37th International Conference on Information Technologies (InfoTech-2023)

TITLE: FEATURE SELECTION USING BIO-INSPIRED OPTIMIZATION FOR IOT INTRUSION DETECTION AND PREVENTION SYSTEM

Presented by:

Richa Singh

USIC&T, G.G.S.I.P.U., Delhi, India

Content

- Introduction
- Objective
- Literature Review
- Methods
- Proposed System
- Result
- Conclusion
- References

Introduction

- An Internet of things (IoT) system consists of several interconnected devices that can communicate with each other with minimal human intervention. Security is one of the major concerns in the IoT framework as IoT devices can be easily targeted by attackers for performing malicious activities.
- For ensuring the security of an IoT framework various techniques have been developed, Intrusion Detection and Prevention System (IDPS) is one of them.
- IDPS detects intrusion and take timely measures to prevent them from occurring.
- The detection of abnormal behaviors in the networks such as penetrations, break-ins, or any other form of suspicious activity is called intrusion detection.

Objective

- To reduce the number of features using bio-inspired algorithms.
- To reduce the computation time taken by IDPS.
- To maximize the accuracy of IDPS.

Literature Review

- **An Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms for IoT Intrusion Detection [3]:**
  - The proposed the FS using the bird swarm algorithm(BSA) and gorilla troops optimizer (GTO).
  - However, the hybridized GTO-BSA takes more computation time in detecting optimal features compared to other approaches.
- **Augmented whale feature selection for IoT attacks: Structure, analysis and applications [4]:**
  - Whale optimization (WO) algorithm, with modified using transfer functions is proposed for the FS of IoT-based IDS.
  - The proposed system's performance is better than other similar approaches.
  - However, the proposed work does not outperform for all datasets in term of specificity, and false positive.
- **PSO-Driven Feature Selection and Hybrid Ensemble for Network Anomaly Detection [5]:**
  - The work proposed an anomaly detection system based on ensemble learning.
  - The optimal FS from the network traffic are performed using the PSO algorithm.
  - Afterward, the hybrid ensemble method is used for the classification of reduced data using gradient boosting machine (GBM), and bootstrap aggregation.

Literature Review

- **Intrusion Detection System for IoT Based on Deep Learning and Modified Reptile Search Algorithm [6]:**
  - FS done by reptile search algorithm (RSA) for the IoT-based IDS is proposed in this work.
  - In this paper, first, the features are extracted using deep learning-based CNN.
  - Afterward, RSA is used for selecting essential features. The performance of a system is evaluated against multiple datasets. However, the RSA has slow convergence speed.
- **A Feature Selection Based on the Farmland Fertility Algorithm for Improved Intrusion Detection Systems[7]:**
  - Authors uses a binary form of farmland fertility algorithm for essential FS.
  - Here, a V-shaped function is employed for converting the continuous space to a binary one.
  - The proposed system is evaluated against UNSNB15 and NSL-KDD datasets.
  - Furthermore, to improve the accuracy of proposed work multiple classifiers such as DT, SVM, and KNN are hybridized.

Literature Review

- **Feature Selection Methods for IoT Intrusion Detection System: Comparative Study [8]:**
  - They performs the comparative study of bio-inspired algorithms including SSA, GWO, WO, HHO, used for the FS task of IoT based IDS.
  - The binary and multiclass classification of network traffic is done using the naïve bayes and KNN classifier and their performance are compared against the BoT-IoT dataset

Methods

Harris Hawk Optimization *:* During exploration, the hawks perform a global search for prey, searching for specific locations and constantly observing their environment.

- The update of the hawk's position is dependent on either their family members or random searches in the population area. This process can be mathematically modeled as follows:

$$H(it+1) = \begin{cases} H_{rnd}(it) - rn_3|H_{rnd}(it) - 2rn_4\, H(it)| & if\ q \geq 0.5 \\ H_{pr}(it) - H_{mean}(it) - rn_5\big(LwB + rn_6(UpB - LwB)\big) & if\ q < 0.5 \end{cases} \quad (1)$$

where $H(it)$ is the current hawk position, $H(it+1)$ is the hawk position in next iteration. $rn_3, rn_4, rn_5, rn_6, and\ q$ are all random numbers. The target position that the hawks are trying to reach is denoted as $H_{pr}(it)$. $H_{mean}(it)$ is the hawks mean position, which is calculated as-

$$H_{mean}(it) = \sum_{i=1}^{N} \frac{H_i(it)}{N} \quad (2)$$

where N is the total hawk population.

