International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

BotNet Prediction in Social Media based on Feature Extraction with Classification using Machine Learning Algorithms

Surendra Singh Choudhary[1], Dr. S. K. Ghosh[2], A Rajesh[3], Badria Sulaiman Alfurhood[4], Suresh Limkar[5], Dr. Jasmeen Gill[6]

Submitted: 04/11/2022 **Accepted:** 02/02/2023

Abstract: Botnet threat detection has been a focus of continuing study. Botnet identification using flow-based features has been successfully accomplished using machine learning (ML) approaches.Flow-based features' main drawbacks are their significant processing expense and partial capture of network communication patterns.This research propose novel technique in BotNet prediction among the authorized social media users by machine learning algorithm feature analysis. Information has been gathered here from users of social media platforms also it has been filtered based on unusual activities. Then this filtered data features has been extracted and classified using KNN-Xception architecture where the malicious activity. An assessment of the experimental data has been done with regards of detection accuracy, RMSE, malicious activity rate, recall, mAP. The suggested method accomplished detection accuracy of 96%, RMSE of 61%, malicious activity rate of 39%, recall of 59%, mAP of 61%.

Keywords*: botnet, Machine Learning, social media users, feature analysis, unusual activities.*

1. Introduction

The group of devices that make up a network are linked together via Internet, as well as the variety of these devices grows daily.Networks of financial as well as commercial institutions are undoubtedly constantly at danger for security breaches, This undermines their reputation in addition to costing billions of dollars in repairs and losses [1].A serious issue is the growing number of people who are impacted by malicious software. Botnets have taken centre stage as the primary issue since they provide one of the greatest risks to monitoring equipment. Their capacity to take over corporate mainframes by breaking into any Internet-connected gadget that utilises a electronic video recorder is the reason for their appeal [2]. The term "botnet" refers to a network of hacked host computers used to carry out hostile actions. Examples of such host devices include desktop computers, smartphones, notebook PCs, and tablets.To control bots and plan harmful assaults, the botmaster needs a C&C channel.With the use of various machine learning algorithms, machine learning technology is employed in the study to identify and banish bots from social media networks. Computer programmes run better thanks to ML technology, which also gives them the capacity to learn.It streamlines the laborious documentation required for data entry. In the research, spam identification is simple and makes precise predictions possible [3].

2. Related Works

Random forest is an ensemble learning technique that is mostly utilised for classification and regression mechanisms, according to [4].During training and class output, the research built a large number of decision trees using decision-making procedures. The invention of the intrusion detection technique and the setup of the honeypot are the foundations of botnet tactics.According to the research, the main factor in the bot's detection is its ability to connect to a botnet [5].As a result, [6] suggested a technique for differentiating between valid domain names and algorithm-generated domain names frequently utilized in botnets utilising character distribution in domain names.Similar to this, [7] classified autonomously produced domain names with genuine domain names using C5.0 decision tree approach as well as Bayesian statistics.The unique detection technique Kopis was proposed by [8] to

[1] *Research Scholar, Civil Engineering, Indian Institute of Technology, Roorkee, India.*
schoudhary1@ce.iitr.ac.in, ORCID 0000-0002-6397-4094
[2] *Professor, Civil Engineering Indian Institute of Technology, Roorkee, India*
sanjay.ghosh@ce.iitr.ac.in, ORCID 0000-0001-7849-9313"
[3] *Professor, Department of CSE, Jain(Deemed-to-be University), Bangalore, India. a.rajesh@jainuniversity.ac.in*
[4] *Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Saudi Arabia. bsalfurhood@pnu.edu.sa*
[5] *Associate Professor, Artificial Intelligence & Data Science, AISSMS Institute of Information Technology, Pune, India*
sureshlimkar@gmail.com
[6] *Associate Professor, Department of CSE, RIMT University, Mandi, Gobindgarh, Punjab, India.*
jasmeengill@rimt.ac.in, 0000-0002-5894-0551

identify domain names linked to botnets. References [9] discussed two methods that fall under the categories of intrusion detection systems and honeynets.IDSs are then be classified into two categories: frameworks that rely on signatures and systems that rely on anomalies.Host-based or network-based techniques are additional categories for anomaly-based detection techniques.Honeynets, IDS, data mining techniques, and DNS-based methods are the four broad categories into which other studies have split botnet detection strategies.However, the authors in [10] divided botnet detection methods into supervised and unsupervised categories based on type of machine learning used.

3. System Model

This study describes cutting-edge methods in BotNet prediction among the authorized social media users by machine learning algorithm feature analysis. Information has been gathered here from users of social media platforms also it has been filtered based on unusual activities. Then this filtered data features has been extracted and classified using KNN-Xception architecture where the malicious activity.

