Received: September 28, 2024. Revised: November 26, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Recognition of Threats in Hybrid Wireless Sensor Networks by Integrating Harris Hawks with Gradient Boosting Algorithm

Hussein Ali Rasool[1] Ali Hamzah Najim[2]* Mustafa Hamid Abd Alsadh[2] Hussein Muhi Hariz[3]

[1]*Altoosi University College, Najaf, Iraq Computer Science, Iraq*
[2]*Department of Computer Technical Engineering Imam Al-Kadhim University College (IKC) Al-Diwaniyah, Iraq*
[3]*Department of Computer Techniques Engineering, Mazaya University College, Dhi-Qar, Annasiriyah 64001, Iraq*
* Corresponding author's Email: alihamza@iku.edu.iq

Abstract: Due to the increasing sophistication and complexity of cyber-attacks, particularly in Hybrid Wireless Sensor Networks (HWSNs), digital community infrastructures face significant security challenges. The Gradient Boosting Machine (GBM) is known for its strong predictive capabilities in hazard identification, while Harris Hawks Optimization (HHO), inspired by hawk hunting behavior, enhances the efficient exploration and exploitation of the search space. The proposed method involves pre-processing the data to ensure cleanliness and consistency, followed by the application of HHO and GBM for threat detection, using the NSL-KDD, WSN-DS, and CIDDS-001 datasets. HHO's iterative optimization process accelerates convergence toward optimal solutions, while GBM builds a robust and accurate threat detection model. This advanced approach provides network administrators and security experts with a powerful tool to protect HWSNs from malicious activities, offering real-world applicability. With high detection accuracy and efficiency, it is well-suited to address evolving threats and ensure the availability and integrity of critical infrastructure in modern network environments. Using Python for implementation, the model achieved exceptional results, with 99.6% accuracy on NSL-KDD, 99.1% on CIDDS-001, and 98.9% on WSN-DS when HHO and GBM were combined for threat recognition in HWSNs.

Keywords: Cyberattacks, Gradient boosting algorithm, Harris hawks optimization, Malicious activity, HWSN, Threat recognition.

1. Introduction

The complex integration of traditional wireless sensor networks with new communication enhancements such as satellite communication or mobile networks, so-called hybrid wireless sensor networks can provide improved security [1], flexibility, and robustness due to aggregated HWSN network while compared to a single local competitor However, a mixture of sensor nodes and multiple communication protocols [2] creates a complex ecosystem open to unique attacks. Typical WSNs with new communication enhancements such as satellite image communication or cellular networks, known as hybrid wireless sensor networks (HWSN) Enterprise network [3] threats originate further, including fraud attempts and external factors a environmental conditions include HWSN useful features dynamic [4] and limited No search for prevention And make it more complex [5]. One of the primary issues in HWSNs is the security of communique channels [6]. Due to the wireless nature of communication, HWSNs are liable to eavesdropping, interception, and unauthorized access [7]. Criminals may additionally use protocol-primarily based communique weaknesses as an opening to acquire personal information or interfere with community capabilities [8]. Also, interoperability issues are added about by using the diverse communication technologies utilized in HWSNs, which can make the community prone to compatibility troubles and safety flaws [9]. Further, organizing well enough protection precautions into

