Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN INTELLIGENT MACHINE LEARNING FRAMEWORK FOR ANOMALY AND SECURITY ATTACK DETECTION TO ENABLE SECURE AUTHENTICATION

KALYAN KUMAR DASARI[1],Dr.K. SAHADEVAIAH[2]

[1]Research scholar, Dept. of Computer Science & Engineering, JNTUK, Andhra Pradesh, India.
[2]Professor, Dept. of Computer Science & Engineering, JNTUK, Andhra Pradesh, India.
Email: [1]*dkkumar123@gmail.com*, [2]*ksd1868@jntucek.ac.in*

ABSTRACT

The authentication process entails transfer of sensitive data between the device or account of the user and the authentication server in a bid to identify an identity of the requester. The authentication process needs to be secured against security assaults. Authentication techniques are numerous depending on the requirements of the system. However, any authentication method cannot be fully secure. This study was aimed at identifying the security threats and abnormalities by employing the device and user context in the authentication process. Context has also been used to examine the Denial-of-service as DoS, attacks like brute-force and distributed denial-of-service (DDoS) attacks. Extensive simulations happening on the standard dataset is CIC-IDS2018 were conducted with python programming. The accuracy, recall, f-score precision and also built time of each of four ML classifiers that are decision tree classifier model, Random Forest classifier, k-nearest neighbour classifier and SVM model were computed on the various combinations of splits of data and various features splits. Approximately the tests give the f-scores, recalls, precisions, and accuracy of over 98.0 of brute-force and DoS/DDoS attacks. Discoveries on the tests security and safety analysis and threat modelling protest the projected validation and authentication strategy can be used to raise the level of security of a secure system.

Keywords: *Sensitive Information, Authentication, DDoS attacks, Machine learning classifier, Brute force, Random Forest.*

1. INTRODUCTION

To provide the security the authentication process is the main scenario to verifying the identity of a human individual requesting admittance of a system. Examples of some of the existing user authentication mechanisms are those based on knowledge (e.g. passwords and personal identification numbers), those are created on control (for example of RSA tokens and ATM cards) those founded on inheritance (e.g. biometrics such as fingerprints and face recognition system) in addition those founded on someone's might know in the different social networks. Malware, asterisk loggers, keyloggers, eavesdropping, phishing replay, dictionary, Trojan horse, brute-force, SQL injection, man-in-the-middle and numerous other attacks are so prevalent with the existing authentication mechanisms. Machine learning (ML) is a term that describes a category of algorithms that utilize information with the aim of improving some jobs. The scope of this study is to develop and evaluate a machine-learning-based security framework that uses user and device contextual information to enhance authentication and identify abnormal network activities. The experimental work primarily focuses on detecting DoS/DDoS and brute-force attacks using the CIC-IDS2018 dataset. Four supervised machine learning algorithms Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbour (KNN), and Random Forest are examined using different feature sets and data-splitting configurations. The study also explores the impact of feature selection using the InfoGainAttributeEval method and evaluates the classifiers based on accuracy, precision, recall, F-score, and model-building time. The proposed framework is designed to complement existing authentication mechanisms by providing an additional layer of security rather than replacing them.It can also be employed to condition a model to pass judgement or prediction without the involvement of a human being. Each of these jobs requires a different set of conventional machine learning algorithms. Efficient detection and identification of dynamically shifting cyberattacks through machine learning are one application [1-3]. ML models can train towards the recognize and

neutralise cyberthreats based on data of past attacks and threats. The datasets are vital in machine learning models to be done training and testing process. Data sets mandatory to perform some activities could be either developed internally or acquired through open-source databases. Every quantifiable unit of information that is used to make an analysis is denoted by a feature or trait in a dataset. In Online they remain so many application libraries of machine learning such as Sci-Kit Learn, PyTorch, TensorFlow among many others. In data preparation, classification, regression, clustering, association, and visualisation you can use an open-source program known as python [4]. Its user-friendly UI, as well as extensive selection of machine learning techniques make data mining jobs easier. The authors of this work construct and test ML models by the aid of the Python libraries. This article discussing some of the most popular machine learning classification tools such as SVM and random forest. It takes all the tree predictions and takes the one that has the highest number of votes to come up with the final output. A multi-layer perceptron (MLP) neural network takes the input, output, and hidden layers. To process an input signal, an input layer receives the input signal whereas the output layer performs the actual processing, e.g. classifying the data. The brains of the MLP are the hidden layers, between the outer two layers. The Random Forest model is one probabilistic graphic representation which employs directed edges in a graphical representation to explicitly state the known conditional dependence. Certain practices and studies in the field of security have proposed the use of extra contextual data to make superior security decisions and threat identification. A fundamental aspect of data security is to afford confidentiality towards the huge amounts of data or information produced by the reputed organisations. Several research papers have proposed the use of contextual information within a system to facilitate privacy protection [5-7]. The context awareness of a privacy-preserving system is provided in these works by building an ontology data model. As a context-based approach to access, Al-Turjman et al. proposed healthcare applications based on the Industrial Internet of Things (IIoT) [8]. It decreases the likelihood of identity stealing and loss through in view of the constantly shifting habits of the residents, the access in addition with the practice of patterns. Karabatis et al. [9] developed a new type of intrusion detection system (IDS) based on contextual information in order to recognize multistage assaults (MSAs). As individual stages or phases of an MSA might not be harmful by themselves, the attack is effective when the successive phases are implemented. To detect cyber-attacks, Melo W et al. [10] proposed a new contextual framework to predict attacks through multiple attack prediction models together with intrusion detection systems. The purpose of the contextual framework in this instance was to increase the ability to be able to identify both typical and unusual instances of assault. De-Marcos et al. [11] proposed a physical context-based authentication system in their proposal to secure them against external threat. The authors in this work illustrated how things in a common physical setting can safely communicate with each other and protect against external dangers. The current article goes a step further than the common authentication options we have today to consider user and device context during the authentication process to detect security violations and other abnormalities. The credibility of the person requesting authentication is assessed by an independent machine learning network operating on and interpreting the situation of the context. A major limitation is that the experimental evaluation relies primarily on the CIC-IDS2018 dataset, and therefore the results may not fully reflect the behaviour of attacks in continuously changing real-world networks. The study also assumes that the dataset is sufficiently representative of the target environment and that the training data are reliable and appropriately labelled. Part 2 of this paper describes the past related work, Part 3 describes the proposed system, Part 4 describes the experiments and their results using different machine learning algorithms, and lastly Part 5 conclude the proposed system and a discussion about future work.