Methods

- The exploration to exploitation transition in HHO is usually described by the following equation-

$$Es_n = 2E_i - \left(1 - \frac{it}{Max_{it}}\right) \quad (3)$$

where Es_n is escaping energy, E_i is initial energy, 'it' is current iteration. Maximum iteration is denoted by Max_{it}.

Exploitation: During the exploitation, the hawks exploits the prey using different attacking strategies, which includes-

- ***Hard besiege-*** In this strategy, prey is exhausted and does not escape successfully.
- ***Soft besiege-*** In this strategy, prey has enough escaping energy. However, hawks encircle the prey and make them unable to escape.
- ***Hard besiege with Progressive Dive***- Here, hawks attempts to minimize the distance between their own location and prey location.
- ***Soft besiege with Progressive Dive***-Prey has enough escaping energy. Hawks need to find the optimal location to catch the prey.

Methods

- ***Salp Swarm Algorithm:*** SSA is motivated by the hunting habits of salps in the sea and their movement is alike jellyfish. They form a chain while living in a group. The foremost salp in a chain is the leader, and others are followers. Leader position is determined with the following equation-

$$S_n^1 = FP_j - rn_1[(UpB - LwB)rn_2 + LwB] \;\; if \; rn_3 < 0.5 \qquad (4)$$

$$S_n^1 = FP_j + rn_1[(UpB - LwB)rn_2 + LwB] \;\; if \; rn_3 \geq 0.5 \qquad (5)$$

where S_n^1 is the leader position, FP_j food position .rn_2 is a random number between [0,1] used to control the mobility step of the leader. rn_3 controls the switch between two position-updating equations.rn_1 is the control parameter that balances SSA execution. It is defined by the following equation-

$$rn_1 = 2e^{-\left(\frac{4it}{Max_{it}}\right)}$$

- The follower position is determined using the following equation-

$$S_n^m = \frac{1}{2}(S_n^m - S_n^{m-1}) \qquad (6)$$

where, S_n^m is the mth follower in nth dimension.

