DDoS Attack Detection Using Machine Learning and Improved Clustering Algorithm

Fatima R. Hamade, Marwah Habiban, and Ali Abdulkarem Habib Alrammahi*
Department of Computer Science, Faculty of Computer Science and Mathematics, University of Kufa, Najaf, Iraq
Email: fatimar.hamade@uokufa.edu.iq (F.R.H.), marwa.habiban@uokufa.edu.iq (M.H.),
alia.alramahi@uokufa.edu.iq (A.A.H.A.)

Manuscript received September 4, 2024; revised November 17, 2024; accepted December 24, 2024
*Corresponding author

***Abstract*—Distributed Denial of Service (DDoS) attacks have recently emerged as one of the most destructive threats to network systems. This paper aims to develop a technique that efficiently identifies DDoS attacks in networked systems by leveraging improved clustering techniques and machine learning algorithms. This methodology employs a Modified Fuzzy C-Means (MFCM) clustering algorithm to partition the available DDoS attack dataset and integrate a classification algorithm to accurately detect attacks and classify data based on specific network characteristics derived from the transformed data packets. The clustering algorithm predominantly relies on distance measurements derived from fuzzy coefficients, significantly limiting its ability to identify and classify emerging attack scenarios. The current study introduces the integration of the MFCM clustering algorithm with sophisticated classification techniques to enhance accuracy and minimize errors. The efficacy of the modified clustering algorithm was evaluated using the entropy criterion, and a value of 0.99 was attained, demonstrating superior performance relative to traditional algorithms. The training algorithm was rigorously evaluated utilizing established performance metrics, such as accuracy, detection rate, and false positive rate. The results indicate that the accuracy improved consistently across all classification algorithms applied, contributing to an enhanced attack detection rate.**

***Index Terms*—ACK/PUSH-ACK, Distributed Denial of Service (DDoS) dataset, standard clustering method, modified Fuzzy C-means, training algorithms and evaluation metrics**

I. Introduction

Distributed Denial of Service (DDoS) assaults represent a critical threat to the security and availability of internet services. These attacks aim to disrupt service by overwhelming the target server with excessive messages or requests, thereby exhausting and causing a shutdown [1].

It is essential to enhance the resilience of existing network services against such attacks to mitigate vulnerabilities that could result in significant financial losses for customers and enterprises. Modern attacks often deviate from traditional motives like financial gain or access to confidential information. Instead, many of these attacks are designed to disrupt services, rendering them inaccessible to users for as long as the attacker desires [2–4]. A single bot master can command numerous bots (zombies) to generate a massive influx of messages, thereby overwhelming network bandwidth and leading to a complete service shutdown in DDoS attacks [5, 6].

Detecting DDoS attacks is a more effective security approach than detecting individual attackers. On the other hand, DDoS attacks evolve alongside the advancements in security measures [7].

Data mining becomes crucial for understanding system-generating data. However, due to the volume and complexity of the data, automatic approaches are often necessary. In cases where data are unlabeled, clustering techniques are indispensable for assigning data to relatively similar groups.

The Combination of Fuzzy C-Means (FCM) [2] and Possibilistic C-Means (PCM) [3] led to the development of Possibilistic Fuzzy C-Means (PFCM). By relaxing the constraint on the typicality values, PFCM has been proposed as a more reliable variant of the Fuzzy Possibilistic C-Means (FPCM) [5]. In this paper, we further enhance the existing algorithm by introducing a mathematical formulation to calculate the membership function, resulting in a new version termed Modified Fuzzy C-Means (MFCM).

Collecting and labelling data initially, which represents data analysis, enhances the accuracy and efficiency of training machine learning algorithms. Accurate data collection reduces prediction errors and improves algorithms' overall performance.

The proposed clustering algorithm complements traditional classification algorithms by balancing the dataset, allowing for better extraction of target information. Each cluster represents an attack class, whether malicious or benign, thereby improving the efficiency of the classification algorithms.