2. RELATED WORK

In behavioral biometrics, the actions of the user when using the device are taken as biometric data. It is the science of human behaviour that is concerned with the quantification of observable patterns. A number of systems and devices have started to employ behavioral biometrics [12-15]. In this research offer the behavioral biometricsas the best scenarios of user verification. A lot of the supervised and unsupervised ML models are still exploited to discovery the distributed denial-of-service (DDoS) attacks. To classify the different types of distributed denial of service (DDoS) attacks, Saini PS et al. [16] used the Naive Bayes, UDP-Flood, and Smurf algorithms on the data. The highest rate of accurateness in the MLP classifier

was 98.63 percent on the dataset comprising of 27 characteristics with the same dataset as to detect DDoS attacks.

Obeidat et al. [17] used the J48, MLP, and Random Forest algorithms, and Naive Bayes. Among the four algorithms that were experimented, J48 proved to have the highest accuracy rate at 98.64 percent. Kurniabudi et al. [18] in their work considered the precision of the algorithms of the Random Forest and the MLP in terms of recognizing denial-of-service (DoS) attacks based on the 2017 dataset by the Canadian Institute for Cybersecurity Intrusion Detection Systems (CIC IDS). The dataset has 84 characteristics. The experiments have shown that Random Forest beat MLP across a range of training/testing data split percentages. In intrusion detection, Alsaif et al. [19] were using the CIC IDS 2017 dataset and information gain feature selection method to identify characteristic importance.

As a criterion to differentiate between typical and denial-of-service (DoS) attacks, Wanjau S et al. [20] relied on the CAIDA 2007 dataset and the machine learning techniques of logistic regression and naive bayes. The initial data set of 20,090 records is split into three parts that are one is training set; second one is testing set as well as validation set as of the ratios are 70:20:10. Here, the accuracy of the Logistic Regression is 99-100 percent and Naive Bayes stands at 99-98 percent. Conversely, the CIC IDS 2017 dataset contains more recent kinds of DoS attacks which do not exist in the CAIDA 2007 dataset. Saito et al. [21] proposed a Convolutional Neural Network-based method in order to detect SSH brute force network attacks. In the experiment we used a modified version of the CIC-IDS 2018 dataset of 80 characteristics. We got a 98.3 recall rate, 99.8 accuracy, 99.3 precision rate and 91.8 F1-score once it comes to the recognition of SSH-brute force assaults.

Agghey et al. [22] investigated SSH and FTP brute-force attack detection using a deep learning model known as Long short-term memory (or LSTM). The LSTM and ML algorithms were assessed on CIC-IDS 2017 dataset that includes entire 80 features, in this 80:20 ratios used as training and testing sets. In this study use the LSTM model achieved the accuracy of 99.88% by using the brute-force attacks involving FTP and SSH. Rizk et al. [23] in their research discuss the use of k-nearest neighbour (K-NN), naive bayes (NB), random forest (RF), and decision tree (DT) as machine-learning classifiers to learn more about username enumeration attack detection on the SSH protocol. The results of this KNN model get the 99.93% accuracy, the NB gets the 95.70% accuracy, the random-forest got the 99.92%accuracy as well as the decision tree attain the 99.88% accuracy by the using total of 36,273 raw packets each with 25 characteristics. We have prudently studied on the submission of Random Forest, SVM, Decision tree and KNN machine learning algorithms by the detection of attacks like denial-of-service (DOS). The experiment involved the use of a feature reduction method in the analysis of 80, 20 and 10 features of the CIC IDS 2018 dataset. The dataset was divided 30% and 40%. This study shows the constructed system achieves better than another proposed models on f-score, precision, recall, accuracy and time taken for execution across the amalgamations of the Machine learning models, splitting style of data and features combination.

Sharafaldin et al. [24] used the CIC-IDS2018 dataset to experiment with five classifiers: decision tree, random forest, linear support vector, gaussian naive bayes and multilayer perceptron. Specifically, all of the five classifiers were not very good at the Brute-force assault, which was characterized by the significant imbalance in classes had only 362 data arguments of illegal traffic movements and it was only 10 features were employed. We have done thorough research on how to employ machine learning techniques like a SVM, Random Forest, Decision Tree and K-Nearest Neighbor classifiers are implemented for finding the brute force attacks. The experiment involved the use of a feature reduction methodology to analyze 80, 20 and 10 features of CIC IDS 2018 dataset. The dataset was divided 80% and 20%. The study demonstrates the model generated achieves greater than the existing models in the aspect of precision, f-score recall, accuracy and time taken by the execution of the various amalgamations of the ML systems under investigation, features and data splits.

Existing authentication methods mainly rely on passwords, physical tokens, or biometric information, but these approaches can still be exposed to increasingly sophisticated cyberattacks, including brute-force and DoS/DDoS attacks. Although previous studies have shown that machine-learning techniques can effectively detect network attacks, most of them focus primarily on

network traffic characteristics or specific types of attacks and give limited attention to user and device context during authentication. In addition, using a large number of network features can increase computational complexity and processing time. This highlights the need for an adaptive and efficient security framework that combines contextual information with machine learning to strengthen authentication and identify abnormal activities without compromising detection performance.

3. PROPOSED ALGORITHM

These are the key ideas behind the solution what we are implementing. The details of the mechanism of the authentication process in the processing background information are also included.