Preprocessing:

Dimensional reduction is used to preprocess the data. An Important technique for preprocessing streaming data and large scale classification tasks is dimensionality reduction. Classifiers's efficiency and performance is improved by dimensionality reduction. Gaussian filtering F(x), given as F_g (x), is given as eq. (1):

$$F_g(x) = F(x) + \frac{F''(x)}{2}\sigma^2 + \cdots . + \frac{F^{(2m)}(x)}{\prod_{p=1}^{m} 2p}\sigma^{2m} + \cdots \quad (1)$$

{m = 1,2, ..., F"(x) = second derivative of F(x), and F'^(2m)(x) = 2mth order derivatives of F(x)}

KNN-Xception architecture based feature extraction with classification:

The Xception framework is primarily relied upon the ability to separate the cross-channel correlation as well as spatial correlation processes of the convolutional neural network processing feature map, after which the Inception module is supplanted by the deep volume convolution module, meaning that Each input channel's input channel has its own independent input channel. Figure 1 illustrates the process of performing spatial convolution, point-by-point convolution with a 1 1 size filter, and channel output mapping to newly created channel space.

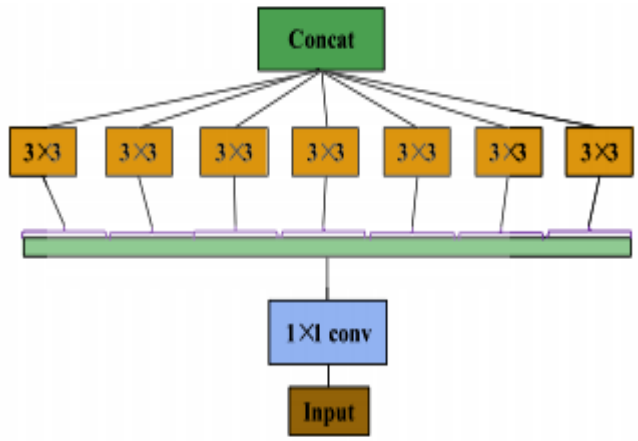

Fig. 1 Xception module

With the same number of parameters, the Xception model outperforms Inception in terms of recognition, demonstrating that the model's advancements are to blame for the performance gain. Various feature maps are produced throughout the process, as illustrated in Figure 1, moreover, individual graph may be merged to convolve different values eq (2):

$$x_j^l = f(u_j^l)$$
$$u_j^l = \sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^b \quad (2)$$

where x_j^l is the output of the l convolution channel of the j layer convolutional layer, and u_j^l is the net activation of the l convolutional channel of the j layer convolutional layer, which is convolved then offset by the preceding held Gx_j^{l-1}. The connecting layer via eq (3) should then apply the aforementioned two-dimensional feature map as its input:

$$x^l = f(u^l)$$
$$u^l = w^l x^{l-1} + b^l \quad (3)$$

where w^l is the weight coefficient of the fully connected layer of layer l, and b^l is the threshold offset term. The training error reduction via eq (4) is then directed in a different direction using the gradient descent approach:

$$\delta^l = \frac{\partial E}{\partial u^l} \quad (4)$$

where δ^l symbolises the function of the squared error with a shift in u^l, and the square of the difference between the intended output and the real output is indicated by E..The class of the new instance is then calculated among the k nearest neighbors according to the most common class. It is important to choose the value of k a priori; different techniques such as cross-validation and heuristics have been proposed to choose it. To prevent tie votes, this value should not be a multiple of the number of groups. KNN's output also largely depends on the measure used to determine the distances between the events. A classification that links the class Y with the qualities j x (1 j n) is conceivable using the

KNN technique since it is non-parametric, which implies that the algorithm may be used without making any assumptions concerning the function. Large values of k often minimising the impact of noise on classification while obscuring class borders. The KNN accounts for proximity by assuming that similar items happen. In other words, similar things are in close proximity to one another.

Algorithm of BotNet detection:

1: dataProcessing (dataset)
2: unitToDrop← 25%
3: Parse data to predefined format
4: Define token dictionary
5: repeat
6. / Parse data to format" /
7. for row ← 1, rotes do
8: Convert text to tokenised integer format

9: Index tokenised text

10. Create dictionary of tokenised text ind

11: Pad data arrays with 0 s to max 25

12: Inject additional tokenised features into

13: end for

14: until return dataset

15: Split Training and Test based on unitToDrop

16. TrainAndValidate

17: TrainAndValidate (trainingData, testData)

18: 18adel ← –

20: loss ← mae

21: optimiser ← Adam

22: epochs ← 100

23: Create new BLSTM/L.STM unit

24: Add LSTM unit to model

25: Develop new Dense Layer

28: Compile model using Optimiser and Loss

29: repeat

30: /FFitModel[4]

32: Calculate Loss, Validation Loss

33: Calculate Accuracy and Validation Accuracy

35: until All epochs completed

36: Return Loss, ValLoss, Acc. ValAce

4. Performance Analysis:

Our strategy is being implemented using Python-based software. Several potent Python libraries are used in our experiment (Pandas, NumPy, scikit-learn, and Matplotlib).