Proposed System

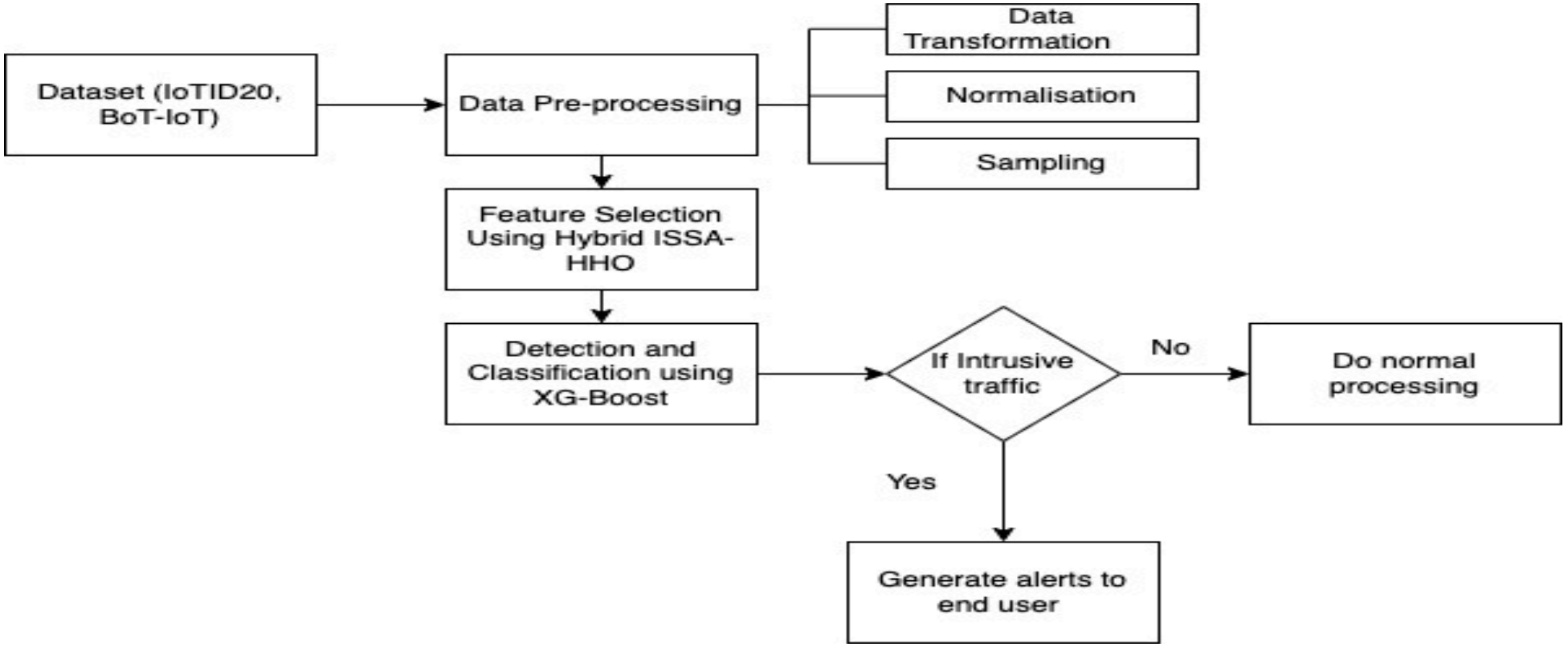

Figure 1. Proposed System framework

Proposed System

IoTID20 Dataset

- Smart home devices such as laptops, tablets, Wi-Fi cameras, smartphones, and other devices are used to generate network flow data for the IoTID20 dataset.
- These devices are divided into categories of victim devices i.e. EZVIZ Wi-Fi camera, SKT NGU, and the attacking devices including smartphones, tablets, etc. all other IoT devices. This labeled dataset includes 83 features.

Table 1. IoTID20 Dataset instances

Category	Instances	Category	Instances	Category	Instances
Mirai	415677	DoS	59391	MITM ARP	35377
Scan	75265	Normal	40073	Spoofing	

BoT-IoT Dataset

- The synthetic testbed configuration includes IoT services, analytical tools, and network platform with attacking and normal virtual machines. This labelled dataset with over 72 million records includes 46 features and four attack categories.

Table 2. BoT-IoT Dataset instances

Category	Instances	Category	Instances	Category	Instances
DDoS	1926624	DoS	1650260	Reconnaissance	91082
Theft	79	Normal	477		

Proposed System

Data Preparation

- Categorical feature values are transformed into numerical values using the Label encoder function.
- Standard scaler function is used for the normalization.
- The datasets are unbalanced. Therefore, random sampling is used to make the datasets balanced.

Feature selection using hybrid ISSA-HHO: The hybridization of ISSA-HHO is as follows-

- ***Exploration Part:*** The exploration part of ISSA is modified using exploration phase of HHO algorithm.
- The trigonometric functions with HHO exploration equations are used to enhance the exploration phase of the ISSA.
- During exploration, the salps performs the global search without falling into local optima. Here, HHO exploration equations are used to perform the global search.

Proposed System

- The opposition based learning is used for the hawks population initialization. Afterward, hawks position is updated as-

$$SH(it+1) = \begin{cases} H_{rnd}(it) - rn_3 \times \sin(rn_0) \times |H_{rnd}(it) - 2rn_4\, H(it)| \;\; if\; q \geq 0.5 \\ H_{pr}(it) - H_{mean}(it) - \; rn_5 \times \cos(rn_0) \times \big(LwB + rn_6(UpB - LwB)\big) \; if\; q < 0.5 \end{cases} \quad (7)$$

where $rn_0 = \propto -it \times \frac{\alpha}{Max_{it}}$, $\propto = 2$, $H_{rnd}(it)$ is the random hawks position obtained using opposition based learning. $q, rn_3, \; rn_4, rn_5$, and rn_6 are the random numbers. $H_{pr}(it)$ is the prey location. $H(it)$ is current hawks position.

Proposed System

Exploitation: During exploitation, local search to obtain optimal solution is performed.

