Wasit Journal for Pure Science

Journal Homepage: https://wjps.uowasit.edu.iq/index.php/wjps/index
e-ISSN: 2790-5241 p-ISSN: 2790-5233

Hybrid Machine Learning Model for Feature Selection in DDoS Attack Detection in Cloud Environments Using Convolutional Neural Networks and Genetic Algorithms

Sabah Mohammed Fayadh[1]*, Ekhlas Ghaleb Abdulkadhim[2] Elham Mohammed Thabit A. Alsaadi[3]

[1]Southern Technical University, Al Nassriyah Technical Institute, Department of Computer Systems, IRAQ.
[2] University of Kerbala, Collage of Tourism Sciences, Karbala, IRAQ.
[3]University of Kerbala, College of Computer Science and Information Technology, Department of Information Technology, Karbala, IRAQ.

*Corresponding Author: Sabah Mohammed Fayadh

DOI: https://doi.org/10.31185/wjps.616
Received 10 November 2024; Accepted 20 December 2024; Available online 30 March 2025

ABSTRACT: As cloud environments are very vulnerable to such threats and are in focus for sophisticated cyberattacks, including the DDoS attack, there is a rising trend in distributed networks as the popularity of cloud computing increases. These attacks usually employ techniques like botnets, spoofing, and multi-vector attacks, and therefore are making becomes increasingly difficult to detect. This paper also presents an adaptive hybrid AI model which employ CNN with GA to select the features of the best option and the detection of DDoS attack. Thus the GA is applied to optimize the feature selection over network traffic and the CNN is then used to learn and extract the spatial and temporal patterns. Subsequently, for testing, our approach undergoes normalization, dimensionality reduction, and feature extraction before the model is tested on CIC-IDS-2017 and CIC-Darknet-2020 datasets. An AI model is presented that performs better than pure AI models like SVM, random forests, and decision trees and surpasses other hybrid methods like HMM and LSTM with a 99.98% detection rate. Results have shown the efficiency of the suggested model in terms of scalability, reliability and operational ability to work in real time, which can further implied that it is a potential solution for DDoS attack in cloud. In future work, it is planned to enhance this framework to include a wider array of attack methods as well as refine the efficiency of the algorithm for use in real time applications with limited computing capabilities.

Keywords: DDoS Detection, Convolutional Neural Networks, Genetic Algorithm, Cloud Security, Feature Selection, Cybersecurity, Real-Time Detection, Network Intrusion Detection Systems.

1. INTRODUCTION

Today, cloud computing significantly helps the businesses to store and process the immense volume of data. It is as scalable and affordable as it can be for sectors in healthcare, finance [1], entertainment. Despite this wide spread adoption, such clouds are more susceptible to cyber threats, particularly the attacks of Distributed Denial of Service (DDoS), which throws malicious traffic in such an amount that it overloads the system resources and makes it functional in such a way that authorized users are prevented from using them and create huge damage especially on cloud environment [2,3]. However, as DDoS attacks turn their heads, making use of botnets, IP spoofing, and even multivector, the traditional

approaches of defense are finding themselves becoming a has-been of the info age. The scalable nature of cloud resources [4,5] makes it difficult to address these problems through signature based and rule based methods, as they frequently fail to detect novel attacks, accompanied by high false negatives and potentially long delays in detection. To overcome this DDoS detection is becoming a trend with Artificial Intelligence (AI) and Machine Learning (ML) [6]. Even though traditional ML model, i.e., SVMs and Random Forests; however, Lstm and Convolvenet neural network have shown better results compared to traditional ML methods due to their capability to learn the discerning patterns in the network traffic. Real time detection however remains impractical for network traffic data with high dimensionality [7,8]. In this study, an adaptive hybrid AI model of CNNs and GA is proposed as a better DDoS detection in cloud environments. By dimensionally reducing computational overhead and improving detection accuracy, the model can provide real-time performance across large scale networks [9]. The focus of this research is an efficient DDoS detection and prevention system which could be extended to handle other broader cyber threats in the cloud environments [10].

DDoS attacks are a leading attack vector to cloud network security and can lead to service disruption and financial and reputational damage [11]. What has made cloud services more vulnerable to such attacks are their increasing scalability and increasing sophistication; bots, IP spoofing, multi-vector attacks [12,13]. Traditional DDoS detection methods, including rule based systems and signature matching, are not adequate in identifying these emerging threats [14]. High false positive rates and easy to bypass novel attack patterns make these methods fail.