II. Related Works

Researchers have long focused on DDoS attack detection systems. Consequently, many clustering and classification algorithms have been developed and enhanced to improve accuracy or reduce the time required for detection. Previous works related to the topic will be discussed for further analysis.

Kumar and Venu [8] provide a valuable examination of Botnet Distributed Denial of Service (DDoS) attack detection using machine learning technology, notably the K-means method. It is advised to use the UNBS-NB real-

time datasets for experimental analysis. In this experimental study, we compare the performance of K-means algorithms with numerous other machine learning methods, including Decision Trees (DT), Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Naive Bayes (NB).

Madathi *et al.* [9] presented a classification algorithm to control DDoS attacks in Software-Defined Networking (SDN), particularly suited for high-frequency applications. The algorithm prevents unauthorized users from dealing with the control unit. They used the K-Nearest Neighbor (KNN) machine learning algorithm to process the DDoS detection dataset and predict whether incoming or regular requests are DDoS attacks. The KNN algorithm achieved 96% accuracy on the input dataset, which contains 17 columns and nearly 1000 rows.

Mandala *et al.* [10] proposed a predictive system to detect DDoS attacks using the classification data mining methods, using the Random Forest Classifier (RFC) to create the model. The system was applied to the modern CICIDS 2018 dataset from the Canadian Institute of Cybersecurity for Intrusion Detection System. Performance analysis confirmed that the model achieved a detection accuracy of 93% without tuning, which improved to 97.2% with the tuning process.

Maslan *et al.* [11] used data gathered from DDoS attack events to do a study employing machine learning techniques for DDoS attack detection. The researchers used Five machine learning techniques to train and evaluate the dataset following feature selection: Naive Bayes (NB), RFC, Neural Network (NN), SVM, and KNN. The study compared these techniques and demonstrated that the Random Forest Classifier achieved the highest accuracy, nearly 98.4 %.

Bindra and Sood [12] introduced DDoS attack detection using the CICIDS 2017 dataset, concluding that supervised machine learning algorithms are more effective for network classification. After pre-processing the data, the RFC outperformed other classifiers in terms of accuracy but required a longer time. The performance evaluation, conducted by accuracy and Receiver Operating Characteristic (ROC) metric, showed that RFC consistently provided better results compared to other classifiers like Logistic Regression (LR), KNN, Gaussian NB, and Linear SVM, achieving accuracy rates exceeding 96%.

Khanbabaei *et al.* [13] proposed a new model leveraging data mining techniques for clustering and classification to identify competitive, knowledge-intensive behaviors within large datasets. The model uses K-means clustering for data grouping and the decision tree algorithm for classification, enabling the extraction of hidden patterns from high-volume data. The results demonstrate that the proposed model effectively enhances competitive and knowledge-intensive processes, showcasing the potential of data mining methods to improve performance in large-scale processes.

Bista and Chitrakar [14] constructed a system for detecting DDoS attacks that integrates a clustering algorithm and classification methods. The proposed algorithm uses heuristics clustering and the NB classification method. Performance evaluation metrics were used to evaluate the system, including accuracy, true positive rate, and false positive rate. The system was tested on two datasets: the first dataset is the Center for Applied Internet Data Analysis (CAIDA), a research unit at the University of California San Diego 2007 dataset, and the second dataset is the Defense Advanced Research Projects Agency (DARPA) 2000 dataset.

Most prior studies have focused on classification algorithms and used either the (K-means) or (heuristic) as clustering algorithms. In contrast, the current research introduces a modified clustering method that has not been used previously, aiming to achieve better results than earlier methodologies. A comprehensive summary of the aforementioned prior works is presented in Table I.