3.1. Working Procedure

Background context checking and analysis of the user and device will be performed along with authentication. To establish the credibility score of user or device when authenticating, it is possible to construct a prediction model or classifier using previous data with the assistance of context analysis. A machine learning network model is construct based on a data collection that is in a state of constant change. The machine learning analysis can be performed on-cloud, and controlled by a centralized network so that the technological stack is update very normal and easy updated without disturbing and effecting the end users. In any network system to system the framework/background of the users and devices are different manner because of able to fit the contextual factors to the requirement of your system. Among the numerous contextual factors, which could be are:

Repetition of Network address- Internet Protocol (IP) reputation intelligence assists in warding off malware sources, malicious or suspicious content among internet users. Authentication can be denied to an organization where the IP address of the machine of a user is recognized to be of bad actors. The user may have indicators of a vulnerable system in both software and applications, including infrequent updates of operating system security patches, the absenteeism of the anti-malware programs or the existence of the malicious software that is keyloggers. Context-based authentication can be used to track the current geographical location of the user to understand whether an authentication request might be originating in a particular area. A combination of user geographic physical location and their login-logout historical data also be used to help avoid unauthorized access. One of the travel situations that can trigger suspicion would be when a user logs in with a Canadian IP at 9 a.m., and then he/she logs in with a Dubai IP at 10.00 a.m. on the same day. The habits of a user can be learnt by monitoring their activities on a device over time, as by capturing the dynamics of their keystrokes and mouse movements. Any further use which is too inconsistent in its patterns of activity can arouse suspicion. Information of the network which is used to recognize abnormal behavior on the network.

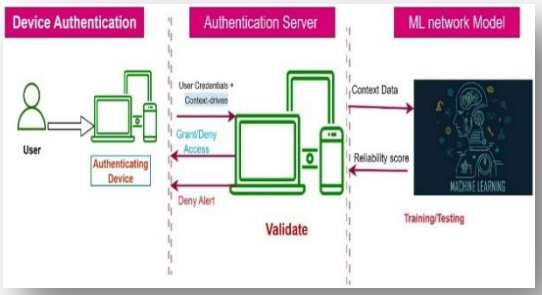

Figure 1: Proposal Model for Authentication

3.2 User/device Registration

A number of methods of authentication already exist that can be used to verify the user or using device. That techniques are intelligent based that are a private identification number and a password authentication, possession type authentication techniques that are RSA token and ATM cards, parental based authentication that are biometrics, such as finger prints or facial recognition as well as the someone you recognize in the social networks. The verification scheme-related confidential information is stored at the server-side upon registration to confirm the identity of the client seeking access to the resources. The context-based authentication method proposed can be altered to meet the different extant authentication schemes. It will be done in the background and will be invisible to users (or possible enemies) of a secure scheme.

3.3 Authentication Process

A device or a user has to be authentic in order to entree diverse structures and sensitive information. This is therefore of dominant position to ensure that the verification stage is not vulnerable to unwary security attacks. There is an exchange of information among the verification and validation server and the device/user as part of the authentication procedure to authenticate the individuality. The verification procedure as shown in Figure 1 is set to support user and device authentication.

3.3.1 Primary Stage

During the authentication procedure as exposed on Figure 1, user credentials in the form of usernames and passwords are used. However, the basic infrastructure of the given system will be similar to the other authentication procedures.

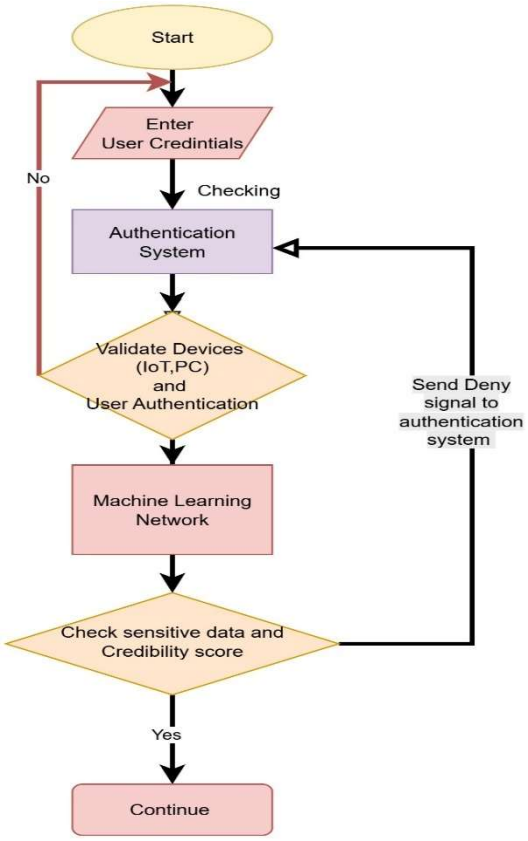

Figure 2: Flowchart Of Authentication Process

When a user logs in, information about the person and the device they are using is collected and sent to the authentication server along with their credentials. As part of the authorization process, standalone devices provide confidential information about the device, such as its digital certificate along with contextual information, to the authentication server. The authentication server authenticates the device/user using the confidential data that matches an individual user/ device. The comparison between the private data like user credential and the data stored in the verification/authentication server is done through the registering process. In case of device/user authentication, the verification and validation server identifies the request as illegal or unauthorized. Happening on the other scenario, the user is authorized to the system/resources provided the confidential information is correct.

3.3.2 Secondary Stage Authentication

The server of authenticating data sends the resulting background data of the device/user towards a separate ML network in the background. This model is an autonomous body it puts the data it receives in perspective. This makes the ML network model easy to place in a shared online cloud system since the technological stack is updated not affecting the end users. The strategy of this data outsourcing makes it possible to analyze voluminous datasets on the strong computational capabilities of a confidential ML network, which improves the work of detection. A range of contextual parameters, such as the standing of the IP address, network information, the behavioral pattern of the users, location information and device software information, could be measured, contingent on the requirements of the structure. The contextual analysis gives the device or user an acceptability/credible score, which might help determine towards if the requesting entity is real. Based on the trustworthiness score, the authentication server can either deny an authentication request or notify the user know which one has been requested.