Machine Learning for DDoS Detection

Historical data and learning patterns in network traffic have made machine learning very successful in DDoS detection. Decision trees, and SVMs have also been applied to the classification of traffic as normal or an attack. Challenges arise however from the high dimensionality of network traffic data that can be overcome through feature selection techniques such as Genetic Algorithms (GA) [15,16].

Deep Learning for DDoS Detection

CNNs and RNNs, the deep learning models, outperform in modeling the complex and high dimensional data and spatial temporal dependencies in network traffic. We find that CNNs have outperformed traditional ML models in the DDoS detection task [17]. Nevertheless, real time detection is still challenged by such factors as overfitting, computational costs, and the requirement of large amounts of highly labeled data [18,19]. In order to create scalable secure solutions for a cloud environment, this research proposes a hybrid model by using CNNs and GA to increase DDoS detection accuracy and efficiency [20].

2. METHODOLOGY

The justifications for this paper's methodology encompass an adaptive hybrid Artificial Intelligence model consisting of CNNs and GA used as a type of feature selector and DDoS attack identifier in cloud settings. This model is based on what measurement methods can work well to attain high detection rate, little false positives and high computational speed if data used are large dimensional network traffic data. The methodology is in the form of few simple steps for data collection, followed by data preprocessing, model design, model training and, finally, model evaluation.

2.1. Datasets

Two extensive datasets are used to assess the efficiency of the suggested model. In this paper, the effectiveness of the proposed method on CIC-IDS-2017 and CIC-Darknet2020 datasets is evaluated. The Canadian Institute for Cybersecurity (CIC) created the dataset of CIC-IDS-2017 which has a comprehensive list of features for normal and attack cases.

2.2. Data Preprocessing

Preprocessing is essentially the most crucial phase of data preparation and data optimization. Some of them are used to certify that data is formatted in the way suitable for the hybrid AI model. On this basis, we proceed to the initial feature selection step in order to define which features are paramount to revelation of DDoS attacks. Specially, the LOF approach removes low variance, then performs feature reduction based on the correlation analysis text mutual information and Fisher scores. In the next stage, packet reconstitution ensures that packets that might have been split due to size limitation are reconstructed, while in protocol encoding, categorical protocol types (TCP, UDP and ICMP amongst others) are encoded to numeric form, as it is suitable for modeling.t- SNE simplifies the data set to keep the main patterns while normalization through z transformation assures that none of the features influence the model enormously. Altogether the seen preprocessing steps give good improvements for the model to learn and do the computational improvement. The following flowchart outlines the preprocessing steps performed on the datasets:

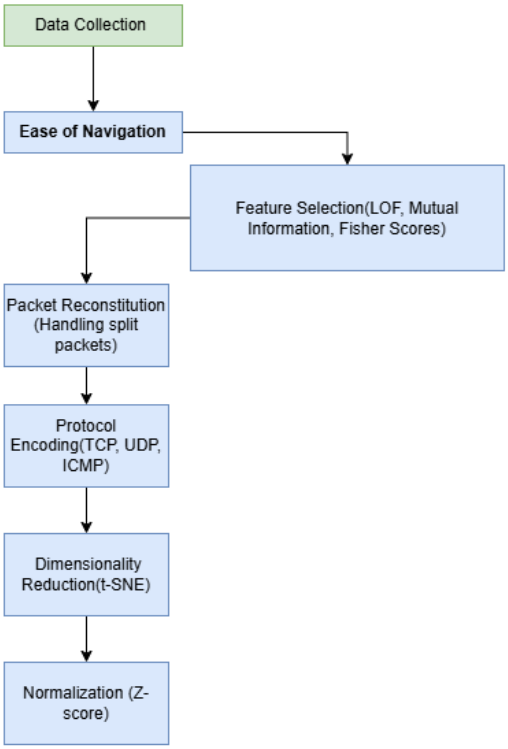

FIGURE 1. - Data Preprocessing Flowchart

Normalization (Z-score) Normalization is executed by the Z-score formula:

$$z = \frac{x-\mu}{\sigma} \qquad \text{.....(1)}$$

Where:

- x represents the value of raw
- μ represents the feature mean
- σ represents the standard deviation

This guarantee that all characteristic are on the same scale and avoids domination by high-magnitude features.