impact is extremely hard due to the intrinsic useful resource boundaries of sensor nodes, which include low quantities of reminiscence, processing capability, and strength [10]. Traditional cryptographic algorithms might not be feasible for nodes due to resource constraints. As a result, security solutions that are energy- and lightweight-efficient and specifically designed for HWSNs must be developed [11]. HWSNs are vulnerable to threats from the environment and physical attacks along with cyber risks. Data acquired by the network may get compromised in terms of integrity and confidentiality due to physical interference or node attack [12]. In addition, environmental factors such as noise, interference, and inefficiencies can affect sensor network performance and reliability, resulting in erroneous data or network failures to detect and prevent these attacks A, the reason it is complete in terms of HWSN architecture, what methods are used for communication and required functionality [13]. HWSNs can be made more resistant to malicious activity by incorporating advanced anomaly detection techniques, machine learning, and intrusion detection algorithms [14]. Furthermore, HWSNs require that strong key management, authentication, and encryption protocols are used to protect data integrity and confidentiality [15]. Therefore, since HWSNs through the integration of sensor nodes and communication technologies are exposed to a wide range of security risks it is important to identify and combat these threats to ensure availability, [16] security and reliability of data in HWSNs By combining the computational power of the Gradient Boosting Algorithm (GBM) with the available intelligence of Harris Hawks in HWSN In the wild, Harris Hawks are known for their ability to hunt. Teamwork and rigorous investigation are used to locate and arrest the suspects. The main objective is to improve the threat detection capability of HWSNs by leveraging the predictive capabilities of GBM by mimicking the Harris Hawks collective hunting technique [17]. The behavioural characteristics of Harris Hawks hybridized with GBM offer several advantages in identifying risk factors in HWSN. In Harris Hawks' cooperative search style, people organize their actions to maximize search achievement citation and efficiency. Like this, the GBM is a reliable and accurate risk detection model with the help of combining several novel sensitivity prediction capabilities using ensemble mastering techniques. It can create an intensive threat popularity framework that could apprehend and address a whole lot of safety problems in HWSNs using combining those complementary strategies. In addition, the utility of Harris Hawks' observational skills improves HWSNs' situational consciousness, bearing in mind the early detection and mitigation of threats. Because of their terrific vision and robust spatial awareness, Harris Hawks can understand possible threats in their surroundings and stumble on minute modifications. The recommended approach will increase the energy of computational techniques in HWSNs by way of utilizing the power and robustness of herbal cognition. Through the integration of GBM's predictive modeling strategies with the collaborative-looking strategies of Harris Hawks, may additionally develop a scalable and resilient structure to guide threat prediction in HWSNs. The aim of the proposed Gradient Boosting Algorithm and Harris Hawks integration is to revolutionize the threat detection surroundings in hybrid wireless sensor networks by presenting a singular aggregate between organic and computational strategies. HHO is a nature-inspired optimization algorithm modeled after the cooperative hunting strategy of Harris hawks. These hawks exhibit a unique behavior in which they collaborate to surround and capture prey, employing both exploration and exploitation tactics to increase their hunting success. In HHO, this behavior is translated into an optimization technique where the algorithm mimics the hawks' intelligent, dynamic strategies to search for the global optimum in complex problem spaces. By adjusting the balance between exploration (searching new areas) and exploitation (refining known areas), HHO effectively navigates the solution space, making it a powerful tool for solving a wide range of optimization problems, including those in engineering, artificial intelligence, and cybersecurity.

The problem statement shows that the traditional security features often depended on rule-primarily based systems and conventional intrusion detection methods. However, the evolving nature of cyber threats posed giant challenges for these techniques. Conventional techniques struggled to evolve to the continuously changing hazard landscape and lacked the sophistication required to efficiently discover and mitigate complicated attacks, which include Distributed Denial of Service conditions. As a result, there may be a growing demand for advanced and flexible safety structures able to efficiently address the complexity and variety of contemporary cyber threats. To address this, the proposed approach makes a speciality of the Identification of Threats in HWSNs by way of combining the Gradient Boosting Algorithm with HHO. By making use of algorithms for optimization and device mastering, this method attempts to improve safety protocols in HWSNs and offers an efficient response to the converting threat situation.

The key Contributions are,

- Harris Hawks Optimization (HHO) and Gradient Boosting Machine (GBM) proposal for threat recognition analysis is the focus of this research. NSL-KDD is used as the benchmark dataset because of the diverse attack types it contains. It has a range of attacks, which makes it suitable to use in assessing the effectiveness of threat detection in real time.
- The research applies Data cleaning; it identifies and rectifies errors, treats missing data, and eradicates duplicated records enhancing the credibility of the dataset. Feature scaling techniques of Min-Max normalization and Standardization guarantee that the normalized values always fall within an ideal ideal range to boost up the general reliability and compatibility of a model.
- In this research, Gradient Boosting, a highly effective ensemble method, is used to categorize possible threats through creating models with decreased probabilities of error in consecutive levels.
- The HHO algorithm is used in the study imitates the hunting behavior of hawks and enhances the search processes with the appropriate balance between exploration and exploitation. This metaheuristic approach optimally updates the hawk positions through leader positions, probability criteria and fitness function.

The flow of this proposed work is constructed as follows. Section 2 includes earlier research for Threat Recognition in HWSN. Section 3 discussed the problematic statement. Section 4 discussed about the Proposed Hawks GBM system for Threat Recognition in HWSN. Section 5 presents the outcomes and discussion of the Proposed HawksGBM. Finally, section 6 carries the conclusion of the paper.