3.4 Intrusion Detection Model

In this paper, a detailed analysis and modelling of an intrusion detection system based on the anomaly take remained conducted towards identify security attacks by utilizing the contextual information of users and devices. Specifically, a variety of ML models take remained to develop then evaluated on the way to identify the brute-

force as well as DoS/DDoS attacks. The flow chart of proposed model is constructed and shows the series of actions, as illustrated in the Figure 2.

3.4.1 The Data Collection and Preprocess

In this paper, this dataset is utilized, the CIC-IDS 2018 Canada dataset, to examine the brute force as well as DoS/DDoS attacks. This data comprises benign then also regular attacks similar to realistic data on the real-world. The data remained created towards create a real-world contextual traffic in the network. It is naturalistic, benign background traffic generating and profiling the abstract behaviour of human interactions based on the B-profile system [25-27]. The dataset covers the many data files which reflect every attack, such as brute-force and DoS/DDoS. The rows in a data file identify the data elements whereas the columns identify the features. The second step in the model is preprocessing operation. During these phases, the quality of the data is safeguarded by means of application of techniques, including the handling of missing values, anomaly elimination, and conversion of the information into the appropriate format. Fashionable the data files of DoS/DDoS in addition of brute-force attacks, every data item is described by 80 features. Fwd Header Length is found to be a replica of the 80 features. This dataset contains any duplicates then remove the duplicate feature and perform the data analysis using 80 features.

3.4.2 Selection of Features

To train, test and validate the ML model, some of the dataset's attributes are carefully chosen and employed in the feature assortment stage. The dataset CIC-IDS2018 comprises the 80 features that all are used to do an exhaustive analysis. In addition to a thorough analysis of the dataset, a feature selection method has been implemented to evaluate the features and select the most relevant top-ranking features for this study. The extracting method of features minimizes the used data popular into a more dimensional data to very less dimensional data space. The accuracy of the model can also be enhanced due to the decreasing number of misleading data, and the time of calculation is lower since there are fewer features. This paper uses the ranking method that is InfoGainAttributeEval method to assign the ranks to every feature in the dataset, and is one of the many feature ranking methods that can be offered. The InfoGainAttributeEval algorithm uses the features to be selected based on the contribution of each feature to the decrease in the aggregate entropy.

3.4.3 Classification, Training and Testing

This proposed paper talks about some of the most popular machine learning classification algorithms like the Support vector machine (SVM), Decision Tree, KNN and Random Forest. The training of a model, or model training, refers to the process of recognizing patterns within the data by using the training dataset have chosen by ML algorithms. The result of this process is a ML model which is trained by using 80 percent in used CIC-IDS2018 dataset on brute-force and DDoS/DoS attacks as well as testing process use the 20 percent in the dataset for these experiments. Another experimental scenario is the same data files are splitting into the 30% and 40% can be used to conduct the experiment by using the any of this model.

3.4.4 Model Evaluation Metrics

Within this model the accuracy of the trained model was measured using the used dataset the ML classifiers assessed using confusion matrix which are positive and negative numbers. The classifiers notation is TN is true negative; FP is false positive; TP is true positive and FN is false negative.

$$\text{Overall Acc}=(TP+TN)/(TP+TN+FP+FN) \quad (1)$$

$$\text{Precision}=TP/(TP+FP) \quad (2)$$

$$\text{Recall}=TP/(TP+FN) \quad (3)$$

$$\text{F-Score}=(2*(P*R))/(P+R) \quad (4)$$

4 EXPERIMENTS AND RESULTS

A great number of experiments were introduced to determine DoS/DDoS and bruteforce attacks. This part provides a detailed description of these approaches. In this work conduct the experiments and implement the ML models using the Python. The generated ML models are analyzed by a different types of performance metrics such as f1-score, precision, accuracy and model-built time. All experiments have been carried out to configure

the system in such a way that the performance times of various systems can be compared. System setup: 64-bit windows operating system, x64 processor. Property Processor: Intel (R) Core (TM) i7-1235U-13th generation-1.30 GHz,16GB RAM, it is assumed that the data is available in a single uniform file or relation, and every data point is represented by a given number of attributes. Moreover, it is supposed that the trained sample will be a illustrative example of the populace to reduce the chances of bias in the developed model. This research uses the CIC-IDS2018 dataset which is deemed to be reflective of the real-world data. This data set comprises a more variety of modern DoS/DDoS and brute-force attacks, both genuine and also malicious.

4.1 DoS/DDoS Attack

In such a cyberattack, the attacker attempts to stop people from using a machine or network resource by messing up the services of a host. Most of the time, this is done by sending too many requests to the target machine or resource in an attempt to break the systems so that important requests can't be met.

4.1.1 Analysis of Dataset

At first both the benign as well as DoS/DDoS attacks data file was used to analyze by using the CIC-IDS2018 dataset. The dataset consists of 692,703 records and 80 features, which are labelled as Heartbleed, DoS Slowhttptest, DoS slow Loris, DoS Golden Eye, DoS Hulk and Benign. In generally the name "DOS" attacks are used in this paper to represent the many types of DoS attacks that will be analyzed. Then, the converted data set will consist of two different types of labels, the label DOS and the label BENIGN. The total number of 692,703 files has 440,031 entries of BENIGN and 252,672 entries of DOS. The DoS/DDoS attacks are analyzed and classification by using the four most prominent machine learning models that are SVM, Random Forest, KNN and Decision tree were implementing in Python through TensorFlow libraries. To generate all the model 80% of dataset is used to train the model and the 20% of dataset is used to test the model. The Python source code will create the confusion matrix for each group of the training and testing data for each classification model. The dataset is split into 30% and 40% of divisions are applied to analyze the DoS/DDoS attacks on all models on CIC-IDS 2018 dataset. Consequently, the model is trained on 80-percent and 20-percent cases respectively, whereas the rest of the cases are used to test the model. The dataset has 80 features that experience an in-depth analysis of DoS/DDoS attacks. To estimate the most important features, the InfoGainAttributeEval ranking strategy was applied so as to rank the 80 features. Here finding in the dataset are topmost 20 features as well as top ranked 10 features are analyzed and the DoS/DDoS attacks. This dataset contains the 80 features, top 20 features, and top 10 features are compared and analyzed on various metrics including performance metrics. The resulting confusion matrix of the generated SVM, Decision tree, KNN and Random Forest and is measured by using the same equations, i.e., 1, 2, 3, and 4 respectively with 40% and 30% splits of data and different types of feature groups like 10,20 and 80.