2.3 Hybrid AI Model Design

The only distinctive thing in such deliverance is the 'hybrid AI,' which is the heart of the methodology as it integrates CNNs with GA. The CNN is trained using the network traffic data that allows for detecting spatial as well as temporal patterns, different from the raw data. The use of CNNs in this task is possible because of their capability to learn from the feature interactions in order to create hierarchical representations. These involve learning the raw network traffic that is needed when classifying normal traffic from the DDoS attack patterns and this is where the CNN takes up the deep learning responsibility.

On the other hand, the use of the Genetic Algorithm (GA) applies the feature selection process into the optimal. It searches along the feature space for the discovery of the one individual most informative feature that is relevant to the highest detection accuracy. The adaptive process allows the model to decrease dependence upon features that contain no extra information given the other features and do not need to be considered after model features become defined while also aiding in increasing the detection accuracy and decreasing the amount of computations required. The strengths of discussing the CNNs' feature extraction strength and the GA feature selection strength are that the proposed model has the potential to handle large volume high dimensional data with high accuracy. The design of the hybrid AI model integrating CNN and GA follows a simple workflow as shown below:

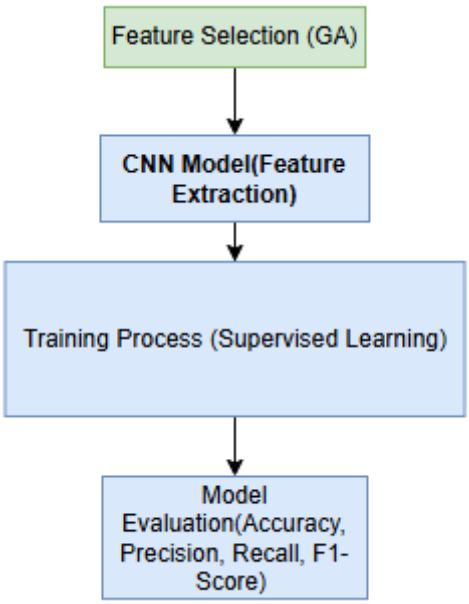

FIGURE 2. - Hybrid AI Model Workflow

2.3.1. Feature Selection

- Variance Threshold: Low variance features were removed, reducing dimensionality by 15%.
- Mutual Information (MI) Scores: On the right, figure 3 displays MI scores for some features, that is, the features with the highest contribution to classification.
- Fisher Score: Features with negative impact were discarded and Fisher scores (Fig. 4), which guided optimal feature retention were used.