2. Related works

Ragab et al [18] proposed a method for identifying Distributed Denial of Service (DDoS) crimes in IoT contexts, called the Piecewise Harris Hawks Optimizer with an Optimal Deep Learning Classifier (PHHO-ODLC). The PHHO-ODLC method functions primarily through a three-step, painstakingly built process, with each stage purposefully created to handle important areas of detection of distributed denial of service and categorization inside the Internet of Things networks. To improve classification accuracy, the algorithm first uses PHHO to carefully sort through the multitude of accessible features and identify the most relevant ones. The PHHO algorithm may be used initially to choose pertinent characteristics and improve classification performance. The DDoS attack classification algorithm can then be implemented through the ABiLS™ network. Finally, the GWO is used to determine the hyperparameters of the ABiLS™ network. The main disadvantage seems to be the need to examine the robustness of the PHHO-ODLC approach to large IoT networks and attacks to further enhance IoT security, systems that evolve can respond to changing DDoS attack techniques and incorporating anomaly detection techniques is important.

An enhanced Harris Hawks optimization (HHO) was developed by Zhang et al [19]. Research to find superior solutions for feature selection and global optimization problemsAn enhanced Harris Hawks optimization (HHO) was developed by Zhang et al. Research to find superior solutions for feature selection and global optimization problems. This technique is an effective optimizer that was motivated by how Harris' hawks attempt to capture rabbits. The initial version occasionally tends to stall at the local optimal solutions. Due to this, a new HHO known as IHHO is developed, which expands the application domains and enhances the optimizer's search capability by incorporating the salp swarm algorithm (SSA) within the original HHO. The three phases of the HHO optimizer's update stage, which is carried out to update each hawk, are: creating hybrid individuals based on HHO-based individuals and SSA-based individuals; updating the search agent considering HHO's mechanisms and greedy selection; and modifying the population based on SSA to produce SSA-based population. To test the effectiveness of the suggested optimizer, a sizable number of experiments on a variety of functions are conducted. According to the overall findings, the suggested IHHO will preserve a better balance between exploration and extraction while offering a faster rate of convergence. Furthermore, a more robust binary IHHO is also built as a wrapper-based FS strategy by the suggested continuous IHHO. Furthermore, the study uses popular benchmark datasets from UCI to compare the binary IHHO that is produced with other FS techniques. The experimental findings show that compared to existing wrapper FS approaches, the suggested IHHO has higher accuracy rates. Overall investigation and analysis support the increase in IHHO as a result of SSA's appropriate exploration capacity.

Systems for detecting intrusions are designed to identify various types of assaults that are outside the

firewall's capability. The IDS categorizes the system's typical and unusual characteristics based on its properties. Numerous machine learning-based intrusion detection systems have been developed thus far. To improve classification performance, the feature selection procedure is crucial. To choose the best features, a DL-based feature selection process is therefore provided in the study work of Simon et al. [20] To categorize deep characteristics and identify assaults in the Internet of Things network, the decision tree method is employed as a classifier. The decision tree algorithm and CNN are combined to develop the proposed IDS. In the proposed model vs the traditional model comparison, the evaluation will be conducted as validation decision-making. Many researchers use the benchmark NSL-KDD set to investigate the relationship between security intrusion detection and data visualization. The decision tree algorithm and CNN are combined to develop the proposed IDS. The proposed model and the traditional model are compared through a validation decision. The benchmark NSL-KDD data set has been used for experimentation. The proposed design uses a decision tree approach as a classifier instead of an absolute neural network.

The study by Lee et al. [21] presents a new penetration predictor that combines Temporal Convolutional Network (TCN) and Attention-CNN-BiLS™ (ACBL) frameworks. When it comes to analyzing the temporal and geographical features of network traffic data, the ACBL and TCN models perform exceptionally well. This integration using neural network architectures improves both model accuracy and performance. Moreover, a new method based on the biological behaviour of dung beetles and the use of TDBO is used to find optimal model hyper criteria and optimize feature selection criteria to improve the performance of the model, using priority ranking of Random Forest algorithm combined with feature selection criteria obtained by TDBO. This ensures the best selection. The combination of the TDBO algorithm and the ACBLT model adds complexity to the overall intrusion detection system. The study uses the UNSW-NW15 data set to evaluate the performance of another ID model, the TDBO-ACBLT model. When compared to other algorithms, TDBO performs exceptionally well in feature selection in terms of parameter optimization accuracy. Compared to popular ML models, the accuracy of the suggested model is greater. The main drawback is during implementation, especially in resource-constrained environments where computational resources are limited.