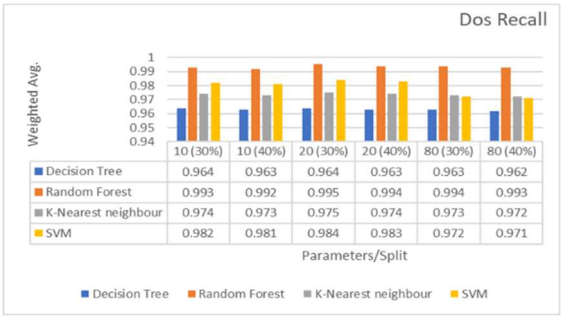

Figure 3: Recall of DoS/DDoS Attacks

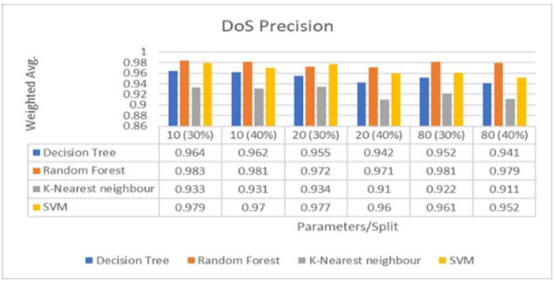

Figure 4: Precision of DoS/DDoS Attack

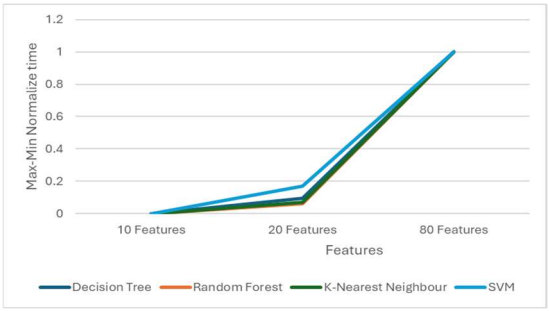

Figure 5: Accuracy of DoS/DDoS Attacks

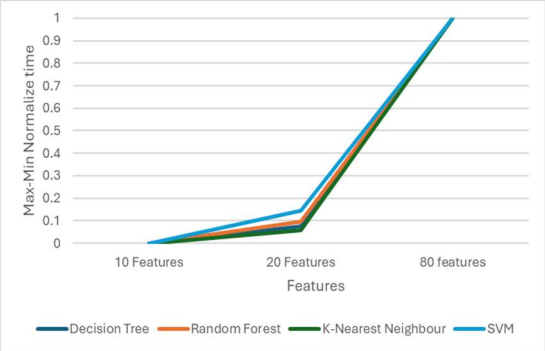

Figure 6: Time for Min-max Normalization with a 30% Split of Dataset of Dos/DDoS Attacks

A graphical representation of measuring the different performance metrics is recall, precision and accuracy accomplished in the Figures shown in 3, 4 and 5. The models built in time is calculated in seconds of the 30 percent and 40 percent of the data division combinations of ML algorithms and features are made known in the below Table 1. The time it took to build the model is changed to Min-Max Normalization time so that the results can be compared. In this normalization data scaling technique lowest value is 0, highest value is 1 and remaining values are between these range. The table 2 represents the Min-Max Normalization values of different ML models of different data divisions and various features of data split. The figures 6 and 7 indicate the Min-Max normalized time of the 30 and 40 percent data divisions respectively.

Table 1: Model Built Time (seconds) of DoS/DDoS Attacks.

Features	10		20		80	
Split %	30 %	40%	30%	40 %	30 %	40 %
Decision Tree	110.7	116.99	243.43	283.3	857.4	945.5
Random Forest	127.2	132.29	263.34	283.3	894.5	944.4
K-Nearest Neighbour	278.2	289.24	384.44	394.4	1209	1484
SVM	110.3	134,98	256.34	267.4	890.5	987.5

Table 2: Min-Max Normalization of Data Splits and Features of Dos/DDoS Attacks

Data Split	30% Data Split			40% Data Split		
Features	10	20	80	10	20	80
Decision Tree	0	0.09563	1	0	0.07355	1
Random Forest	0	0.05952	1	0	0.09575	1
K-Nearest Neighbour	0	0.06751	1	0	0.05784	1
SVM	0	0.16847	1	0	0.14453	1

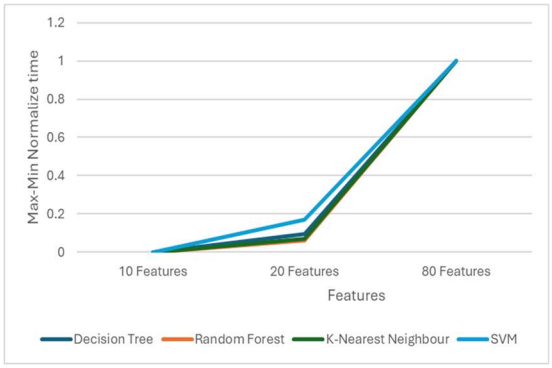

Figure 7: Time for min-max normalisation with a 40% split of dataset of Dos/DDoS attacks.