- The LeF is used to control the step size of follower salps. Therefore, the salp follower position is updated as-

$$SH_n^m = \frac{1}{2}(SH_n^m - SH_n^{m-1}) + LeF(Dim) \qquad (8)$$

- where SH_n^m is the current salps position and SH_n^{m-1} is the previous salp positon. $LeF(Dim)$ is levy flight function obtained from equation -
- After that, local search technique is applied to avoid local optima and improve solution.

$$LeF = \frac{ur \times \sigma}{|vr|^{\frac{1}{\beta}}}, \sigma = \left(\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times \sin\left(\frac{\beta-1}{2}\right)} \right)^{\frac{1}{\beta}}$$

- The value of $\beta = 1.5$ and ur and vr are the random values. Dim is problem dimension.

Proposed System

Classification and Detection: The features selected as provided as input to this phase for classifying intrusive traffic using the XGBoost classifier.

Prevention: In the prevention phase of IDPS, the system first detects and classifies any intrusive or abnormal traffic on the network.

- If the classified traffic is determined to be normal, it is allowed to proceed through the network without any intervention.
- However, if the traffic is classified as intrusive or abnormal, an alarm is generated to notify the end user or security personnel.
- This proposed IDPS is placed before the firewall of the end user system.
- In case of any anomaly detected, rules at the firewall are updated which helps the end user to block the incoming

Result-(BoT-IoT dataset)