TABLE I: SUMMARY OF PREVIOUS WORKS

No	Paper info.	Year	Algorithms	Accuracy
1	A. A. Kumar and N. Venu [7]	2023	Mathematical model + SVM+ NB	97.2%
2	Madathi *et al*. [8]	2022	KNN	96%
3	S. Mandala *et al*. [9]	2022	RFC	97.2%.
4	A. Maslan *et al*. [10]	2020	RFC+SVM+KNN	98.4 %
5	Bindra and Sood [11]	2019	logistic regression, KNN, Gaussian NB and linear SVM + RFC	96%
6	Khanbabaei *et al.* [12]	2019	K-means clustering method with classification algorithms	96.2%
7	S. Bista and R. Chitrakar [13]	2018	clustering algorithm+ NB	97.5%.

III. PROPOSED METHOD

The proposed system can be illustrated through the general diagram shown in Fig. 1. The process begins with a dataset of network traffic information. Nine features were selected as most relevant for the clustering and classification process. Following this, the system transitions to the clustering and classification stage, which will be elaborated on in detail in the following sections.

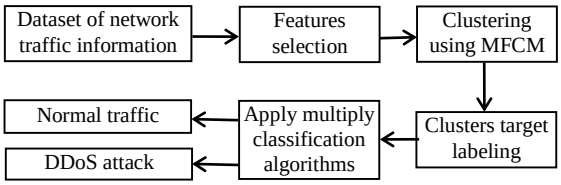

Fig. 1. Block diagram of proposed DDoS attack detection using machine learning and improved clustering algorithm.

A. Clustering using Modified Fuzzy C-means (MFCM)

In this paper, which aims to classify attacks, the initial step in the training process is to use clustering technology to identify the target classes on which the machine learning algorithm will base its training and prediction. One of the factors affecting the accuracy of the clustering process is the choice of distance measures or the calculation of membership functions. Many standard membership functions include the following:

1) Linear membership

The most basic and popular types of membership functions for fuzzy clustering are linear ones. Each data point is assigned a value that decreases linearly in proportion to its distance from the cluster center. The membership value of a data point increases as it gets closer to the center and decreases as it moves farther away. Although linear membership functions are simple and easy to use, they may fail to fully capture the complex and nonlinear relationships between data points and clusters. Fig. 2 shows the representation of linear membership functions [15].

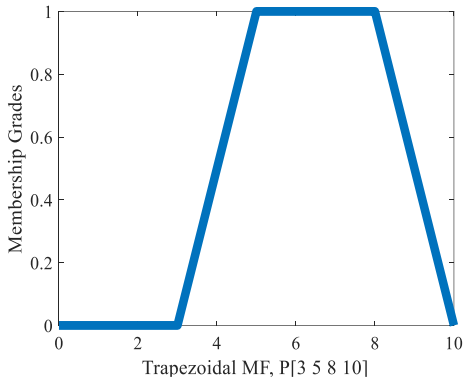

Fig. 2. Linear model for membership function.

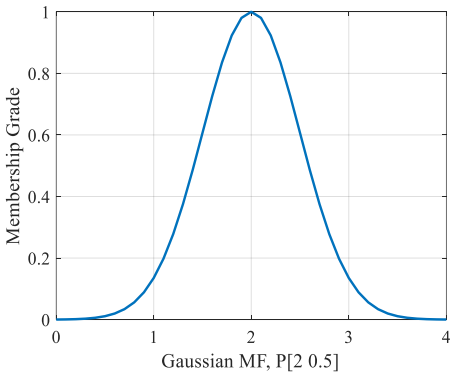

Fig. 3. Gaussian model for membership function

2) Gaussian membership

Another standard membership function for fuzzy clustering is the Gaussian membership function. Each data point is assigned a bell-shaped value based on its distance from the cluster center. The membership value of a data point decreases as the distance from the center increases until it hits zero. Unlike linear membership functions, Gaussian ones are less sensitive to noise and outlier values. They can thus better capture the natural distribution of data points in the vicinity of different clusters, as shown in Fig. 3 [16].