Dataset description:Cybersecurity breaches are frequently linked to the harmful activity of devices infected with malware that creates botnets. We want to create a machine learning model that can discriminate between the bad actions of several botnet families. This study makes use of two datasets with actual botnet traffic: CTU-13 and IoT-23.

Table-1 Analysis of the suggested method in comparison to the current method

Parameter	IDS	DNS	BotNet_FE_MLA
Detection Accuracy	91	93	96
RMSE	55	58	61
Malicious activity rate	32	35	39
Recall	45	52	59
mAP	49	56	61

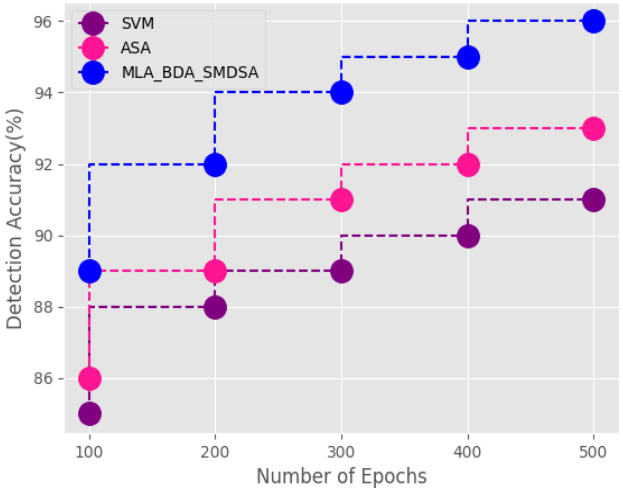

Fig. 2 Comparison of detection accuracy

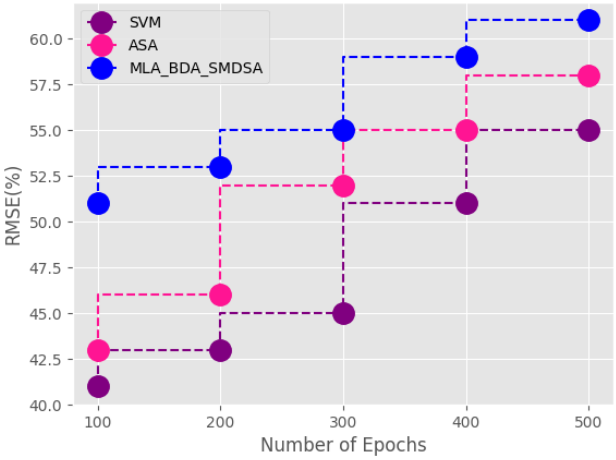

Fig. 3 Comparison of detection RMSE

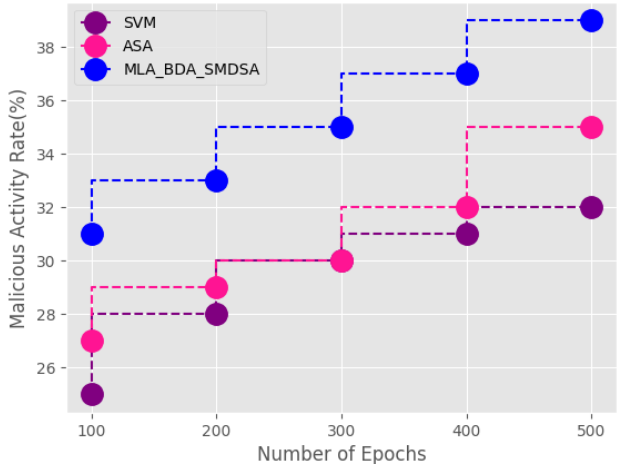

Fig. 4 Comparison of detection malicious activity rate

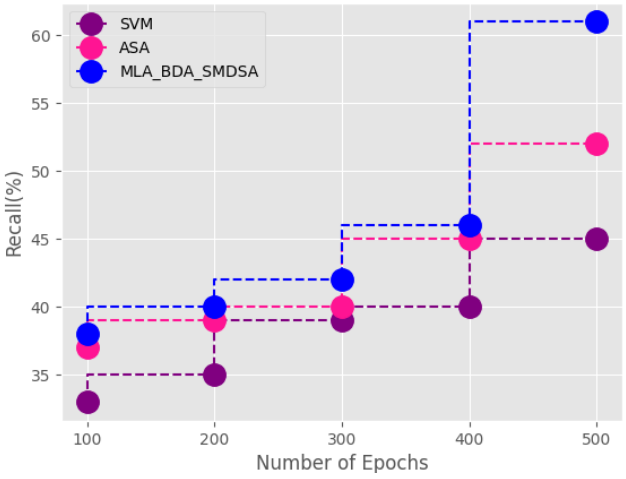

Fig. 5 Comparison of detection recall

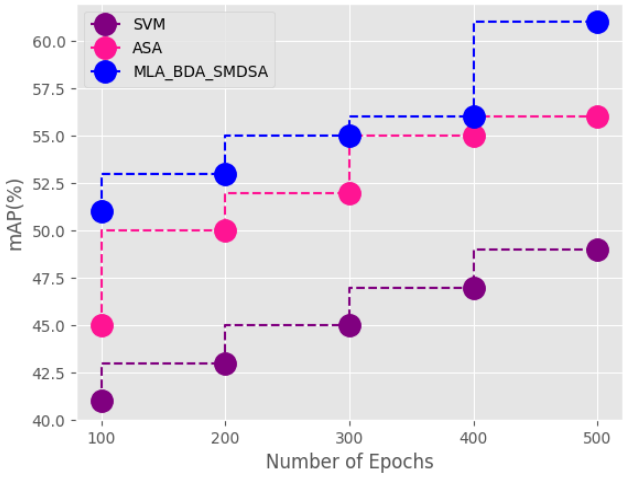

Fig. 6 Comparison of detection mAP

The above table-1 shows comparison of both the suggested and current methods predicated upon BotNet Detection using machine learning strategy. The parametric evaluation in this instance has been completed in perspective of detection accuracy, RMSE, malicious activity rate, recall, mAP. The recommended approach accomplished detection accuracy of 96%, RMSE of 61%, malicious activity rate of 39%, recall of 59%, mAP of 61%.

5. Conclusion

This research proposed novel technique in BotNet prediction among the authorized social media users by machine learning algorithm feature analysis. Here the input has been filtered and its features has been extracted and classified using KNN-Xception architecture where the malicious activity.Gaussian filtering removes noise in the data. But Gaussian filter also distorts the signal. Edge position displacement, edges vanishing, also phantom edges are raised because of using Gaussian filtering as preprocessing for edge detections.The proposed technique attained detection accuracy of 96%, RMSE of 61%, malicious activity rate of 39%, recall of 59%, mAP of 61%.

