Grenze International Journal of Engineering and Technology, June Issue

Enhancing Intrusion Detection Systems: Unleashing the Power of Hybrid Optimization Approaches

Jeevasurya V[1], Kishore Kumar R[2], Sreekanth S[3] and Hemashree P[4]
[1-4]Coimbatore Institute of Technology, Coimbatore, India
Email: {jeevanandsurya, kishorekumar0813, sreekanthsubramanian78}@gmail.com, hemashree@cit.edu.in

***Abstract*—Intrusion detection is a vital defense mechanism that protects computer systems, networks, and data from unauthorized access, attacks, and malicious activities. It enables real-time or near real-time detection of security breaches, allowing immediate action to mitigate the impact of intrusions. By monitoring network traffic, system logs, and user activities, Intrusion Detection Systems (IDS) aid in identifying and responding to threats promptly. Moreover, IDS plays a crucial role in analyzing the nature and severity of intrusions, facilitating the implementation of appropriate countermeasures, and enhancing overall security. Compliance with security standards and regulations is also ensured, maintaining customer trust and privacy protection. Traditional IDS approaches often struggle to effectively detect sophisticated and evolving cyber threats. To address these limitations, this article proposes a novel approach to enhance IDS performance through the use of hybrid optimization techniques. The proposed work uses the backpropagation algorithm, a widely used method for training artificial neural networks (ANNs), along with optimization algorithms like Artificial Bee Colony (ABC) and Harmony Search (HS), to optimize intrusion detection performance by selecting the most effective features while reducing resource requirements. The integration of backpropagation neural networks with the ABC and HS algorithms presents a promising solution for enhancing IDS. These hybrid optimization techniques not only enhance detection accuracy but also address the crucial issue of feature selection, reducing IDS complexity and computational requirements. This research contributes to the development of more robust and efficient IDS systems capable of detecting and mitigating advanced cyber threats in real-time network environments.**

Keywords: **Intrusion Detection System, Feature Selection, Artificial Bee Colony Exploration, Harmony Search, Meta-Heuristic, Backpropagation.**

I. INTRODUCTION

Intrusion detection is an indispensable defense mechanism in the realm of cybersecurity. With the increasing prevalence of unauthorized access attempts, attacks, and malicious activities targeting computer systems, networks, and valuable data, the need for robust Intrusion Detection Systems (IDS) has become paramount. These systems enable the timely detection of security breaches, empowering organizations to take immediate action and mitigate the impact of intrusions. This research serves to underscore the vital importance of intrusion detection in today's ever-evolving cyber threat landscape. By monitoring various aspects such as network traffic, system logs, and user activities, IDS plays a pivotal role in swiftly identifying and responding to traffic, system logs, and user activities, IDS plays a pivotal role in swiftly identifying and responding to potential threats. This

Grenze ID: 01.GIJET.9.2.383

capability is critical in ensuring the security and integrity of systems and networks, as well as protecting sensitive data from unauthorized access or modification. Furthermore, the IDS go beyond mere detection; they also assist in analyzing the nature and severity of intrusions. This analysis provides valuable insights for implementing appropriate countermeasures and fortifying the overall security posture of the organization. Compliance with security standards and regulations is another key aspect facilitated by intrusion detection, ensuring that organizations maintain the necessary measures to protect customer trust, uphold the privacy, and meet legal requirements. To optimize the performance of intrusion detection, researchers have explored the utilization of advanced techniques. These techniques include machine learning algorithms such as Support Vector Machines (SVM), Clustering techniques, and other classification algorithms. These algorithms, coupled with Meta-Heuristic optimization approaches like Bat Search, Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), and many other optimization techniques, enhance the efficiency and effectiveness of intrusion detection by identifying and selecting the most relevant features. This optimization contributes to reducing resource requirements and improving overall performance. The proposed work focuses on using Meta-Heuristic techniques, namely Artificial Bee Colony (ABC) and Harmony Search (HS), for selecting important features that contribute to detecting intrusion in a system. Using the selected features, a Backpropagation deep neural network model is built to detect an intrusion, given the selected features about a particular system in place.