3) Sigmoidal membership

Sigmoidal membership functions in fuzzy clustering assign each data point an S-shaped value calculated based on the distance to the cluster center. The closer the member data point is to the center, the higher its membership value is, and the further it is from the center point, the lower it will be until it reaches some minimum or maximum value. The sigmoidal membership functions can tackle various challenges with overlapping and nested clusters and abrupt changes and transitions among clusters, as shown in Fig. 4 [17, 18].

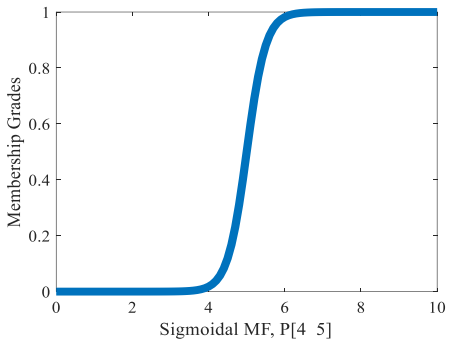

Fig. 4. Sigmoidal model for membership function.

4) Exponential membership

The Generalized Bell-Shaped (GBell) membership function is a fuzzy membership function characterized by a bell-shaped curve. This versatile function is widely used in fuzzy logic systems across various applications. The GBell function is defined by three parameters: a, b, and c, which control the shape, center, and width of the bell curve, respectively, as shown in Fig. 5 [19, 20].

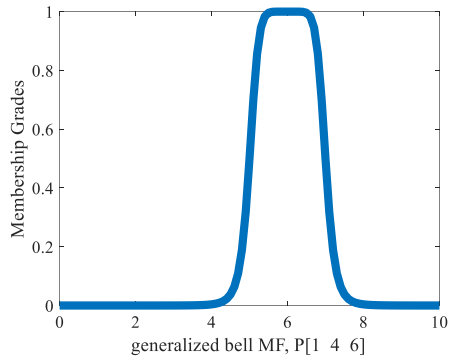

Fig. 5. Exponential model for membership function.

The goals of the investigation and the properties of the data must be taken into account when choosing the most suitable membership function for fuzzy clustering. There is no ideal membership function; different membership functions may provide disparate outcomes and interpretations. Metrics that assess the quality and effectiveness of the clustering can be employed to determine the optimal membership function. These include compactness, which evaluates how tightly the data points are packed around the cluster centers; separation, which considers how perfect clusters are isolated from one another; and validity, which assesses how good clusters reflect the underlying structure and patterns of the data.

This paper proposes the use of logistic regression to define the membership function:

$$\mu_{i,j} = \ln\left(\frac{\mu_{i,j}}{1-\mu_{i,j}}\right) = b_0 + b_1 x_{1i} + \ldots + b_5 x_{5i} + \epsilon \quad (1)$$

where μ_{ij} Is the membership function, $b_0, b_1, \ldots, b_5$ Represent fuzzy association parameters, x_i Is the position

i in the data matrix, and ϵ Denotes the error term. Suppose $b_5 = (\text{pl}, \text{pu}, \text{pm}, \text{pr})$ In equation (1), then the predictable outputs become trapezoidal fuzzy numbers, where pl, pu, pm, and pr represent the left, right middle, left middle, and right points [21].

B. Dataset

The proposed approach uses a DDoS attack dataset with 23 features. The data types include 19 features of type int64 and four features of type object. The memory usage is 26.5+ MB for 151200 entries, as detailed in Table II [22].

TABLE II: DATASET FEATURES

Variable No.	Feature name	Data type	Variable No.	Feature name	Data type
1	ip.src	object	12	ip.flags.rb	int64
2	ip.dst	object	13	tcp.seq	int64
3	tcp.srcport	int64	14	tcp.ack	int64
4	tcp.dstport	int64	15	frame.time	Object
5	ip.proto	int64	16	Packets	int64
6	frame.len	int64	17	Bytes	int64
7	tcp.flags.syn	int64	18	Tx Packets	int64
8	tcp.flags.reset	int64	19	Tx Bytes	int64
9	tcp.flags.push	int64	20	Rx Packets	int64
10	tcp.flags.ack	int64	21	Rx Bytes	int64
11	ip.flags.mf	int64	22	Label	Object