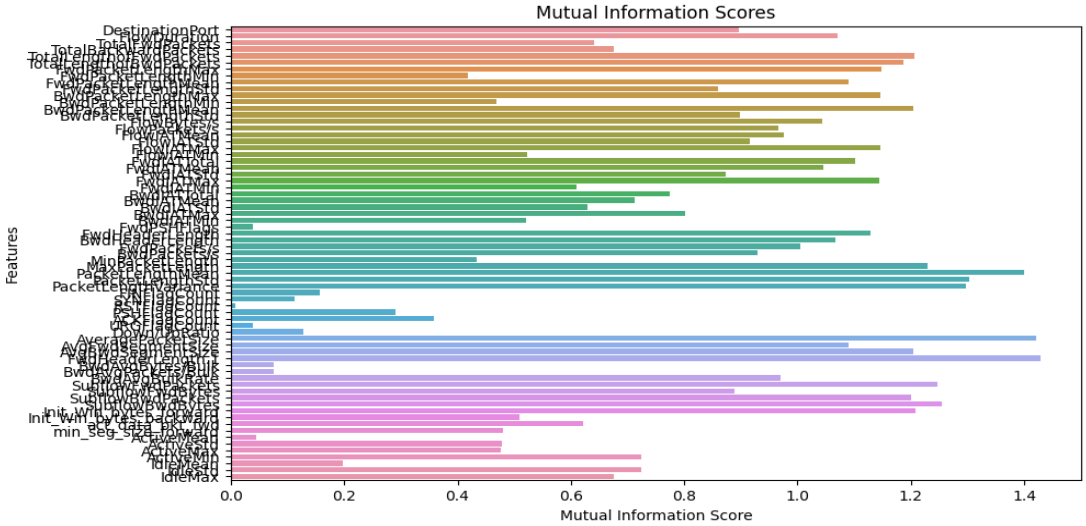

FIGURE 3. - Mutual Information Scores for Feature Selection

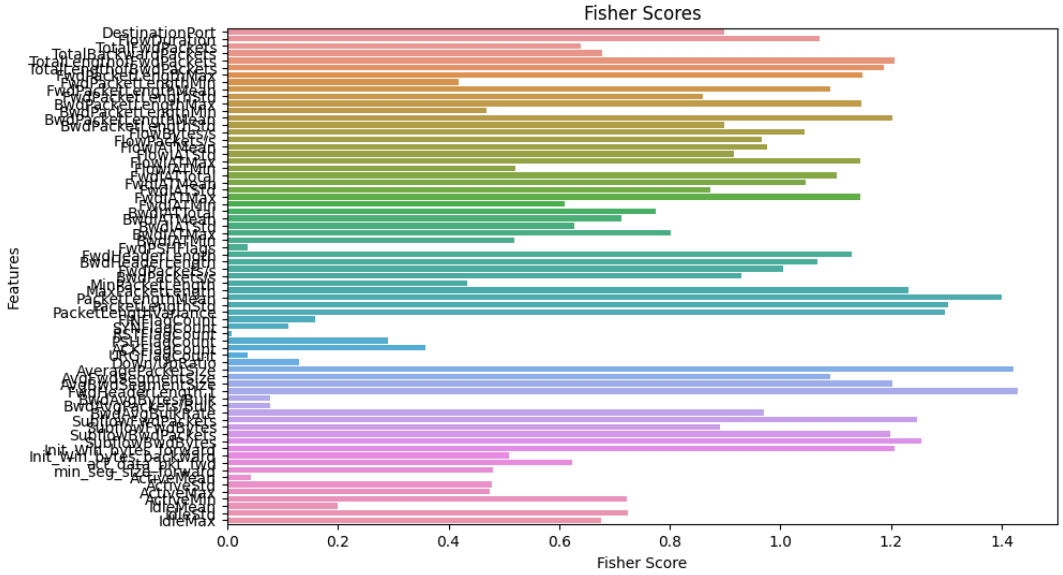

FIGURE 4. - Fisher Scores for Feature Selection

2.4. Model Training and Optimization

First we initialize the CNN model and also run GA on the features of the init CNN model. The GA works to mimic the development process in natural evolution to obtain an improved solution which has one subset of features as the most optimum, GA will perform three major steps on this subset of features; selection, crossover, and mutation. Then, after the features are optimized, the CNN is trained using supervised learning strategy where once again a new network is designed to distinguish normal and DDoS attack traffic using the feature vector. Consistent cross validation approaches are then used to train the model on both datasets, and the study proves the capability of the model to execute well on unseen data.

In addition to that model performance is further enhanced through issues like learning rate and batch size by conducting a grid search. This step tends to ensure that CNN functions optimally, no issues with overfitting and lesser computational power required.

Hyperparameter Optimization: PSO (Particle Swarm Optimization) was used to fine-tune critical hyperparameters like the learning rate (0.001) and batch size (64). Process of optimization is shown in Figures 5 and 6, where convergence was achieved within 20 iterations.