II. Related Work

In the work proposed in [1], the authors have used a hybrid Bat Meta-Heuristic Algorithm with a Support Vector Machine (SVM) classifier to perform feature selection and reduce the dimensionality to improve the IDS. The NSL-KDD dataset was used for building and evaluating the system which outperformed existing systems with a reduced number of features. [2] works based on Network-Based Intrusion Detection Systems (NIDSs) which are network traffic patterns that are used for the prediction of possible attacks. The model used throughput (TP), End-To-End Delay (ETED), and Packet Delivery Ratio (PDR) as performance metrics. For the experiment, an NS2-based NISC network architecture generator coded in Python3 to simulate an attack and normal scenarios in NS2 with configurable nodes and dynamic clusters are used. A low-rate TCP denial of service (DoS) attack on cluster-based network architecture is implemented. To analyze the behavior a NICS-based testbed, which comprises scalable clusters of nodes of various network architectures and platforms is used. In [3], the Cuckoo Search initially, the number of population nests, and the iterations are initialized. After each iteration, the hybrid search is used to update the population of nests. It updates the population nest based on the conditions that the lowest quality nests are updated randomly based on the global search, and the remaining nests are updated locally using levy's flight. Java platform and WEKA 3.9 tool is used for classification. In [4], proposes a Cyber Intrusion Detection System (CIDS) using a Multi-Objective Binary Bat Algorithm (MOBBA) for feature selection and an Enhanced Bat Algorithm (EBA) for parameter optimization in neural networks. The authors aim to design a more efficient and accurate CIDS by integrating metaheuristic algorithms with deep learning techniques. The proposed CIDS uses MOBBA to select the most relevant features from the input dataset, which reduces the dimensionality of the dataset and speeds up the training process of the neural network. EBA is then used to optimize the parameters of the neural network to improve its accuracy. Likewise, in [5] an intrusion detection system for network security using an elite clone artificial bee colony and backpropagation neural network was proposed. The authors used an Elite Clone Artificial Bee Colony (ECABC) to optimize the weights of the Backpropagation Neural Network (BPNN). The results showed that the proposed system outperforms other methods in terms of detection rate and false alarm rate. In [6], an Improved Binary Whale Optimization Algorithm (IBWOA) is used to achieve a better dimension reduction to get better accuracy for feature selection of intrusion detection. The slow convergence of the WOA algorithm is overcome by introducing PSO within its updating mechanism. In [7], the authors propose an MGWO, which is a Modified Gray Wolf Optimization algorithm to improve the efficiency of the IDS. MGWO improves the initialization phase, by utilizing filter and wrapper approaches to include important features in the early iterations. Similarly, in [8], a novel routing algorithm was introduced, leveraging the principles of Differential Evolution (DE) to enhance the LEACH routing protocol. The aim was to cater to the specific demands of monitoring applications in outdoor environments, such as meteorological, hydrological, and wetland ecological settings. By harnessing the efficient and rapid search capabilities of DE, the proposed algorithm focused on optimizing the multi-objective selection of cluster heads while mitigating blind nodes. This approach aimed to enhance energy efficiency and system stability. Through simulations, it was observed that the newly proposed LEACH routing algorithm exhibited superior performance. It effectively extended the operational lifespan of the system and improved the overall

quality of the wireless sensor networks deployed. The research presented in [9] focuses on the effective utilization of cuckoo search for automated cryptanalysis or attacks on substitution ciphers. Specifically, the study enhances a previously proposed attack based on a genetic algorithm that targets the simple substitution cipher. A comparative analysis is conducted among different attack algorithms, including cuckoo search, genetic algorithm, enhanced genetic algorithm, tabu search, and scatter search. Among these approaches, cuckoo search demonstrates the highest level of performance. In [10], a highly effective and precise method for detecting Distributed Denial of Service (DDoS) attacks is introduced, utilizing the Lion optimization algorithm. This proposed detection technique demonstrates robustness in identifying DDoS attacks, even when the attack traffic is minimal in magnitude. To assess its performance, a comparison is made with state-of-the-art techniques. The experimental results validate the superiority of the proposed method, showcasing its ability to outperform existing state-of-the-art approaches with an impressive accuracy rate of 96%. This highlights the efficacy and reliability of the proposed method in detecting DDoS attacks, making it a significant advancement in the field of intrusion detection. In [11], a novel time synchronization technique inspired by the behavior of spider monkeys was proposed for large-scale Vehicular Ad-Hoc Networks (VANETs). The main objective was to enhance packet delivery time synchronization while minimizing energy consumption. The technique utilized a metaheuristic-stimulated framework based on the natural behavior of spider monkeys. The proposed technique was employed to investigate Sybil attacking strategies in VANETs. In the paper [12], a method called PSO-LightGBM was introduced for intrusion detection. This method combines particle swarm optimization (PSO) with gradient descent, utilizing LightGBM for feature extraction from the data. The extracted features are then fed into an OCSVM (One-Class Support Vector Machine) algorithm to detect and classify malicious data.