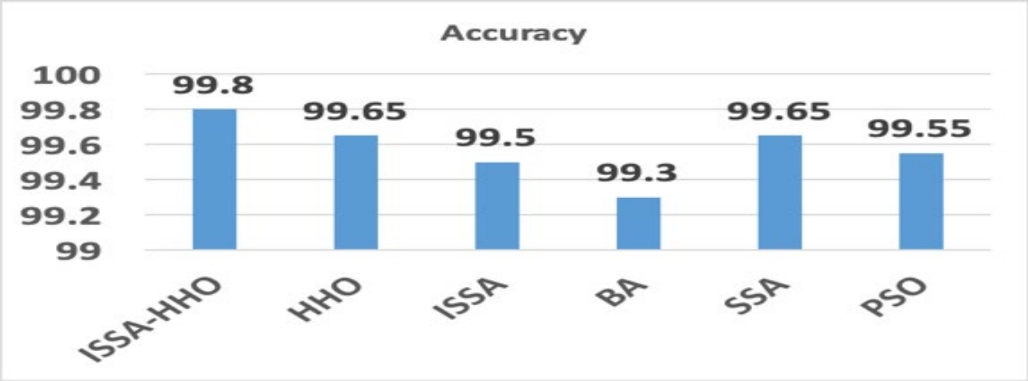

Figure 2. Accuracy

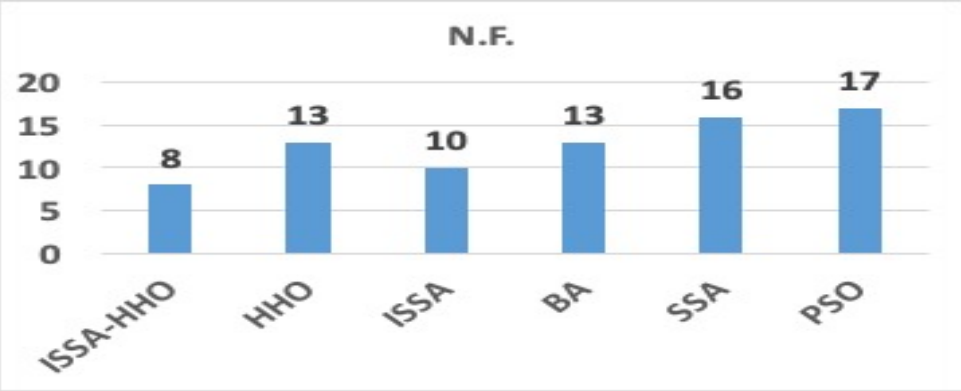

Figure 3. Number of Features

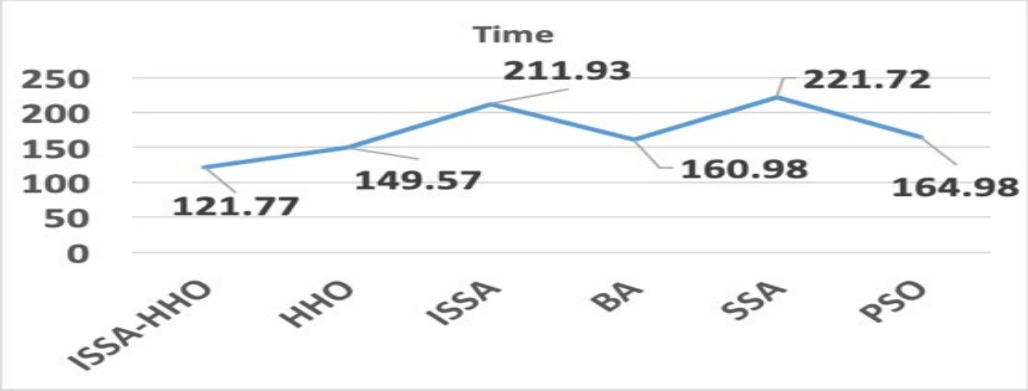

Figure 4. Time

Figure 5. Recall

Result-(BoT-IoT dataset)

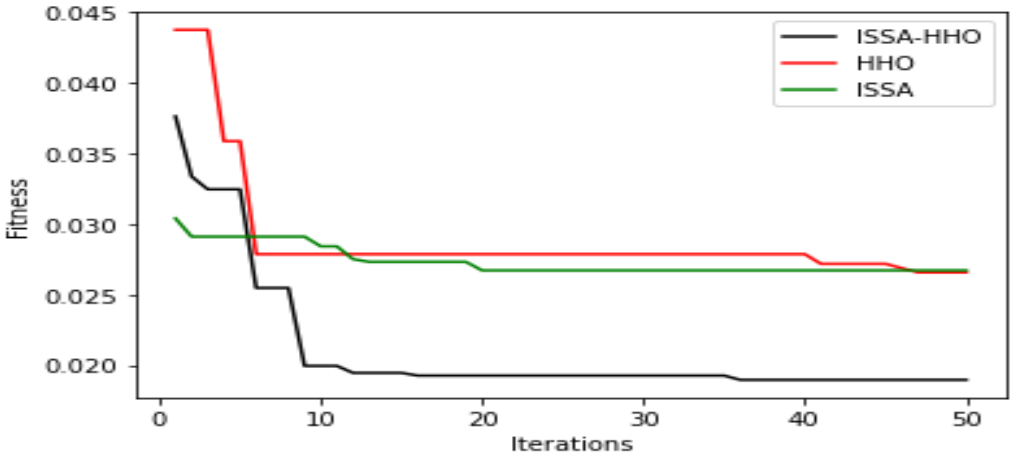

Figure 6. Convergence Curve

Result-(IoTID20 dataset)