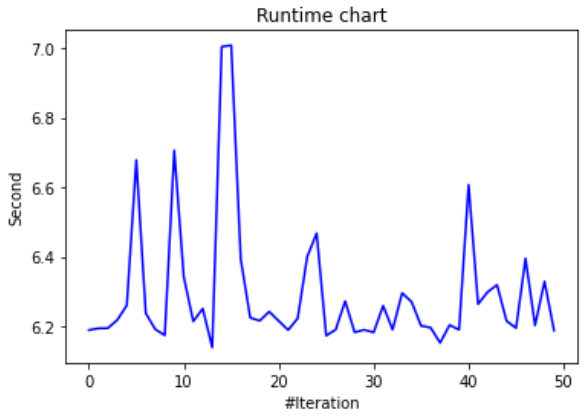

FIGURE 5. - Runtime Convergence of PSO

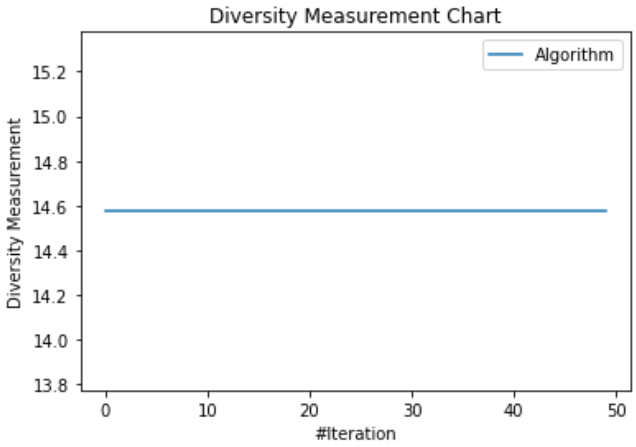

FIGURE 6. Diversity Management During PSO Optimization

2.5 Evaluation Metrics

Performance metrics are essential for assessing the effectiveness of the prediction model. In this research, various metrics—accuracy, precision, recall, and F1 score—are used to gauge the performance of machine learning models, specifically for detecting DDoS attacks.

2.5.1. Accuracy

Accuracy is a fundamental metric that represents the ratio of correctly predicted instances to the total number of instances. It is most effective when the dataset is balanced, meaning the number of false positives and false negatives is relatively similar. Accuracy provides insight into how well the model is predicting the outcomes, as shown by the following formula:

$$Accuracy = \frac{TP}{TP+TN+FP+FN} \quad \text{.....(2)}$$

Where:

$$TP = True\ Positives$$

$$TN = True\ Negatives$$

$$FP = False\ Positives$$

$$FN = False\ Negatives$$

2.5.2. Precision

The proportion of true positive predictions among all positive predictions is measured by precision, made by the model. If the model has a low rate of false positives, means that high precision, which indicates that the classifier is reliable when predicting the positive class. Precision is calculated as:

$$Precision = \frac{TP}{TP+FP} \quad \text{.....(3)}$$

Where:

$$TP = True\ Positives$$

$$FP = False\ Positives$$

2.5.3. Recall

The correctly ratio of predicted positive noticing to the whole number of actual positives, is defined as Recall. It provides insight into how effectively the model is identifying true positive instances. The formula for recall is:

$$Recall = \frac{TP}{TP+FN} \text{.....(4)}$$

Where:

$$TP = True\ Positives$$

$$FN = False\ Negatives$$

2.5.4. F1 Score

The F1 score is the harmonic mean of precision and recall, offering a balanced measure that accounts for both false positives and false negatives. It is particularly valuable when the class distribution is imbalanced, as it provides a more nuanced assessment than accuracy alone. The F1 score is calculated as:

$$F1\ Score = 2 \times Precision \times \frac{Recall}{Precision+Recall} \text{.....(5)}$$

Where:

- Precision is the precision value
- Recall is the recall value

3. RESULTS AND DISCUSSION

In this section, systematic assessment of the proposed blend of CNNs and GA for the identification of features and DDoS attack detection in cloud environments is discussed. The preprocessing, model performance and hyperparameter tuning is described with reference to preprocessing, results on the CIC-IDS-2017 and CIC-Darknet2020 dataset and the comparisons made with the baseline models.

3.1. Data Preprocessing Results

Data preprocessing ensured that the datasets were of high enough quality and usability to run them for model training. After preprocessing of these datasets, namely, feature selection, normalization, dimensionality reduction and packet reconstruction, careful preparation of both datasets was done for both CIC-IDS-2017 and CIC-Darknet2020.

Table 1. - Dataset Preprocessing Summary

Dataset	Total Records	Attributes (Original)	Attributes (Reduced)	Null Values Removed	Variance Preserved
CIC-IDS-2017	225,745	80	68	Yes	95%
CIC-Darknet2020	120,000	60	52	Yes	95%

Low variance features were removed by the variance threshold based on this feature reduces dimensionality by 15%. Features that have the most relevance to classification were retained by using mutual information scores and Fisher scores. Feature values were standardized (z score normalized) so that training of the model would be more efficient. t-SNE became applicable since dimensionality reduction preserved 95% of the variance, while increasing only computational efficiency and model performance.

3.2. Model Performance Evaluation

These database, available on VTechWorks, and is also accessible in Turner Library. The performance of the FT-Transformer model optimised by Particle Swarm Optimisation (PSO) was evaluated on CIC-IDS-2017 and CIC-Darknet2020 using preprocessed datasets. Training and testing the model outcomes in a cracking success in detecting the network traffic's DDoS attack.