III. Motivation

Being able to detect and stop attacks on computer systems makes creating an intrusion detection system a crucial challenge in the realm of cybersecurity. Known risks are identified using rule-based or signature-based techniques in conventional methodologies. These techniques might not, however, be successful in identifying new or undiscovered risks. The intrusion detection system can be improved to obtain better accuracy in detecting intrusions while lowering the false positive and false negative rates by applying a metaheuristic optimization method. Two well-liked meta-heuristic optimization algorithms that have been effectively used in diverse domains are Harmony Search and Artificial Bee Colony. A population-based algorithm called Harmony Search imitates musical improvisation in order to produce novel answers by using a musical metaphor. It has been shown to be effective in refining the solutions generated by other optimization algorithms. Using Harmony Search followed by Artificial Bee Colony for feature selection in intrusion detection can result in better optimization results than using either algorithm alone. Harmony Search can be used to generate a subset of relevant features, and Artificial Bee Colony can be used to refine these features to further improve the accuracy of intrusion detection.

IV. Solution Methodologies

The methodology used in this research uses the NSL-KDD dataset, which is a widely used benchmark dataset for testing intrusion detection. The data has 42 features, all corresponding to various values of the system at different times, which also includes the time of intrusion. All these features are not necessarily important for detecting intrusions. Therefore, it is enough to use a subset of important features to build the model that will detect intrusions. Meta-heuristic algorithms can be used for feature selection. In this research work, two meta-heuristic algorithms, namely, Harmony Search (HS) and Artificial Bee Colony (ABC) are used. A hybrid approach that combines both algorithms for feature selection (HS for exploration and ABC for exploitation) is also considered in the proposed work. The results of all these methods are evaluated on the basis of the Area under the curve of the Receiver Operating Curve and the False Positive Rate. Using the selected features, a backpropagation sequential model was built for predicting intrusion.

A. Need for Feature Selection

The NSL KDD dataset is a widely used benchmark dataset in the field of intrusion detection. It was developed as an improvement over the original KDD Cup 1999 dataset to address certain limitations. This dataset comprises network traffic data with various features representing network connections, allowing researchers to develop and evaluate intrusion detection systems. The NSL KDD dataset includes features such as protocol type, service, source and destination IP addresses, flags, and various statistical attributes. These features provide valuable information about network connections and can be used to classify them as normal or malicious. Feature

selection plays a crucial role in intrusion detection using the NSL KDD dataset. By selecting the most relevant and informative features, the performance of intrusion detection algorithms can be improved. Feature selection reduces computational complexity, enhances efficiency, and helps focus on the most discriminative attributes for accurate intrusion detection. Researchers utilize the NSL KDD dataset to train and evaluate machine learning models and intrusion detection algorithms. By leveraging feature selection techniques, they can identify the most influential features for classifying network connections as normal or intrusive. This process improves the accuracy and effectiveness of intrusion detection systems, allowing for better identification and response to potential security breaches.

B. Backpropagation Model

The Backpropagation algorithm [15] is a widely used method for training artificial neural networks (ANNs) and has proven to be effective in various classification tasks, including intrusion detection. The complex features present in the NSL KDD dataset, such as protocol type, service, IP addresses, and statistical attributes, require a robust model like backpropagation to capture the underlying patterns and relationships. Backpropagation allows the ANN to iteratively learn and adjust the weights of its connections based on the error signal obtained during training. By training a backpropagation model on the NSL KDD dataset, the ANN can learn to recognize and classify complex network connections as either normal or malicious. The model's ability to handle non-linear relationships and complex feature interactions makes it well-suited for intrusion detection, where identifying intricate attack patterns is crucial. The backpropagation model excels in learning hierarchical representations of the data, enabling it to uncover hidden features and distinguish between different classes of network connections. Through multiple iterations of forward and backward propagation, the model can optimize its weights to minimize classification error, ultimately improving the accuracy and performance of intrusion detection.