4.1.2 Implication

Figures 3, 4 and 5 show that the recall, precision and accuracy values of the SVM and Random Forest algorithm is better than the Decision tree and the KNN models in all combinations of features and division of data. When compared to the all implemented four models the SVM and Random Forest models are found to be more successful with recall, precision, and accuracy (Fig. 3,4 and Fig. 5) and can be run in nearly 5 to 8 times less by 10 and 20 features than 80 features. It is therefore desirable to assume that when trying to develop a ML model to perceive DoS/DDoS attacks it is worth considering 20 features although there are 80 features, and with different metrics in mind including performance metrics. Figures 6 and 7 indicate that the Random Forest algorithm is greater accuracy than the SVM algorithm in the aspect of 30% and 40% data division. The model of the ML algorithm that was trained on the Random Forest algorithm and takes into consideration 20 features is then chosen for

detecting DoS/DDoS attacks is the best authentication solution.

4.2 Brute-Force Attack

A brute force attack is a technique used in hacking whereby trial and error are used to decipher invalid logon credentials. Usually, a combination of automated programs and tools are used to work through all possible combinations of user credentials.

4.2.1 Analysis of Dataset

The dataset CIC-IDS2018 is used the input data which describes the brute force attack and benign traffic. The dataset contains 445,909 records and 80 features that are classified as either FTP-Patator, Benign and SSH-Patator. In this article, the many brute-force attack categories are represented in a single form as Brute-Force attacks so that they can be analyzed collectively. In turn, the transformed dataset can be viewed as having two categories of labels in it, namely, BENIGN and Brute-Forces. This dataset contains the 432074 BENIGN records and 13835 Brute-Force records by the total of 4,45,909 records.

The Brute-Force attacks are analyzed and classification by using the four most prominent machine learning models that are SVM, Random Forest, KNN and Decision tree were implementing in Python through TensorFlow libraries. To generate all the model 80% of dataset is used to train the model and the 20% of dataset is used to test the model. The Python TensorFlow library will create the confusion matrix for each group of the training and testing data for each classification model. The dataset is split into 30% and 40% of divisions are applied to analyze the Brute-Force attacks on all models on CIC-IDS2018 dataset.

Consequently, the model is trained on 80-percent and 20-percent cases respectively, whereas the rest of the cases are used to test the model. The dataset has 80 features that experience an in-depth analysis of Brute-Force attacks. To estimate the most important features, the InfoGainAttributeEval ranking strategy was applied so as to rank the 80 features.. This dataset contains the 80 features, top 20 features, and top 10 features are compared and analyzed on various metrics including performance metrics. The resulting confusion matrix of the generated SVM, Decision tree, KNN and Random Forest and is measured by using the same equations, i.e., 1, 2, 3, and 4 respectively with 40% and 30% splits of data and different types of feature groups like 10,20 and 80. A graphical representation of measuring the different performance metrics is recall, precision and accuracy accomplished in the Figures shown in 8, 9 and 10. The models built in time is calculated in seconds of the 30 percent and 40 percent of the data division combinations of ML algorithms and features are made known in the below Table 3. The time it took to build the model is changed to Min-Max normalization time so that the results can be compared. In this normalization data scaling technique lowest value is 0, highest value is 1 and remaining values are between these range. The table 4 represents the Min-Max normalization values of different ML models of different data divisions and various features of data split. The figures 11 and 12 indicate the Min-Max normalized time of the 30 and 40 percent data divisions respectively.

4.2.2 Implication

With Figures 8, 9, and 10, it can be seen that the values of recall, precision, and accuracy of SVM, Decision tree, KNN and Random Forest and are greater in almost all feature and data split combinations.

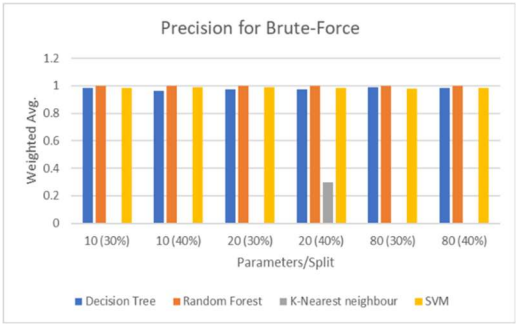

Figure 8: Recall of Brute-Force Attacks.

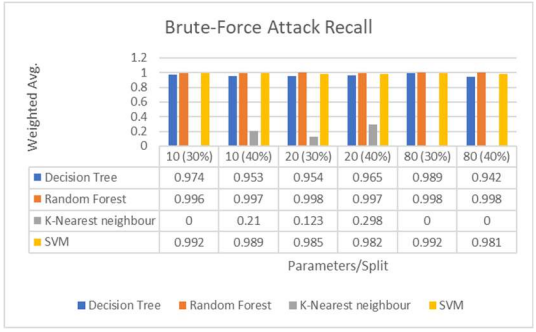

Figure 9: Precision of Brute-Force Attacks

Journal of Theoretical and Applied Information Technology
31st August 2026. Vol.104. No.16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The recall, precision and accuracy of Fig. 8, 9 and 10 indicate that the metrics of all feature and data split combinations are significantly greater. The recall, precision, and f-score value of KNN is, however, astronomically low (0 in many cases). It is clear that the recall, precision and accuracy of the SVM and Random Forest algorithms are more than 99 percent in using 10, 20, and 80 features. The algorithms reveal the optimum performance based on 10 features, 20 features and 80 features, as shown in Figures 8,9 and 10.

Table 3 Model built time (seconds) of bruteforce attack

Features	10		20		80	
Split %	30 %	40%	30%	40%	30%	40%
Decision Tree	11.7	12.99	24.43	23.34	87.43	95.54
Random Forest	57.22	62.29	63.34	83.32	94.45	94.43
K-Nearest Neighbour	178.24	289.24	284.44	394.43	609.43	784.45
SVM	10.29	9.98	25.34	27.44	50.45	67.45

Table 4. Min-Max Normalization of Data Splits and Features of the Brute-Force Attack.