The data are imported from the DDoS attack dataset using Panda's library and then saved using Panda's data frame. The resulting data frame includes some null values handled and removed during the pre-processing step. The Nine essential features were selected during the feature selection step, including ip.src, ip.dst, tcp.srcport, tcp.dstport, ip.proto, frame.len, packets and bytes. These features represent various attributes such as ip- source address, ip- destination address, tcp- protocol source port, tcp- protocol destination port, ip- protocol number, frame length, number of sent packets, and length of the message by byte, respectively. It is worth noting that the selection process was based on the relationships between these features.

C. Data Classification

In classification algorithms, a clustering step comes before the classification process to improve the guessing process's speed and accuracy. Along with the suggested technique, MFCM, several important clustering algorithms, such as fuzzy c-means, heuristic clustering, and hierarchical clustering, were selected based on the examination of earlier research. The clustering process establishes a dependency relationship between the specified features, thereby improving the classification algorithms' efficiency in balancing datasets. The next step involves feeding the classification algorithms with the clusters obtained from the proposed clustering process and labelling the clusters based on the previously specified features. A set of classification algorithms that depend on the probabilistic relationship between the dataset's features were chosen, including Naïve Bayes, KNN and decision tree [23–25].

IV. RESULTS ANALYSIS

The evaluation of the clustering process involves selecting the results from one of the standard clustering methods and comparing them with the proposed method to feed the generated clusters into the classification algorithms. The entropy coefficient is used to evaluate the accuracy of the clustering algorithm given by

$$H_C(P) = \frac{1}{N}\sum_{l=1}^{c}\sum_{i=1}^{n}\left|\mu_{l,i}.\ln\mu_{l,i}\right| \quad (2)$$

where $H_c(P)$ Is entropy coefficient, N is the number of items for the clustered set, c is the number of clusters, and μ_{li} It is the membership function.

These indicators have the following properties:

- In case when the resulting partition is the most uncertain, that is, $\mu_{l,i}$ for all i=1, 2, ⋯, n and l= 1, 2, ⋯, c. The exponents take the values $H_c(P)$=ln$c$.
- The range of the valued entropy of the partition by the inequality $0 \le H_c(P) \le \ln c$.

The performance of clustering algorithms is according to the entropy coefficient, as shown in Table III.

TABLE III: EVALUATION OF CLUSTERING ALGORITHMS

Clustering algorithm	Entropy measure
Heuristic clustering	0.3479
hierarchical clustering	0.8711
Fuzzy C-means	0.9928
Modified fuzzy C-means	0.9928

According to the entropy coefficient in Table III, the modified algorithm is the best for the clustering process, as it becomes more efficient when the entropy coefficient is closer to 1.

The accuracy, precision, true and false positive rates are used to evaluate the performance of the classification models using metrics called the confusion matrix, as shown in Table IV [26].

TABLE IV: CONFUSION MATRIX

TN (True negative)	FP (False positive)
FN (False negative)	TP (True positive)

The confusion matrix represents the correct predictions and incorrect predictions, given as

$$\text{Correct predictions} = TN + TP \quad (3)$$

$$\text{Incorrect predictions} = FP + FN \quad (4)$$

Another way to evaluate the performance of the classification models is through the classification report, which provides metrics such as accuracy, precision, and F1-score. The accuracy of the models can be calculated as given in (5):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Precision is defined as the percentage of truly predicted positive class samples among all predicated positive samples by the classifier and calculated as given in (6):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

