	👍 May not meet the real-time requirement. 👍 Low Deployment Cost 👍 Achieve a well-balanced tradeoff among various parameters.
Drawbacks	👎 High level of communication and computation overheads 👎 Heavyweight 👎 Difficulties to obtain better performance
Description	With the advent of the Internet of Things (IoT), the security of the network layer in the IoT is getting more and more attention. The traditional intrusion detection technologies cannot be well adapted in the complex Internet environment of IoT. For the deep learning algorithm of intrusion detection, a neural network structure may have fine detection accuracy for one kind of attack, but it may not have a good detection effect when facing other attacks. Therefore, it is urgent to design a self-adaptive model to change the network structure for different attack types. This paper presents an intrusion detection model based on improved genetic algorithm (GA) and deep belief network (DBN). Facing different types of attacks, through multiple iterations of the GA, the optimal number of hidden layers and number of neurons in each layer are generated adaptively, so that the intrusion detection model based on the DBN achieves a high detection rate with a compact structure. Finally, the NSL-KDD dataset was used to simulate and evaluate the model and algorithms. The experimental results show that the improved intrusion detection model combined with DBN can effectively improve the recognition rate of intrusion attacks and reduce the complexity of the neural network structure. Through GA, the optimal individuals can be generated by iterations. DBN can effectively process high complex and high dimensional data, and the classication results are very good. In this paper, the improved genetic algorithm are com- bined with the deep belief networks, GA performs multiple iterations to produce an optimal network structure, DBN then uses the optimal network structure as an intrusion detection model to classify the attacks. In this way, facing different types of attacks, the problem of how to select an appropriate neural network structure when using deep learning methods for intrusion detection is solved, thus it improves the classi- cation accuracy and generalization of the model, and reduces the complexity of network structure. This method has many advantages: on the one hand, the specic network structure generated for specic attack types is higher in classication accuracy than other network structures, which can reach more than 99% of detection rate.

Literature Survey 10	
Title	A Novel Multimodal-Sequential Approach Based on Multi-View Features for Network Intrusion Detection
Authors	Haitao He , Xiaobing Sun , Hongdou He , Guyu Zhao , Ligang He and Jiadong Ren
Published Year	2019
Efficiency	High Effective with Complex Problems Lowering the Complexity Threshold Simplicity and Explainability.
Drawbacks	Maximizes the complexity of the problem Cannot be implemented real time Unsuitable for large scale scenarios.
Description	Network intrusion detection systems (NIDS) are essential tools in ensuring network information security, and neural networks have become an increasingly popular solution for NIDS. However, with the gradual complexity of the network environment, the existing solutions using the conventional neural network cannot make full use of the rich information in the network traffic data due to its single structure. More importantly, this will lead to the existing NIDS have incomplete knowledge of the intrusion detection domain, and making it unable to achieve a high detection rate and good stability in the new environment. In this paper, we take a step forward and extract the different level features from the network connection, rather than a long feature vector used in the traditional approach, which can process feature information separately more efficiently. And further, we propose multimodal-sequential intrusion detection approach with special structure of hierarchical progressive network, which is supported by multimodal deep auto encoder (MDAE) and LSTM technologies. By design the special structure of hierarchical progressive network, our approach can efficiently integrate the different level features information within a network connection and automatically learn temporal information between adjacent network connections at the same time. Based on the three benchmark datasets from 1999 to 2017, including NSL-KDD, UNSW-NB15, and CICIDS 2017, we investigated the performance of our proposed approach on the task of detecting attacks within modern network. The experimental results show that the average accuracy of this method is 94% in binary classification and 88% in multi-class classification, which is at least 2% and 4% super than other methods respectively, and demonstrated that our model has excellent stability. Moreover, we further explore the multimodality and complementarity in traffic data, the experimental results show that the performance of detection model can be further improved in the range 2% to 5% when using our MDAE model to process the features of traffic data. Network trafc is a formal representation of complex net- work behavior and contains rich feature information, but single consideration for intrusion detection cannot make full use of the rich information in network trafc.

Conclusion

The study reviews the datasets developed in the field of Intrusion Detection System (IDS). These datasets have been used for performance evaluation of the EL and DM based IDS. The study revealed that there is a need to update the underlying dataset to identify the recent attacks in the field of IDS with improved performance. This is because the attackers execute attack by using varied processes and technologies. Moreover the pattern of executing different attacks simulates the need to have datasets with realistic network scenarios. To fulfill the requirement of building an intrusion detection dataset with realistic network traffic and updated network attacks CSE-CIC-IDS on AWS datasets have been introduced. This paper reviews the characteristics of these datasets and also discusses a few shortcomings.