C. Artificial Bee Colony

In the Artificial Bee Colony (ABC) [14] Algorithm, the colony comprises three types of bees: employed bees, onlooker bees, and scout bees. Employed bees are assigned to specific food sources and perform local searches to improve the quality of their associated solutions. Onlooker bees observe the dances performed by employed bees to select promising food sources. Scout bees, initially, discover all the food source positions randomly. During the optimization process, both employed and onlooker bees exploit the nectar of food sources, gradually depleting their resources. As a result, employed bees may become exhausted. In such cases, the employed bee that exploited the exhausted food source transforms into a scout bee and begins searching for new food sources once again. In essence, an employed bee becomes a scout bee when its food source is depleted. In ABC, each food source represents a potential solution to the problem being solved, and the nectar amount associated with a food source represents the quality or fitness of that solution.

D. Harmony Search

Harmony Search (HS) [13] is a meta-heuristic search algorithm that draws inspiration from the improvisation process of musicians to find harmonious solutions. In recent years, HS has garnered considerable attention due to several advantages it offers. Implementing HS is straightforward, and the algorithm has demonstrated fast convergence to optimal solutions. Moreover, HS is capable of finding sufficiently good solutions within a reasonable computational time. These merits have made the HS algorithm a popular choice for optimizing various engineering problems across different domains.

V. EXPERIMENTAL SETUP AND RESULTS

A. Data

In the intrusion detection industry, the NSL KDD dataset is a frequently used benchmark dataset. Compared to the original KDD Cup 1999 dataset, it offers a more accurate picture of network traffic. Important elements of the collection include flags, source and destination IP addresses, protocol type, service, and flags. Researchers can create and assess efficient intrusion detection systems because these properties capture crucial information about network connections. The NSL KDD dataset's wide range of characteristics makes it an invaluable tool for researching and addressing the difficulties of intrusion detection in complicated network environments. The dataset has a total of 43 features of which 42 are dependent variables and 1 feature is the independent target variable.

B. Setup

The experiment was conducted using Google Colab with runtime type as GPU. 30% of the original dataset is used for evaluating the performance of the model being built. The harmony search algorithm uses the values of 10, 0.95, and 0.5 for Harmony Memory Size (HMS), Harmony Memory Considering Rate (HMCR), and Pitch Adjusting Rate (PAR). HMS refers to the number of harmonies, or candidate solutions. HMCR determines the probability of selecting a value from the memory. PAR determines the probability of adjusting a selected value from the memory. Artificial Bee colony uses the values 20 and 5 for the number of employee bees and the number of scout bees respectively. The employee bees in ABC are responsible for exploiting the search space by visiting the food sources (candidate solutions) and assessing their quality. The scout bee's role in ABC is to explore the search space globally and introduce new solutions to the population.

C. Model

TABLE I. MODEL

Appearance		
Layer (type)	**Output Shape**	**Param**
dense (Dense)	(None, 64)	1152
dropout (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 128)	8320
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 512)	66048
dropout_2 (Dropout)	(None, 512)	0
dense_3 (Dense)	(None, 128)	65664
dropout_3 (Dropout)	(None, 128)	0
dense_4 (Dense)	(None, 1)	129

The model architecture is shown in Table I. The Backpropagation model was built in Python using TensorFlow. The same architecture is used for building the model on all three feature selection methodologies, and their performance is compared. The algorithm is run for around 25 epochs. The model building started with high loss and slowly converged to get low loss.

D. Results

TABLE II. RESULTS

Technique	**Features**	**Accuracy**	**False Positive Rate**
Harmony Search	14	95.45%	4.17
ABC	26	96.97%	1.32
Harmony Search + ABC	8	96.08%	0.65

Table II, shows the results of the experiments conducted. The original number of features available in the training data is 42. Using Harmony Search, these features were reduced to 14. The backpropagation model built using these features gave 95.45% accuracy with a False Positive Rate (FPR) of 4.17. ABC reduced the features to 26, which gave 96.97% accuracy and an FPR of 1.32. With the help of the hybrid optimization, which

combined both the HS and ABC algorithms extracted the most prominent features, reducing it to 8, which balanced the performance and cost of the model. From the results, we can see that the combined Hybrid approach can produce better results in terms of FPR in comparison with the other methodologies. It is able to achieve low FPR with only a few features compared to the other techniques. The results show improvement over the existing feature selection-based methods in terms of FPR. The Receiver Operating characteristic (ROC) curve of all three techniques is shown in Figures 1, 2, and 3. respectively, based on which the FPR is calculated.