References

[1] McDermott, C. D., Majdani, F., &Petrovski, A. V. (2018, July). Botnet detection in the internet of things using deep learning approaches. In *2018 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

[2] Ibrahim, W. N. H., Anuar, S., Selamat, A., Krejcar, O., Crespo, R. G., Herrera-Viedma, E., & Fujita, H. (2021). Multilayer framework for botnet detection using machine learning algorithms. *IEEE Access*, *9*, 48753-48768.

[3] Sengupta, T., De, S., & Banerjee, I. (2021, July). A closeness centrality based p2p botnet detection approach using deep learning. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.

[4] Lee, S., Abdullah, A., Jhanjhi, N., &Kok, S. (2021). Classification of botnet attacks in IoT smart factory using honeypot combined with machine learning. *PeerJ Computer Science*, *7*, e350.

[5] Kundu, P. P., Truong-Huu, T., Chen, L., Zhou, L., & Teo, S. G. (2022). Detection and Classification of Botnet Traffic using Deep Learning with Model Explanation. *IEEE Transactions on Dependable and Secure Computing*.

[6] Al-Sarem, M., Saeed, F., Alkhammash, E. H., & Alghamdi, N. S. (2021). An Aggregated Mutual Information Based Feature Selection with Machine Learning Methods for Enhancing IoT Botnet Attack Detection. *Sensors*, *22*(1), 185.

[7] Shareena, J., Ramdas, A., & AP, H. (2021). Intrusion detection system for iot botnet attacks using deep learning. *SN Computer Science*, *2*(3), 1-8.

[8] Owen, H., Zarrin, J., & Pour, S. M. (2022). A Survey on Botnets, Issues, Threats, Methods, Detection and

Prevention. *Journal of Cybersecurity and Privacy*, *2*(1), 74-88.

[9] Zamani, A. K., &Chapnevis, A. (2022). BotNet Intrusion Detection System in Internet of Things with Developed Deep Learning. *arXiv preprint arXiv:2207.04503*.

[10] Sarker, I. H. (2021). Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. *SN Computer Science*, *2*(3), 1-16.