	30% Data Split			40% Data Split		
	10 Features	20 Features	80 Features	10 Features	20 Features	80 features
Decision Tree	0	0.09563	1	0	0.07355	1
Random Forest	0	0.05952	1	0	0.09575	1
K-Nearest Neighbour	0	0.06751	1	0	0.04584	1
SVM	0	0.14647	1	0	0.13353	1

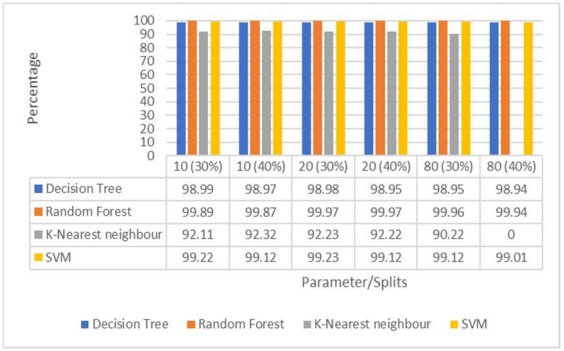

Figure 10: Accuracy of Brute-Force attack.

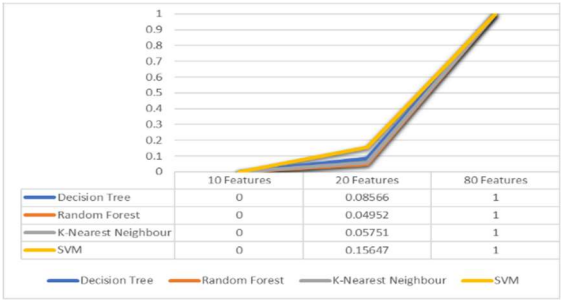

Figure 11: Time for Min-Max Normalization with a 30% Split of Dataset of Brute-Force Attack

The algorithms demonstrate optimal performance with 10 features, followed by 20 and 80 features, as illustrated in Figs. 10 and 11. Therefore, when developing a machine learning model to identify brute-force assaults, it is advantageous to prioritize 10 features over 20 or 80 features, taking into account a variety of metrics, such as performance metrics. The max-min normalization graphs are shown of all ML algorithms are comparable for both 30% and 40% of the data

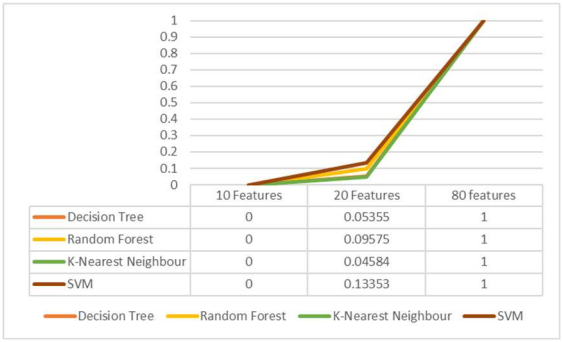

Figure 12: Time for Min-Max Normalization with a 40% Split of Dataset of Brute-Force Attack.

division, as evidenced by figure 11 and figure12. Consequently, the proposed authentication solution allows for the selection of an ML algorithm model that is generated using either the SVM or Random Forest algorithm and takes into account ten features in order to identify brute-force attacks.

5. CONCLUSION

This study presents a novel context-aware machine learning framework that strengthens conventional authentication by combining user and device contextual information with anomaly and attack detection. Its key contribution is the integration of contextual analysis with ML classifiers to identify DoS/DDoS and brute-force

attacks using the CIC-IDS2018 dataset. Among the evaluated models, Random Forest and SVM demonstrated strong detection performance, while feature selection helped reduce computational requirements. The findings demonstrate the potential of the proposed approach to provide an adaptive and additional layer of protection against evolving cyber threats. Future work will focus on evaluating diverse attack types, datasets, contextual parameters, and real-time network environments.Random forest model with 10 feature selection and 30% split data is achieved highest accuracy when to DoS/DDoS. The results and experiments presented that the Random Forest classifier and the support vector machine classifier using 20 features with 40% split getting more accuracy than other ML models in identifying brute-force attacks. By analyzing contextual information in addition to user credentials, the proposed framework can support more adaptive authentication decisions and help identify potentially unauthorized or abnormal access attempts. The reduction in the number of features while maintaining effective classification performance also indicates the potential for developing more efficient security solutions. However, the findings should be interpreted within the scope of the CIC-IDS2018 dataset and the attack categories considered in this study.

6. CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

7. AUTHOR CONTRIBUTIONS

Kalyankumar Dasari: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Writing – Original Draft.
Dr. K. Sahadevaiah: Supervision, Project Administration, Validation, Writing – Review & Editing.

8. ACKNOWLEDGMENTS

The authors would like to thank the Department of Computer Science & Engineering, JNTUK, Andhra Pradesh, India, for providing the research facilities and resources that supported this work.

REFERENCES:

[1] J. Kim, J. Kim, H. L. T. Thu and H. Kim, "Long Short Term Memory Recurrent Neural Network Classifierfor Intrusion Detection," in IEEE International Conference on Platform Technology and Service (PlatCon),2016.

[2] Buczak AL, Guven E (2016) A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Surv Tutor 18(2):1153–1176. https:// doi. org/ 10. 1109/ COMST. 2015. 24945 02

[3] Kim, Aechan& Park, Mohyun& Lee, Dong. (2020). AI-IDS: Application of Deep Learning to Real-Time Web Intrusion Detection. IEEE Access. 8. 70245-70261. 10.1109/ACCESS.2020.2986882. Desai A, Rai S (2012) Analysis of machine learning algorithms using WEKA. Int J Comput Appl ICWET 2012(13):32–37

[4] Perera C, Zaslavsky A, Christen P, Georgakopoulos D (2014) Context aware computing for the internet of things: a survey. IEEE Commun Surv Tutor 16(1):414–454. https:// doi. org/ 10. 1109/ SURV. 2013. 042313. 00197

[5] Gheisari, Mehdi & Pham, Viet &Alazab, Mamoun & Zhang, Xiaobo & Fernández-Campusano, Christian & Srivastava, Gautam. (2019). ECA: An Edge Computing Architecture for Privacy-Preserving in IoT-Based Smart City. IEEE Access. 7. 155779-155786. 10.1109/ACCESS.2019.2937177.