The recall represents the true positive rate, which means the proportion of samples that are truly predicted out of all the results predicted to be positive, and the recall is calculated as given in (7):

$$\text{Recall}=\frac{TP}{TP+FN} \quad (7)$$

F1-score combines both precision and recall into a single value, where the best value is 1.0, and the worst value is 0.0 and is calculated as given in (8):

$$\text{F1-SCORE}=2\frac{Precision\times Recall}{Precision + Recall} \quad (8)$$

Another measure to evaluate the performance of classification is the F-beta Score, along with the false positive rate, which can be calculated as (9) and (10) respectively:

$$\text{F-beta Score}=(1+B^2)\frac{\text{Precision}\times\text{Recall}}{\text{Precision}\times B^2+\text{Recall}} \quad (9)$$

$$\text{False positive rate (FPR)}=\frac{FP}{TN+FP} \quad (10)$$

As mentioned earlier, the paper's main objective is to establish a relationship between features to extract the target feature that will be used for training. All machine learning algorithms chosen to test the efficiency of the proposed method were evaluated in two ways.

1) Implement a machine learning algorithm without clustering

The straightforward approach of training algorithms on the original dataset, where eight features were chosen, and the ninth represents the target feature, the basis for training. It is worth noting that machine learning algorithms cannot process textual data, so the address feature data was converted into two values (0, 1) to predict the presence of an attack. The final accuracy is shown in Table V.

TABLE V: EVOLUTION OF IMPLEMENT MACHINE LEARNING ALGORITHM WITHOUT CLUSTERING

Algorithm	Accuracy	precision	Recall	F1-score	F-beta Score
NB	0.7710	0.6958	0.6944	0.6902	0.6942
KNN, k = 2	0.8557	0.80533	0.8053	0.8053	0.8053
Decision Tree	0.8500	0.32704	0.3270	0.3270	0.3270
AdaBoost	0.8701	0. 83513	0.8135	0.8153	0.8121

2) Implement machine learning algorithms with clustering

As in the first method, eight features will be selected from the original dataset, while the target feature and balanced dataset will be processed using the proposed clustering method. Standard algorithms will be employed to compare the performance of classification algorithms as given in Table VI to Table VIII.

Based on the results obtained from using the clustering algorithm (MFCM) with the classification algorithms, the proposed algorithm is superior to the others, as shown in Table VIII.

When comparing the accuracy results from Table IV with the results of Table VIII, it is evident that the classification algorithms' performance significantly improved after utilizing the modified clustering algorithm. The classification accuracy with this type of data increased from 0.97% to 0.99%, thereby enhancing the reliability of the proposed method.

TABLE VI: IMPLEMENT MACHINE LEARNING ALGORITHMS WITH HEURISTIC CLUSTERING

Algorithm	Accuracy	precision	Recall	F1-score	F-beta Score
NB	0.8896	0.9730	0.9065	0.9386	0.9071
KNN, k = 2	0.9956	0.9983	0.9968	0.9975	0.99681
Decision Tree	0.9998	1.0	0.9998	0.9999	0.9998
AdaBoost	0.9998	0.9998	1.0	0.9999	0.9999

TABLE VII: IMPLEMENT MACHINE LEARNING ALGORITHMS WITH FUZZY C-MEANS CLUSTERING

Algorithm	Accuracy	precision	Recall	F1-score	F-beta Score
NB	0.9890	0.9871	0.9931	0.99	0.993
KNN, k = 2	0.9689	0.9280	0.9780	0.9590	0.9680
Decision Tree	0.9998	1.0	0.9997	0.9998	0.9999
AdaBoost	0.9996	0.9994	1.0	0.9997	0.9999

TABLE VIII: IMPLEMENT MACHINE LEARNING ALGORITHMS WITH MFCM

Algorithm	Accuracy	precision	Recall	F1-score	F-beta Score
NB	0.9907	0.9853	0.9828	0.9840	0.9828
KNN, k = 2	0.9989	1.0	0.9980	0.9990	0.9980
Decision Tree	0.9992	0.9979	0.9994	0.9986	0.9994
AdaBoost	0.9998	0.9998	1.0	0.9999	0.9999