Almuqren et al [21] introduced the Gradient-Based Optimizing using Hybrid Deep Learning (BSSHN-GBOHDL) approach, which is a Blockchain-Assisted Secured Home Automation System designed to detect harmful behaviors in home automation settings. The BSSHN-GBOHDL approach makes use of an aggregate of records preliminary processing, hybrid neural network (HDL)-based totally pastime categorizing, and Gradient-Based Optimizer (GBO)-primarily based hyperparameter optimization to stumble on illicit sports. It also uses blockchain technological advances to enhance records privacy. The efficacy of the BSSHN-GBOHDL approach is proven by trying out the statistics set NSL-KDD, reaching the most pleasurable accuracy of 98%. This demonstrates the effectiveness of integrating BC with DL-based pastime popularity in home automation, outperforming exceptional current techniques. The fundamental contribution is the improvement of a decentralized, secure, and protected machine that makes use of devices network intelligence to successfully emerge as aware of and neutralize dangers. The BSSHN-GBOHDL model protects user privacy, strengthens the defense architecture of home automation networks, and fosters teamwork in the face of emerging security risks by incorporating BC technology. All things considered, the method presents a viable way to improve the privacy and safety of smart home settings. While blockchain technology can enhance data confidentiality through encryption and decentralized storage, it also raises privacy concerns. All data stored on the blockchain is immutable and visible to all network participants, potentially compromising the privacy of smart home users.

Ali et al [23] utilized dense convolutional neural networks to tackle IoT access concerns, specifically Distributed Denial of Service (DDoS) cyberattacks. These networks use evolutionary approaches to optimize their network efficiency by learning from the collection of intrusion data and unusual activity to recognize and monitor threats. By using neurons to evaluate complex hypotheses, the sparse network reduces the amount of interference involved in the distribution of IoT information, the attempt to distinguish between threat and standard patterns using artificial neural networks through a training pattern was better than the typical defense of a computer network. IGA-BP stands for Leakage Detection System and Active Networks, where the network-based approach depends on auto-encoder networks and improved genetic algorithms to detect infiltration and leakages. MATLAB is used for confirmation of proper algorithm execution which provides as good as 98.98% detection rate and consequently up to 99.29% accuracy of the procedure

with the minimum amount of processing resources. Data acquisition from the internet (IoT) using the DDoS Attack dataset is based on the current framework which can identify the threats while detecting the attack. Attributes which are the ones extracted and the development of sparse matrices process from the training and validation phases as well, improve the ability to recognize the activities in the stage of detecting. The price of detection (98.98%) and accuracy (99.29%) are going to be avoided by the application MATLAB at the price of low calculation complexity. Both strategies demonstrate how well evolutionary algorithms and sparse convolution networks can identify and mitigate IoT intrusion hazards, reducing computing complexity and errors while achieving excellent detection rates and accuracy. The computational speed of the proposed methods may not be optimal for real-time intrusion detection in high-speed IoT networks. Processing large volumes of data in real time requires efficient algorithms and hardware resources to meet strict latency requirements.

Elhariri et al [24] utilized two primary components of the implied approach to breaking severity detection in structural health monitoring (SHM) are deciding on features utilizing a hybrid filter-wrapper via multipurpose enhanced salp swarm optimization and extracting features through constructed feature development and CNN-based deep feature training. Ten UCI data and 4 datasets containing crack image information are used to train and verify the system. In comparison to traditional classification methods, experimental results show substantial gains in crack quality identification performance, with rises in detection average precision and the F-value of about 37% and 31%, respectively. Furthermore, when compared to employing the complete set of features, the suggested method yields an impressive reduction rate of about 67 per cent in the selected feature collection across all evaluated datasets. The suggested method performs better in terms of both computational time and decreased feature rate. The results demonstrate improvements in acceptable, slight, and extreme crack identification, respectively, when employing VGG16 learned features as opposed to fused hand-crafted features. The research is significant because it explores and illustrates the effects of multi-feature reduction in dimensionality using hybrid filter-wrapper and multi-objective optimization techniques for feature selection, with a focus on fracture severity identification in SHM.