Future Work

In the future we focus on studying the performance of these datasets with various ML and DM techniques along with incorporating feature engineering and data sampling to address the shortcomings of these datasets.

References

- , K. Zheng et al., Algorithms to speedup pattern matching for network May , . doi: , ., /j.comcom., ., ., .
- , Y. Xu and H. Zhao, Intrusion detection alarm ltering technology based on ant colony clustering algorithm, in Proc. , th Int. Conf. Intell. Syst.
- , L. Dhanabal and S. P. Shantharajah, A study on NSL-KDD dataset for intrusion detection system based on classication algorithms, Int. J. Adv.
- , T. Wang, G. Zhang, A. Liu, M. Z. A. Bhuiyan, and Q. Jin, A secure IoT service architecture with an efcient balance dynamics based on cloud and Jun. , .
- , T. Clouqueur, V. Phipatanasuphorn, and P. Ramanathan, Sensor deploy- ment strategy for detection of targets traversing a region, Mobile. Netw.
- , W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Energy-efcient communication protocol for wireless microsensor networks, in Proc. , rd Annu. Hawaii Int. Conf. Syst. Sci., Aug. , p. , .
- , F. L. Fessant, A. Papadimitriou, A. C. Viana, C. Sengul, and E. Palomar, A sinkhole resilient protocol for wireless sensor networks: Performance Jan. , .
- , L. Nishani and M. Biba, Machine learning for intrusion detection in , T. Clausen, Optimized link state routing protocol, IETF Internet Draft, Fremont, CA, USA, Tech. Rep. , Jul. , .
- , G. Indirani and K. Selvakumar, A swarm-based efcient distributed intrusion detection system for mobile ad hoc networks (MANET), Int.
- , S. Ganapathy, P. Yogesh, and A. Kannan, Intelligent agent-based intru- sion detection system using enhanced

multiclass SVM, Comput. Intell.

» , L. Ljung, System Identication: Theory for the User. Upper Saddle River, NJ, USA: Prentice Hall, , .

» , H. H. Pajouh, G. Dastghaibyfard, and S. Hashemi, Two-tier network anomaly detection model: A machine learning approach, J. Intell. Inf.

» , S. Park, S. Seo, and J. Kim, Network intrusion detection using stacked Oct. , . , H. Saxena and V. Richariya, Intrusion detection in KDD, dataset using SVM-PSO and feature reduction with information gain, Int. J. Comput.

» , A. Kim, M. Park, and D. H. Lee, AI-IDS: Application of deep learning to Apr. , .

» , L. Thurner, A. Scheidler, F. Schfer, J.-H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, PandapowerAn open-source python tool for convenient modeling, analysis, and optimization of electric power Nov. , .

» , Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, DAEN: Deep autoencoder networks for hyperspectral unmixing, Jul. , .

» , L. Zhao, Z. Wang, X. Wang, and Q. Liu, Driver drowsiness detection using facial dynamic fusion information and a DBN, IET Intell. Transp.

» , X. Sun and W. Xu, Fast implementation of DeLongs algorithm for comparing the areas under correlated receiver operating characteristic Jul. , .

» , S. F. Jilani, Q. H. Abbasi, and A. Alomainy, Inkjet-printed millimetre- , G., ., .

» , M. Kumar and A. K. Singh, Distributed intrusion detection system using blockchain and cloud computing infrastructure, in Proc. , th Int. , ., /ICOEI, ., ., .

» , W. Wang, X. Du, D. Shan, R. Qin, and N. Wang, Cloud intrusion detection method based on stacked contractive auto-encoder and support vector , /TCC., ., .

» , Z. Yan and Y. Xu, A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system, , ., /TPWRS., ., .

» , C. Li, J. Wang, H. Wang, M. Zhao, W. Li, and X. Deng, Visual-texual emotion analysis with deep coupled video and danmu neural networks, , ., /TMM., ., .

» , K. Zhu, Z. Chen, Y. Peng, and L. Zhang, Mobile edge assisted literal multi-dimensional anomaly detection of in-vehicle network using LSTM, , ., /TVT., ., .

» , B. Riyaz and S. Ganapathy, A deep learning approach for effective intru- , P. Mishra, V. Varadharajan, U. Tupakula, and E. S. Pilli, A detailed inves- tigation and analysis of using machine learning techniques for intrusion , st Quart., , doi: , ., /comst., ., .

» , K. Nugroho, E. Noersasongko, Purwanto, Muljono, and H. A. Santoso, Javanese gender speech recognition using deep learning and singu- lar value decomposition, in Proc. Int. Seminar Appl. Technol. Inf. TIC., ., .