V. CONCLUSION

In this paper, machine learning algorithms were applied to detect DDoS attacks, both with and without clustering techniques. The entropy coefficient was used to evaluate the performance of clustering techniques, with the proposed MFCM clustering method demonstrating superior results. Combining clustering techniques with classification algorithms enhances predictive accuracy, particularly for algorithms that rely on the relationship between features. The accuracy rate 0.99 reported in Table VIII represents a significant improvement, surpassing previous methods like K-means, which achieved a maximum accuracy of 0.97.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Conceptualisation, Ali Abdulkarem Habib Alrammahi; the methodology, Ali Abdulkarem Habib Alrammahi; software, Fatima R. Hamade and Marwah Habiban; wrote the manuscript, Fatima R. Hamade and Marwah Habiban; writing review and editing, Ali Abdulkarem Habib Alrammahi; proofread the paper, Ali Abdulkarem Habib Alrammahi. All authors had approved the final version.

REFERENCES

[1] M. A. Al-Shareeda, S. Manickam and M. Ali, "DDoS attacks detection using machine learning and deep learning techniques: Analysis and comparison," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 930–939, Apr. 2023 .

[2] R. R. Papalkar and A. S. Alvi, "Analysis of defense techniques for DDos attacks in IoT—A review," *ECS Trans*., vol. 107, no. 1, pp. 3061–3068, Apr. 2022 .

[3] Z.-X. Ji, Q.-S. Sun, and D.-S. Xia, "A modified possibilistic fuzzy c-means clustering algorithm for bias field estimation and segmentation of brain MR image," *Computerized Medical Imaging and Graphics*, vol. 35, no. 5, pp. 383–397, 2011.

[4] H. Alyasiri, J. A. Clark, and D. Kudenko, "Applying cartesian genetic programming to evolve rules for intrusion detection system," in *Proc of the 10th Int. Joint Conf. on Computational Intelligence*, 2018, pp. 176–183.

[5] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 517–530, 2005.

[6] A. A. Zaidan, B. B. Zaidan, M. Y. Qahtan *et al.*, "A survey on communication components for IoT-based technologies in smart homes," *Elecommun. Syst.*, vol. 69, pp. 1–25, Mar. 2018 .

[7] F. Musumeci, A. C. Fidanci, F. Paolucci, F. Cugini and M. Tornatore, "Machine-learning-enabled DDoS attacks detection in P4 programmable networks," *Journal of Network and Systems Management*, vol. 30, no. 1, pp. 1–27, Jan. 2022.

[8] A. A. Kumar and N Venu, "Machine learning algorithms' significance in identifying botnet DDoS," *High Technology Letters*, vol. 29, no. 11, pp. 199–213, 2023 .

[9] M. Madathi, R. Harini, R. Monikaa and N. Gowthami, "Detection of DDoS attack in SDN environment using KNN algorithm," *IJRAR-International Journal of Research and Analytical Reviews*, vol. 9, no. 2, pp. 252–257, Apr. 2022 .

[10] S. Mandala, A. Ramadhan, M. Rosalinda, S. M. Zaki and E. Weippl, "DDoS detection using information gain feature selection and random forest classifier," in *Proc. of 2022 2nd Int. Conf. on Intelligent Cybernetics Technology & Applications*, Dec. 2022 . doi: 10.1109/ICICyTA57421.2022.10038126

[11] A. Maslan, K. M. B. Mohamad and F. B. M. Foozy, "Feature selection for DDoS detection using classification machine learning techniques," *IAES International Journal of Artificial Intelligence*, vol. 9, no. 1, pp. 137–145, Mar. 2020 .

[12] N. Bindra and M. Sood, "Detecting DDoS attacks using machine learning techniques," *Automatic Control and Computer Sciences*, vol. 53, no. 5, pp. 419–428., 2019

[13] M. Khanbabaei, M. Alborzi, F. M. Sobhani and R. Radfar, "Applying clustering and classification data mining techniques for competitive and knowledge-intensive processes improvement," *Knowledge and Process Management*, vol. 26, no. 2, pp. 123–139, 2019 .