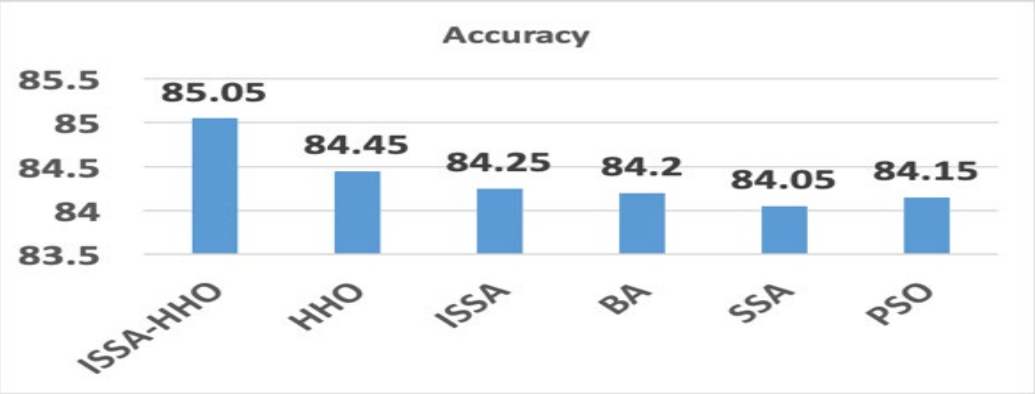

Figure 7. Accuracy

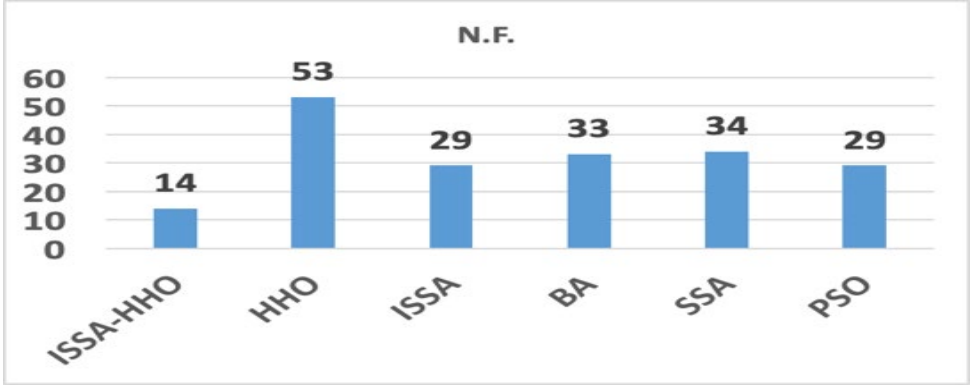

Figure 8. Number of features

Figure 9. Time

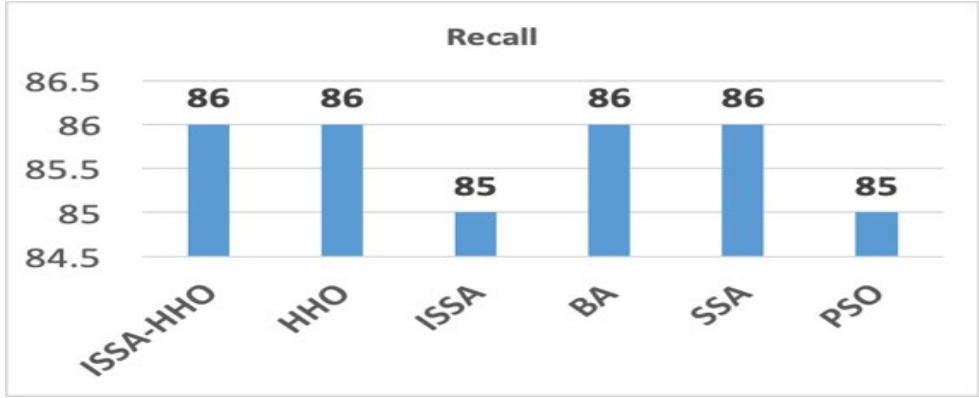

Figure 10. Recall

Result-(IoTID20 dataset)