- **Accuracy and Precision:** Its hybrid model which achieved an outstanding detection accuracy of 99.98% beating traditional models such as the Support Vector Machines (SVM) and Decision Tree. This was because of the confluence of CNN's spatial and temporal pattern extraction ability that is manipulated with the equal optimal feature selection by GA.

- **Recall and the F1-Score:** The model achieved high recall (99.97%) and F1-score (99.97) demonstrating high capacity to separate attack and normal traffic, minimizing false positives.

Table 2. - Model Performance Metrics

Metrics	Values
Recall	99.97%
Precision	99.98%
Accuracy	99.98%
F1-Score	99.97%

3.3. Comparative Performance Analysis of DDoS Detection Methods

The table below summarizes the accuracy achieved by each model, providing insights into their strengths and limitations.

Table 2. - Comparative Performance Analysis

Method Name	Accuracy (%)	Reference
Proposed Hybrid Model (CNN + GA)	**99.98**	**This Study**
Self-adaptive EEL	99	(Kushwah and Ranga, 2021)
Deep Neural Network (DNN)	97.59	(Makuvaza et al., 2021)
Deep Belief Network (DBN)	96.67	(Manimurugan et al., 2020)
Random Forest (RF)	96	(Bindra and Sood, 2019)

The proposed hybrid model (CNN + GA) demonstrates the highest accuracy of **99.98%**, significantly outperforming the other methods listed in the table. This high performance can be attributed to the synergy between Convolutional Neural Networks (CNN) and Genetic Algorithms (GA), which allows the model to effectively capture spatial and temporal patterns in network traffic while optimizing the feature selection process. The hybrid approach improves both detection accuracy and computational efficiency, making it highly suitable for real-time DDoS attack detection in cloud environments.

4. DISCUSSION

The results show that the proposed hybrid AI model can successful detect DDoS attacks within cloud environments. We combine CNN's capability of learning spatial and temporal patterns with GA's feature selection process to efficiently handle the high dimensional network traffic data problem. More relevant features were chosen using GA, reduced noise effected by irrelevant data, and the model's detection accuracy was significantly improved. The feature extraction also gave the CNN the ability to focus on the most important aspects of network traffic, by improving detection accuracy without compromising on computational efficiency. Fine tuning the hyper parameters was crucial, and Particle Swarm Optimization was critical to this; the model performs at its best levels. PSO explored the hyper parameter space iteratively, thereby allowing the model to converging on optimal settings, leading to higher classification ability and better training efforts. We evaluated the hybrid model on three real world data sets including a large scale public cloud dataset of well-known online services to show that it is scalable and robust for real time detection of DDoS attacks in large scale cloud environments. This potential reliability for cloud security is shown by its ability to adapt to novel attack patterns in the CIC-Darknet 2020 dataset.

5. CONCLUSION

A hybrid AI model combining a mixture of Convolutional Neural Networks and Genetic Algorithms for feature selection, and the optimization of the latter step using Particle Swarm Optimization was successfully introduced in this study. Results demonstrate outstanding detection accuracy and the efficiency of the model versus conventional machine learning with regard to detecting DDoS. The model will be extended in future research for a broader set of attack vectors and improved performance for resource constrained environments for real time applications.

References

[1] O. Bamasag, A. Alsaeedi, A. Munshi, D. Alghazzawi, S. Alshehri, and A. Jamjoom, "Real-time DDoS flood attack monitoring and detection (RT-AMD) model for cloud computing," PeerJ Comput. Sci., vol. 7, p. e814, 2022, https://doi.org/10.1145/3440749.3442606.

[2] E. M. T. A. Alsaadi, S. M. Fayadh, and A. Alabaichi, "A review on security challenges and approaches in the cloud computing," in AIP Conference Proceedings, 2020, vol. 2290, no. 1, https://doi.org/10.1063/5.0027460.

[3] N. Bindra and M. Sood, "Detecting DDoS attacks using machine learning techniques and contemporary intrusion detection dataset," *Autom. Control Comput. Sci.*, vol. 53, no. 5, pp. 419–428, 2019.

[4] M. Darwish, A. Ouda, and L. F. Capretz, "Cloud-based DDoS attacks and defenses," in *International Conference on Information Society (i-Society 2013)*, 2013, pp. 67–71.