The current approaches provide useful frameworks for identifying and classifying anomalies inside the framework, they are constrained by their incapacity to adapt to different attack patterns. It has been shown that the ML-based IDS is a useful method for safeguarding UAV networks because network security is such a significant concern for them. Furthermore, there are additional difficulties because of the dataset that was utilized as well as the processing and memory demands of the current models. The fact that these models use less processing power and memory is one of the primary issues. Thus, this difficulty is addressed in the Cengiz et al. [25] presented in this study. To lessen the dimensionality of the UAV Attack Dataset, a novel dimensional reduction method based on correlation coefficients, information gain, and PCA is presented. Then, a new input recognition algorithm based on genetic algorithm (GA) and ANN is proposed. Optimal artificial weights are generated using genetic programs. The convergence and prediction accuracy of the proposed model are compared with the surface diffusion network and its version. According to this comparison, the proposed model outperforms the other classifiers by at least 6% in terms of prediction accuracy and time efficiency. IDS can be used to detect and stop UAV-related data leaks. This can be done by using an IDS to monitor the UAV communications network and detect any data extraction attempts.

Previous research on security measures for IoT networks, systems health monitoring (SHM), and UAV systems typically relies on rule-based algorithms and traditional intrusion detection methods though great difficulties arise Conventional techniques frequently lack the sophistication required to reliably identify and counteract complex attacks like attacks in SHM systems and UAV networks, or Distributed Denial of Service conditions. As a result, the need for improved and flexible security systems that might efficaciously handle the complexity of present-day cyber threats became turning into increasingly apparent. This consciousness brought on the use of system understandings (ML), optimization algorithms, and emerging technologies inclusive of blockchain to enhance safety features across more than one domain. The proposed approach improves detection in hybrid wireless sensor networks employing the predatory style of Harris hawks and the strength of the Gradient Boosting Algorithm. Through the combination of powerful devices gaining knowledge of methods with optimization strategies, this approach offers a feasible resolution to the problems presented by way of constantly evolving cyber threats. It aims to strengthen the safety of networked devices with the aid of growing intrusion detection structures' accuracy and efficiency through an included technique.

Table 1. Limitations of Intrusion Detection Algorithms and Systems

References	Algorithm	Year	Limitation
[18]	PHHO-ODLC (Piecewise Harris Hawks Optimizer with an Optimal Deep Learning Classifier)	2023	The sensitivity of ABiLSTM network performance to hyperparameters may require extensive experimentation and computational resources.
[25]	Improved HHO (IHHO)	2021	Integration of the Salp Swarm Algorithm (SSA) into the HHO framework adds complexity and managing interactions between the two algorithms may introduce overhead in terms of implementation and computational resources.
[26]	Customized Machine Learning Approaches for IoT Intrusion Detection	2024	The susceptibility of ML-based detection systems to adversarial attacks is not addressed, which could be a significant concern in real-world settings.
[21]	TDBO-ACBLT (Temporal Dependency-based Optimization - Adaptive Convolutional-BiLSTM-Tensor Fusion)	2024	The integration of TDBO algorithm and ACBLT model adds complexity to the intrusion detection system, potentially resulting in implementation challenges, especially in resource-constrained environments.
[22]	BSSHN-GBOHDL (Blockchain-Assisted Secured Home Automation System with Gradient-Based Optimizing using Hybrid Deep Learning)	2023	While blockchain technology enhances data privacy, it also raises concerns regarding the immutability and visibility of data stored on the blockchain, potentially compromising smart home user privacy.
[23]	Evolutionary Sparse Convolution Network (ESCNN)	2022	The computational speed of proposed methods may not be optimal for real-time intrusion detection in high-speed IoT networks, requiring efficient algorithms and hardware resources to meet strict latency requirements.
[24]	Hybrid Filter-Wrapper with Multipurpose Enhanced Salp Swarm Optimization	2020	The method adds complexity to the intrusion detection system, potentially leading to challenges during implementation, especially in resource-constrained environments.
[24]	IDS based on Machine Learning Techniques for UAV Systems	2024	While the proposed model is efficient, it may not address all security vulnerabilities in UAV systems, requiring continual updates and improvements to adapt to evolving threats.

Table 1 shows the Limitations of Intrusion Detection Algorithms and Systems.