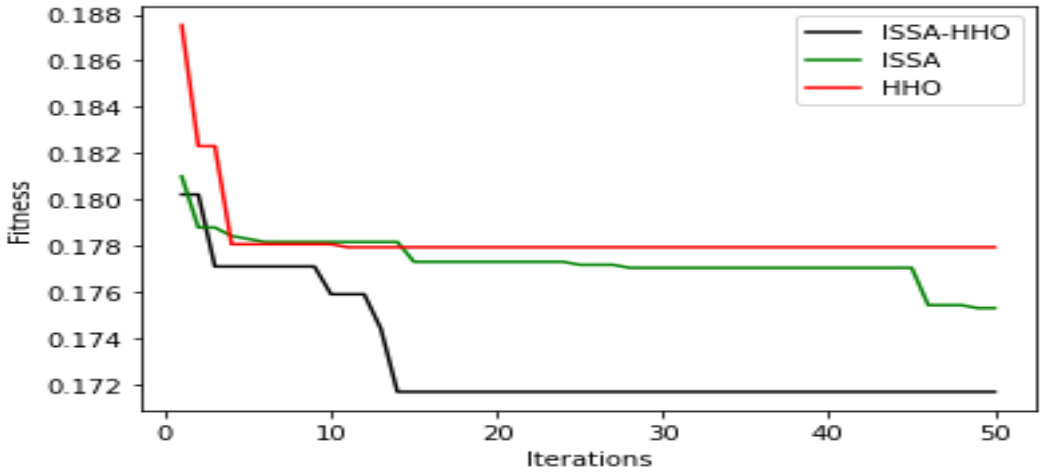

Figure 11. Convergence Curve

Conclusion

- A new approach for FS in an IoT-based IDPS is proposed by combining the improved salp swarm algorithm (ISSA) and the harris hawk optimization (HHO) algorithm.
- To evaluate the proposed system, two IoT-based datasets, BoT-IoT and IoTID20, are used.
- The hybridized FS method is compared against other bio-inspired algorithms such as HHO, ISSA, BA, SSA, and PSO.
- The outcomes indicate that the proposed system achieves superior detection accuracy with minimal computation time and selected features.
- Moreover, the hybridized ISSA-HHO-XGBoost exhibits a more rapid convergence rate than the original ISSA and HHO.

References

1. Adamopoulos,I., Ilias A., Makris, C.,Stamatiou, Y.C. Intelligent Surveillance Systems on the Internet of Things Based on Secure Applications with the IBM Cloud Platform. *International Journal on Information Technologies & Security*, vol. 15, no. 2, 2023.

2. Kumar, R.A., Franklin, J.V., Koppula, N. A Comprehensive Survey on Metaheuristic Algorithm for Feature Selection Techniques. *Materials Today: Proceedings*, 2022.

3. Kareem, S. S., Mostafa, R. R., Hashim, F. A., El-Bakry, H. M. An Effective Feature Selection Model Using Hybrid Metaheuristic Algorithms for IoT Intrusion Detection. *Sensors*, vol. 22, no. 4, 2022, p.p.1396.

4. Mafarja, M., Heidari, A. A., Habib, M., Faris , H., Thaher, T. Augmented whale feature selection for IoT attacks: Structure, analysis and applications. *Future Generation Computer Systems*, vol. 112, 2020, pp. 18-40.

5. Louk, M. H., Tama, B. A. PSO-Driven Feature Selection and Hybrid Ensemble for Network Anomaly Detection. *Big Data and Cognitive Computing*, vol. 6, no. 4, 2022.

6. Dahou, A., Elaziz, M. A., Chelloug, S. A., Awadallah, M. A., Al-Betar, M. A., Al-qaness, M. A., Forestiero, A. Intrusion Detection System for IoT Based on Deep Learning and Modified Reptile Search Algorithm. *Computational Intelligence and Neuroscience*, 2022, pp. 1-15.

References

7. Naseri, T. S., Gharehchopogh, F. S. A Feature Selection Based on the Farmland Fertility Algorithm for Improved Intrusion Detection Systems. *Journal of Network and Systems Management*, vol. 30, no. 3, 2022.

8. Singh, R., Ujjwal. R. L. Feature Selection Methods for IoT Intrusion Detection System: Comparative Study. In *Computational Intelligence. Lecture Notes in Electrical Engineering*, 2023.

9. Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., Mirjalili, S. M. Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software*, vol. 114, 2017, pp. 163-191.

10. Tubishat , M., Idris, N., Shuib, L., Abushariah, M. A., Mirjalili, S. Improved Salp Swarm Algorithm based on opposition based learning and novel local search algorithm for feature selection. *Expert Systems with Applications*, vol. 145, 2020

11. Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., Chen, H. Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, vol. 97, 2019, pp. 849-872.

12. Koroniotis, N., Moustafa, N., Sitnikova, E., Turnbull B. Towards the Development of Realistic Botnet Dataset in the Internet of Things for Network Forensic Analytics: Bot-IoT Dataset. *Future Generation Computer Systems*, vol. 100, 2019, pp. 779-796.

References

13. Ullah, I. Mahmoud, Q. H. A Scheme for Generating a Dataset for Anomalous Activity Detection in IoT Networks. In *Canadian Conference on Artificial Intelligence*, 2020.

14. Kennedy, J., Eberhart, R. Particle swarm optimization. in *Proceedings of ICNN'95 - International Conference on Neural Networks,* Perth, WA, Australia, 1995.

15. Yang, X., Gandomi, A. H. Bat algorithm: a novel approach for global engineering optimization. *Engineering Computations*, vol. 29, no. 5, 2012, pp. 464-483.

Thank You