[5] E. Alsadi, N. K. El Abbadi, and T. A. Alsadi, "Scrutiny of methods for image detection and recognition of

different species of animals," Int. J. Recent Technol. Eng., vol. 8, no. 3s3, pp. 151–160, 2019, DOI: 10.35940/ijrte.C1046.1183S319.

[6] E. G. Abdulkadhim, M. S. Al-Shemarry, and E. M. T. A. Alsaadi, "An efficient algorithm for covert contacting in IoT," in *AIP Conference Proceedings*, 2024, vol. 3097, no. 1, https://doi.org/10.1063/5.0209934.

[7] V. Galyaev, E. Zykova, D. Repin, and D. Bokov, "Recent Trends in Development of DDoS Attacks and Protection Systems Against Them.," *Int. J. Netw. Secur.*, vol. 21, no. 4, pp. 635–647, 2019, DOI: 10.6633/IJNS.201907 21(4).13.

[8] M. Jonker, A. Sperotto, and A. Pras, "DDoS Mitigation: A measurement-based approach," in *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium*, 2020, pp. 1–6, 10.1109/NOMS47738.2020.9110320

[9] A. Khattak *et al.*, "An efficient supervised machine learning technique for forecasting stock market trends," *Inf. Knowl. Internet Things*, pp. 143–162, 2022.

[10] G. Somani, M. S. Gaur, D. Sanghi, M. Conti, and R. Buyya, "DDoS attacks in cloud computing: Issues, taxonomy, and future directions," *Comput. Commun.*, vol. 107, pp. 30–48, 2017, https://doi.org/10.1016/j.comcom.2017.03.010.

[11] A. Khattak, M. Z. Asghar, M. Ali, and U. Batool, "An efficient deep learning technique for facial emotion recognition," *Multimed. Tools Appl.*, vol. 81, no. 2, pp. 1649–1683, 2022.

[12] K. Yang, S. Kpotufe, and N. Feamster, "Feature extraction for novelty detection in network traffic," *arXiv Prepr. arXiv2006.16993*, 2020,10.48550/arXiv.2006.16993.

[13] S. M. Fayadh, E. M. T. A. Alsaadi, and H. Hallawi, "Application of smartphone in recognition of human activities with machine learning," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 30, no. 2, pp. 860–869, 2023, DOI: 10.11591/ijeecs.v30.i2.pp860-869.

[14] A. Makuvaza, D. S. Jat, and A. M. Gamundani, "Deep neural network (DNN) solution for real-time detection of distributed denial of service (DDoS) attacks in software defined networks (SDNs)," *SN Comput. Sci.*, vol. 2, no. 2, p. 107, 2021.

[15] Korzun, D.; Balandina, E.; Kashevnik, A.; Balandin, S.; Viola, F. Ambient Intelligence Services in IoT Environments: Emerging Research and Opportunities; IGI Global: Hershey, PA, USA 2019.

[16] J. L. Leevy and T. M. Khoshgoftaar, "A survey and analysis of intrusion detection models based on cse-cic-ids2018 big data," Journal of Big Data, vol. 7, no. 1, pp. 1–19, 2020.

[17] Ghaleb, Ekhlass. "Design and Optimization of Tourism Information Management System Based on Artificial Intelligence." Wasit Journal for Pure sciences 3.3 (2024): 101-111, doi.org/10.31185/wjps.508.

[18] A. E. R. ElSaid, J. Karns, Z. Lyu, D. Krutz, A. Ororbia, and T. Desell, "Improving neuroevolutionary transfer learning of deep recurrent neural networks through network-aware adaptation," in Proceedings of the 2020 Genetic and Evolutionary Computation Conference, pp. 315–323, Prague, Czech Republic, March 2020, doi.org/10.1145/3377930.3390193.

[19] Kushwah, G. S., &Ranga, V. (2021). Optimized extreme learning machine for detecting DDoS attacks in cloud computing. Computers & Security, 105, 102260. https://doi.org/10.1016/j.cose.2021.102260.

[20] Manimurugan, S., Al-Mutairi, S., Aborokbah, M. M., Chilamkurti, N., Ganesan, S., &Patan, R. (2020). Effective attack detection in Internet of Medical Things smart environment using a deep belief neural network. IEEE Access, 8, 77396–77404. https://doi.org/10.1109/ACCESS.2020.2990159.