3. Hybrid threat recognition with Harris hawks and gradient boosting

The proposed method integrates the satisfactory capabilities of HHO and Gradient Boosting, to cope with the complex difficulties of attack recognition in modern community environments. This aggregate method combines superior algorithms for ML with optimization stimulated by nature to offer an integrated reaction to the changing cyber security danger scenario. By simulating the dynamics of prey seizure and institution collaboration, HHO efficiently guides the search system toward foremost solutions, making it specifically nicely desirable for complex optimization troubles. Complementing HHO, Gradient Boosting is a powerful ensemble learning technique that combines the predictive capabilities of multiple weak learners to build a robust and accurate threat detection model. Fig. 1 shows the Framework of the Proposed Hawks-GBM System.

3.1 Data collection

The designated network attack detection database known as NSL-KDD is used in several assignments to evaluate various machine learning-based

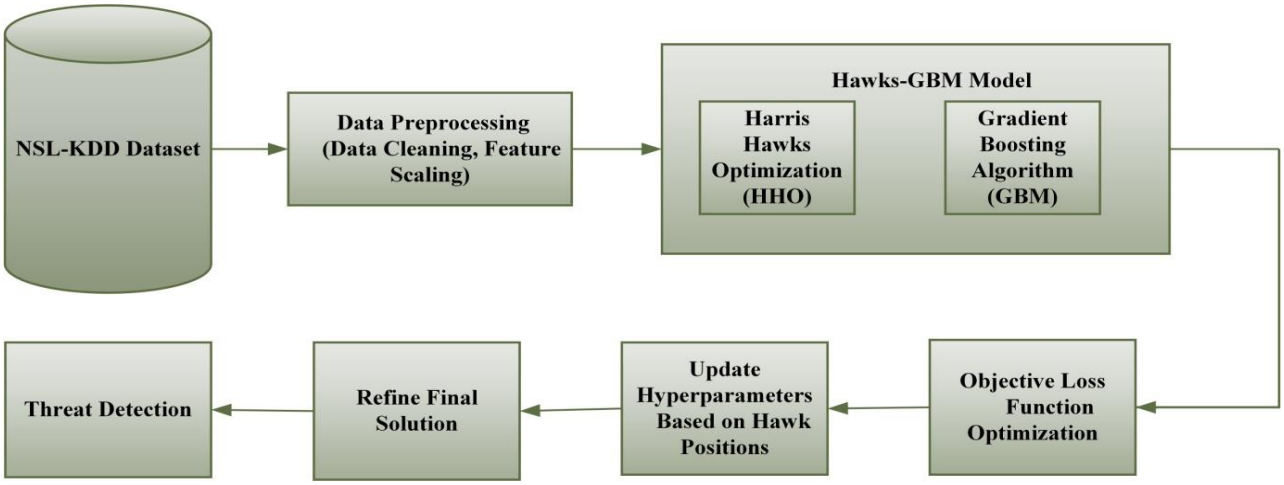

Figure. 1 Recognition of Threats using the Hawks-GBM system

methodologies for generating various threat detection strategies [27]. For the 41 characteristics in the NSL-KDD dataset, four categories of characteristics may be identified: basic, time-based, content-based, and host-based traffic aspects. The primary determinant of these characteristics' value is their distinct, enduring, and symbolic nature. Within the NSL-KDD dataset, instances are labelled into five distinct classes, facilitating the classification of network traffic into different attack categories. These classes include Normal, Denial-of-Service (DoS), Root-to-local (R2L), Probe, and Unauthorized Root (U2R) attacks. Ranging from general anomalies to specific intrusion attempts targeting system vulnerabilities each attack class represents a unique threat scenario.

NSL-KDD is the expanded version of KDD'99, optimized for overcoming its flaws and providing a more accurate assessment of IDS efficiency in network space. It includes several files in different formats: Since QN was based on two data sets of different sizes, and the numbers of positive samples in Ti and Di are not given, KDDTrain+. ARFF` and `KDDTrain+. TXT denotes the set of all training instances and vectors of binary and attack-type labels. For a rather smaller subset, KDDTrain+_20Percent. ARFF` and `KDDTrain+_20Percent. TXT provides 20% of the total training data curated in ARFF and CSV formats. Similarly, `KDDTest+. ARFF` and `KDDTest+. IMT` represents the full test set and `KDDTest-21. ARFF` and `KDDTest-21. Instead of considering records as Text or as TXT, it would exclude texts of high difficulty level to consider files that give less difficulty.