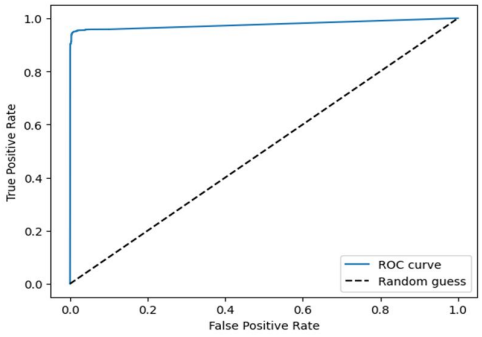

Fig 1. ROC curve under Harmony Search

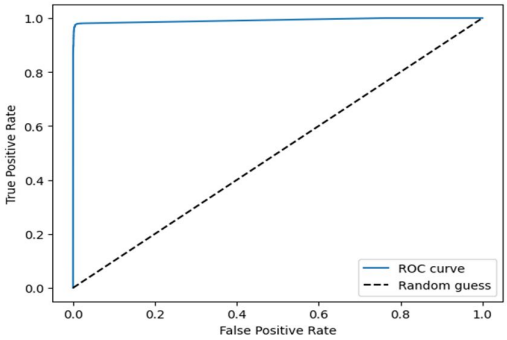

Fig 2. ROC curve under Artificial Bee Colony

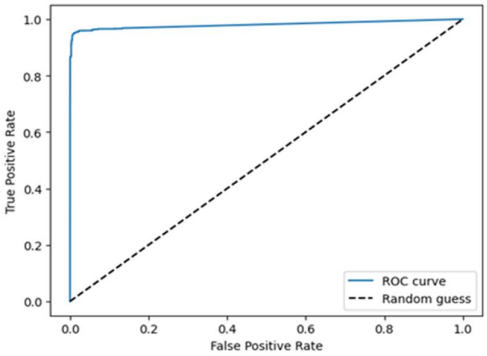

Fig 3. ROC curve under Harmony Search and Artificial Bee Colony

VI. Conclusions

In the process of experimenting with meta-heuristic algorithms like Harmony Search, Artificial Bee Colony, and a hybrid of Harmony Search and Artificial Bee Colony it is observed that the Harmony Search followed by Artificial Bee Colony gives a better performance, even a bit better than the Hybrid Bat Based Algorithm in the way of giving similar results with fewer features, which significantly saves resource and time. The proposed algorithm can be used for building an IDS system, which will only require very few features compared to the

original set of features to achieve reliable results. Harmony Search followed by Artificial Bee Colony can be a promising approach for feature selection in intrusion detection systems, as it can achieve comparable or even better results with a much smaller set of features. The proposed method can be applied to various types of data, including network traffic data and system logs. Overall, the Harmony Search followed by Artificial Bee Colony is a promising approach for feature selection in intrusion detection systems, and it has the potential to significantly improve the efficiency and effectiveness of intrusion detection while reducing the cost and complexity of the system.

FUTURE WORK

The proposed project opens up several avenues for further improvement and exploration. One potential direction for future research involves utilizing diverse datasets to build the IDS. By incorporating different datasets, researchers can evaluate the generalizability and robustness of the proposed technique across various network environments and attack scenarios. Additionally, there is scope for investigating alternative metaheuristic algorithms for feature extraction. Exploring a wider range of algorithms can help identify the most effective ones for capturing relevant features in network traffic data. Furthermore, metaheuristic algorithms can be employed for hyperparameter tuning of the machine learning and Deep learning models. This optimization process can fine-tune the model's parameters and potentially enhance its overall performance. Combining a dedicated algorithm for feature selection with one for parameter tuning can offer an opportunity to evaluate the system's performance comprehensively. Expanding the study to include various machine learning models, beyond the backpropagation neural network, can also be beneficial. Assessing the performance of alternative models can shed light on whether different architectures or algorithms result in improved IDS accuracy and efficiency.

REFERENCES

[1] M. A. Laamari and N. Kamel, "A Hybrid Bat Based Feature Selection Approach for Intrusion Detection," in Bio-Inspired Computing - Theories and Applications, L. Pan, G. Păun, M. J. Pérez-Jiménez, and T. Song, Eds., in Communications in Computer and Information Science. Berlin, Heidelberg: Springer, 2014, pp. 230–238. doi: 10.1007/978-3-662-45049-9_38.