[6] Tao M, Zuo J (2018) Multi-layer cloud architectural model and ontology-based security service framework for IoT-based smart homes. Future Gener ComputSyst 78(3):1040–1051. https:// doi. org/ 10. 1016/j. future. 2016. 11. 011

[7] Kumar, Pardeep & Lee, Hoon-Jae. (2011). Cryptanalysis on two user authentication protocols using smart card for wireless sensor networks. 2011 Wireless Advanced, WiAd 2011. 10.1109/WiAd.2011.5983262.

[8] Aparicio-Navarro FJ, Kyriakopoulos KG, Ghafir I, Lambotharan S, Chambers JA (2018) Multi stage attack detection using contextual information. In: IEEE military communications confer ence, USA

[9] Karabatis, George & Wang, Jianwu&Aleroud, Ahmed. (2016). Towards Adaptive Big Data Cyber-attack Detection via Semantic Link Networks.

[10] Melo W, Machado R, Carmo L (2018) Using physical context-based authentication against external attacks: models and protocols. Secur Commun Netw 2018:1–14. https:// doi. org/ 10. 1155/ 2018/ 65909 28

[11] de-Marcos, Luis & Martínez, José & Junquera Sánchez, Javier &Cilleruelo Rodríguez, Carlos & Pages-Arévalo, Carmen. (2021). Comparing Machine Learning Classifiers for Continuous Authentication on Mobile Devices by Keystroke Dynamics. Electronics. 10. 1622. 10.3390/electronics10141622.

[12] Ackerson JM, Dave R, Seliya N (2021) Applications of recurrent neural network for biometric authentication & anomaly detection. Information. https:// doi. org/ 10. 3390/ info1 20702 72

[13] Samet, Saeed &Ishraque, Mohd Tazim &Ghadamyari, Mehdi &Kakadiya, Krishna & Mistry, Yash &Nakkabi, Youssef. (2019). TouchMetric: a machine learning based continuous authentication feature testing mobile application. International Journal of Information Technology. 11. 625-631. 10.1007/s41870-019-00306-w.

[14] Pryor L, Dave R, Seliya J, Boone ES (2021) Machine learning algorithms in user authentication schemes. In: IEEE international conference on electrical, computer and energy technologies, South Africa

[15] Al-Naymat, Ghazi & Alkasassbeh, Mouhammd& Hawari, Eshraq. (2018). Using machine learning methods for detecting network anomalies within SNMP-MIB dataset. International Journal of Wireless and Mobile Computing. 15. 67-76.

[16] Saini PS, Behal S, Bhatia S (2020) Detection of DDoS attacks using machine learning algo rithms. In: IEEE 7th international conference on computing for sustainable global development, India

[17] Obeidat, Ibrahim &Hamadneh, Nabhan & Alkasassbeh, Mouhammd&Almseidin, Mohammad & Alzubi, Mazen. (2019). Intensive Pre-Processing of KDD Cup 99 for Network Intrusion Classification Using Machine Learning Techniques. International Journal of Interactive Mobile Technologies (iJIM). 13. 70. 10.3991/ijim.v13i01.9679.

[18] Kurniabudi, Stiawan D, Darmawijoyo et al (2020) CICIDS-2017 dataset feature analysis with information gain for anomaly detection. IEEE Access. https:// doi. org/ 10. 1109/ ACCESS. 2020. 30098 43

[19] Alsaif, Omar &Shhatha, Amer. (2025). ENHANCING CYBERSECURITY THROUGH MALWARE DETECTION BASED ON MACHINE LEARNING TECHNIQUE. Kufa Journal of Engineering. 16. 82-100. 10.30572/2018/KJE/160306.

[20] Wanjau S, Wambugu G, Kamau G (2021) SSH-brute force attack detection model based on deep learning. Int J Comput Appl Technol Res. https:// doi. org/ 10. 7753/ IJCAT R1001. 1008

[21] Saito, Satomi &Maruhashi, Koji & Takenaka, Masahiko & Torii, Satoru. (2016). TOPASE: Detection and Prevention of Brute Force Attacks with Disciplined IPs from IDS Logs. Journal of Information Processing. 24. 217-226. 10.2197/ipsjjip.24.217.

[22] Agghey AZ, Mwinuka LJ, Pandhare SM, Dida MA, Ndibwile JD (2021) Detection of username enumeration attack on SSH protocol: machine learning approach. Symmetry. https:// doi. org/ 10. 3390/ sym13 112192

[23] Rizk, Frederic & Rizk, Rodrigue & Rizk, Dominick & Rizk, Patrick & Chu, Chee-Hung. (2025). KAN-MID: A Kolmogorov-Arnold Networks-Based Framework for Malicious URL and Intrusion Detection in IoT Systems. IEEE Access. 13. 160855-160873. 10.1109/ACCESS.2025.3605171.

[24] Sharafaldin I, Lashkari AH, Ghorbani AA (2018) Toward generating a new intrusion detection data set and intrusion traffic characterization. In: International conference on information systems security and privacy, Portugal

[25] Ma, Wei & Wang, Xing & Hu, Mingsheng& Zhou, Qinglei. (2021). Machine Learning Empowered Trust Evaluation Method for IoT Devices. IEEE Access. 9. 65066-65077. 10.1109/ACCESS.2021.3076118.

[26] Ahsan M, Gomes R, Chowdhury MM, Nygard KE (2021) Enhancing machine learning prediction in cybersecurity using dynamic feature selector. J Cybersecur Priv 1(1):199–218. https:// doi. org/ 10. 3390/ jcp10 10011

[27] Zhou, Yangming& Liu, Yangguang. (2014). Correlation analysis of performance metrics for classifier. 487-492. 10.1142/9789814619998_0081.