» , L. Ertz, M. Steinbach, and V. Kumar, Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data, in Proc. SIAM Int.

» , H.-T. Li, C.-Y. Chou, Y.-T. Chen, S.-H. Wang, and A.-Y. Wu, Robust and lightweight ensemble extreme learning machine engine based on , ., /TCSI., ., .

» , C. Alippi and M. Roveri, Virtual k-fold cross validation: An effective method for accuracy assessment, in Proc.

Int. Joint Conf. Neural Netw.

» , A. Aldweesh, A. Derhab, and A. Z. Emam, Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and doi: , ., /j.knosys., ., .

» , S. Gamage and J. Samarabandu, Deep learning methods in net- work intrusion detection: A survey and an objective comparison, , ., /j.jnca., ., .

» , M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, Application of deep reinforcement learning to intrusion detection for supervised prob- , ., /j.eswa., ., .

» , S. Levine, A. Kumar, G. Tucker, and J. Fu, Ofine reinforcement learning: Tutorial, review, and perspectives on open problems, , arXiv:, ., . , A. AbuGhazleh, M. Almiani, B. Magableh, and A. Razaque, Intelligent intrusion detection using radial basis function neural network, in Proc. , ., /SDS., ., .

» , Z. Yang, X. Wei, L. Bi, D. Shi, and H. Li, An intrusion detection system based on RBF neural network, in Proc. , th Int. Conf. Com- , ., /CSCWD., ., .

» , L. Lv, W. Wang, Z. Zhang, and X. Liu, A novel intrusion detec- tion system based on an optimal hybrid kernel extreme learning , ., /j.knosys., ., .

» , T. Poggio and F. Girosi, Networks for approximation and learning, Proc.

» , C.-G. Li, M. Wang, Z.-J. Huang, and Z.-F. Zhang, An actor-critic rein- forcement learning algorithm based on adaptive RBF network, in Proc. , ., /ICMLC., ., .

» , Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, Kitsune: An ensem- ble of autoencoders for online network intrusion detection, , arXiv:, ., .

» , T. Thi Nguyen and V. Janapa Reddi, Deep reinforcement learning for cyber security, , arXiv:, ., .

» , I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion trafc characterization, , ., /, .

» , M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, and S. Ghemawat, Ten- sorFlow: Large-scale machine learning on heterogeneous distributed sys- tems, , arXiv:, ., .

» , H. van Hasselt, A. Guez, and D. Silver, Deep reinforcement learning with double Q-learning, , arXiv:, ., .

» , V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, Playing atari with deep reinforcement learning, , arXiv:, ., . , R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, A comprehensive survey on machine learning for networking: Evolution, applications and research , ., /s, -, -, -, .

» , M. Azizjon, A. Jumabek, and W. Kim, , D CNN based network intru- sion detection with normalization on imbalanced data, in Proc. Int. , ., /ICAIIC, ., ., .

» , J. E. B. Maia, V. R. S. Laboreiro, F. E. Chaves, F. J. A. Maia, T. G. N. Silva, and T. N. Ferreira, Performance comparison between edited kNN and MQ-RBFN for regression and classication tasks, in Proc. , st Brazilian , .

» , . avuolu, A new hybrid approach for intrusion detection using Jul. , .

» , I. S. Thaseen, C. A. Kumar, and A. Ahmad, Integrated intrusion detection model using chi-square feature selection and ensemble of classiers, , I. S. Thaseen and C. A. Kumar, Intrusion detection model using fusion of chi-square feature selection and multi class SVM, J. King Saud Univ.

» , Z. Liu, Z. Lai, and W. Ou, Structured optimal graph based sparse fea- May , Art. no. , doi: , ., /j.sigpro., ., .

» , J. Kim, J. Kim, H. Le T. Thu, and H. Kim, Long short term memory recurrent neural network classier for intrusion detection, in Proc. Int.

» , M. Wang and J. Li, Network intrusion detection system based on con- , .

» , R. Zazo, P. S. Nidadavolu, N. Chen, J. Gonzalez-Rodriguez, and N. Dehak, Age estimation in short speech utterances based on LSTM , .

» , S. Wan, Y. Xia, L. Qi, Y.-H. Yang, and M. Atiquzzaman, Auto- mated colorization of a grayscale image with seed points propagation, TMM., ., .

» , S. Wan, Y. Xia, L. Qi, Y.-H. Yang, and M. Atiquzzaman, Auto- mated colorization of a grayscale image with seed points propagation, TMM., ., .