[2] S. K. Shandilya, B. J. Choi, A. Kumar, and S. Upadhyay, "Modified Firefly Optimization Algorithm-Based IDS for Nature-Inspired Cybersecurity," Processes, vol. 11, no. 3, Art. no. 3, Mar. 2023, doi: 10.3390/pr11030715.

[3] R. K. V. P, P. L. N. Raju, K. V. S. Raju, and A. Kalidindi, "Feature Selection and Performance Improvement of Malware Detection System using Cuckoo Search Optimization and Rough Sets," Int. J. Adv. Comput. Sci. Appl. IJACSA, vol. 11, no. 5, Art. no. 5, Aug. 2020, doi: 10.14569/IJACSA.2020.0110587.

[4] W. A. H. M. Ghanem et al., "Cyber Intrusion Detection System Based on a Multiobjective Binary Bat Algorithm for Feature Selection and Enhanced Bat Algorithm for Parameter Optimization in Neural Networks," IEEE Access, vol. 10, pp. 76318–76339, 2022, doi: 10.1109/ACCESS.2022.3192472.

[5] G. Qi, J. Zhou, W. Jia, M. Liu, S. Zhang, and M. Xu, "Intrusion Detection for Network Based on Elite Clone Artificial Bee Colony and Back Propagation Neural Network," Wirel. Commun. Mob. Comput., vol. 2021, p. e9956371, Sep. 2021, doi: 10.1155/2021/9956371.

[6] H. Xu, Y. Fu, C. Fang, Q. Cao, J. Su, and S. Wei, "An Improved Binary Whale Optimization Algorithm for Feature Selection of Network Intrusion Detection," in 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Sep. 2018, pp. 10–15. doi: 10.1109/IDAACS-SWS.2018.8525539.

[7] A. Alzaqebah, I. Aljarah, O. Al-Kadi, and R. Damaševičius, "A Modified Grey Wolf Optimization Algorithm for an Intrusion Detection System," Mathematics, vol. 10, no. 6, Art. no. 6, Jan. 2022, doi: 10.3390/math10060999.

[8] X. Li, L. Xu, H. Wang, J. Song, and S. X. Yang, "A Differential Evolution-Based Routing Algorithm for Environmental Monitoring Wireless Sensor Networks," Sensors, vol. 10, no. 6, Art. no. 6, Jun. 2010, doi: 10.3390/s100605425.

[9] A. Jain and N. S. Chaudhari, "A New Heuristic Based on the Cuckoo Search for Cryptanalysis of Substitution Ciphers," in Neural Information Processing, S. Arik, T. Huang, W. K. Lai, and Q. Liu, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 206–215. doi: 10.1007/978-3-319-26535-3_24.

[10] D. Arivudainambi, V. K. K.A, and S. Sibi Chakkaravarthy, "LION IDS: A meta-heuristics approach to detect DDoS attacks against Software-Defined Networks," Neural Comput. Appl., vol. 31, no. 5, pp. 1491–1501, May 2019, doi: 10.1007/s00521-018-3383-7.

[11] C. Iwendi, M. Uddin, J. A. Ansere, P. Nkurunziza, J. H. Anajemba, and A. K. Bashir, "On Detection of Sybil Attack in Large-Scale VANETs Using Spider-Monkey Technique," IEEE Access, vol. 6, pp. 47258–47267, 2018, doi: 10.1109/ACCESS.2018.2864111.

[12] J. Liu, D. Yang, M. Lian, and M. Li, "Research on Intrusion Detection Based on Particle Swarm Optimization in IoT," IEEE Access, vol. 9, pp. 38254–38268, 2021, doi: 10.1109/ACCESS.2021.3063671.

[13] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A New Heuristic Optimization Algorithm: Harmony Search,"

SIMULATION, vol. 76, no. 2, pp. 60–68, Feb. 2001, doi: 10.1177/003754970107600201.
[14] D. Karaboga, “Artificial bee colony algorithm,” Scholarpedia, vol. 5, no. 3, p. 6915, Mar. 2010, doi: 10.4249/scholarpedia.6915.
[15] Rumelhart, D., Hinton, G., and Williams, R, “Learning representations by back-propagating errors,” Nature, 323(6088), 533-536. doi:10.1038/323533a0.