[14] S. Bista and R. Chitrakar, "DDoS attack detection using heuristics clustering algorithm and naïve bayes," *Journal of Information Security*, vol. 9, pp. 33–44, 2017. dio: 10.4236/jis.2018.91004

[15] J. Dombi and Z. Gera, "The approximation of piecewise linear membership functions and Łukasiewicz operators," *Fuzzy Sets and Systems*, vol. 154, no. 2, pp. 275–286, 2005 .

[16] I. A. Hameed, "Using Gaussian membership functions for improving the reliability and robustness of students' evaluation systems," *Expert Systems with Applications*, vol. 38, no. 6, pp. 7135–7142, 2011.

[17] S. Gupta, P. K. Biswas, B. Aljafari *et al.*, "Modelling, simulation and performance comparison of different membership functions based fuzzy logic control for an active magnetic bearing system," *The Journal of Engineering*, #e12229, 2023. doi: 10.1049/tje2.12229

[18] K. Kameshwaran and K. Malarvizhi, "Survey on clustering techniques in data mining," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 2, pp. 2272–2276, 2014.

[19] A. A. H. Alrammahi, F. A. O. Sari, and H. A. H. Shamsuldeen, "Analysis of the development of fruit trees diseases using modified analytical model of fuzzy c-means method," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 358–364, 2023.

[20] Y. Gromov, Y. Minin, A. A. H. Alrammahi, and F. A. Sari, "Probabilistic and fuzzy models of the optimal allocation of resources of a network information system," in *Proc. of 2019 1st Int. Conf. on Control Systems, Mathematical Modelling, Automation and Energy Efficiency*, 2019, pp. 353–358.

[21] S. Mustafa, S. Asghar, and M. J. I. J. o. F. S. Hanif, "Fuzzy logistic regression based on least square approach and trapezoidal membership function," *Iranian Journal of Fuzzy Systems*, vol. 15, no. 6, pp. 97–106, 2018.

[22] Y. R. Kumbam, "ACK/PUSH-ACK DDoS," *Kaggle*, 2020.

[23] G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," in *Proc. of 2013 Fourth Int. Conf. on Computing, Communications and Networking Technologies*, July 2013. doi: 10.1109/ICCCNT.2013.6726842

[24] P. K. Vaishnav, S. Sharma and P. Sharma, "Analytical review analysis for screening COVID-19," *International Journal of Modern Research*, vol. 1, no. 1, pp. 22–29, 2021 .

[25] A. A. Al-Rammahi, S. Farah, and F. G. Al-Jelaihawi, "COVID-19 plasma monitoring based on clustering a large set of recovered patient data," in *Proc. of 2020 International Multi-Conference on Industrial Engineering and Modern Technologies*, 2020. doi: 10.1109/FarEastCon50210.2020.9271296

[26] A. Munther, I. Mohammed, M. Anbar and A. Hilal, "Performance evaluation for four supervised classifiers in internet traffic classification," in *Communications in Computer and Information Science*, M. Anbar, N. Abdullah, and S. Manickam, ed., 2020, p. 168–181.

Fatima R. Hamade received a master's degree in computer science from the faculty of computer science and mathematics at the University of Kufa, Iraq. She currently works at the University of Kufa in Najaf, Iraq. Her research interests include information security and data science.

Marwah Habiban received a master's degree in computer science from the faculty of computer science and mathematics at the University of Kufa, Iraq. She currently works at the University of Kufa in Najaf, Iraq. Her research interests include image processing and data science.

Ali Abdulkarem Habib Alarammahi received a master's degree in information technology from Dr Babasaheb Ambedkar Marathwada University in Aurangabad, India, and a PhD degree in data mining from Tambov State Technical University, Russia, in 2022. He currently works at the University of Kufa in Najaf, Iraq. His research interests include data analysis and data science.