The NSL-KDD dataset is more advanced than the KDD'99 data since it filters out duplications in the training data set which may cause bias towards particular classifiers. It also makes sure that there are no repeated records in the test sets so that the results are not skewed and gives a better evaluation of the detection methods. Furthermore, it will be easier to achieve various classification performances since the selection of records in this dataset covers records according to the difficulty level. It means that in creating the final overall table, the employment of this format of a comprehensive and balanced dataset enables formulaic and like-for-like results to be arrived at from one research to another.

The NSL-KDD dataset is used for testing and training, providing standards by which to measure the suggested system. The suggested system's detection performance is further evaluated using the NSL-KDD dataset as a benchmark, employing semi-supervised machine learning algorithms for a class of assaults with 42 characteristics and classification labels. Forty-one attributes are classified into content, host, traffic, and basic features. The dataset has a total record of 148,515 samples sectioned into 80% of training and 20% of testing samples, with four different classes of attacks. The dataset is divided into normal and pathological clusters in order to extract the vector characteristics for training. The vector characteristics are obtained for categorisation as normal and pathological clusters following training. Four groups of attacks—denial of service attacks, R2L, U2R, along with probe category—are grouped into the dataset, which has 23 classes of attack types. The DoS attack prevents authorised users from accessing the network and causes the network service to become congested. The U2R attack sniffs the genuine user's credentials to introduce vulnerabilities into the host system. The R2L remotely introduces flaws into the network host's system. In violation of the security rule, the probe attack searches the network for data extraction and collection. While the other attacks have single linkages, the probe along with DoS assaults has numerous links. [28] Explains

the four attack classes seen in the NSL-KDD benchmark database.

CIDDS-001 (Coburg Intrusion Detection Data Set): An IDS-based data set that was prepared from an academic data set for IDS evaluation. The data was prepared by researchers at HS-Coburg in Germany who developed network flow data in a simulated virtual environment of both good and bad network traffics and supports the evaluation of IDS based on anomaly detection and signature detection. It is available in several versions: V1 is the raw data of CSV format, V2 has cleaned parquet files of correct data types and free of missing records, while V3 organizes it for efficient storage. This is one of the most significant datasets used in research by Markus Ring, Sarah Wunderlich, Dominik Grüdl, Dr. Dieter Landes, and Dr. Andreas Hotho. Researchers using this CIDDS-001 dataset are recommended to cite this work by the original authors in their publication [29].

This dataset is meant for intrusion detection systems generally, and particularly in hybrid IoT and WSN scenarios. In the dataset, information exists on various types of Denial of Service attacks interfering with regular communication and reducing network performance, including Blackhole, Grayhole, Flooding, and Scheduling attacks. The dataset captures different features of these attacks, and this enables researchers to develop machine learning models for real-time threat detection. The threats are crucial to identify in hybrid WSNs since they are the backbone of IoT systems. Here, the compromised nodes or communication breakdown can impact a wide range of applications. WSN-DS provides a basis for enhancing security frameworks in IoT-based WSNs [30].

3.2 Pre-processing using data cleaning and feature scaling

Using HHO and GBM for risk reputation in HWSNs, the NSL-KDD dataset needs to be pre-processed to make sure that the facts are easy, steady, and accurately developed. The technique of identifying and correcting mistakes, missing computation, or anomalies inside the dataset is known as data cleansing. Data cleaning guarantees that the dataset is free of errors and noise when used for hazard reputation in HWSNs. The dataset is improved by the techniques like suggest, median, or mode imputation. Removing duplicates to eliminate redundancy will enhance a particular statistic's integrity. Outliers get separated, and they are not granted the ability to skew the tests thus leading to balanced quality in the assessment.

Feature scaling is also vital in the data pre-processing phase, which involves scaling numerical features to the same range. As for the NSL-KDD data set, characteristic scaling would include Min-Max Normalization and Standardization amongst other normalization techniques. Min-Max Normalization encompasses scaling numerical capabilities to a standardized range, typically between zero and 1, which may be done through the usage of the following Eq. (1).

$$K_{normalized} = \frac{K - K_{min}}{K_{max} - K_{min}} \quad (1)$$

Where K represents the initial value of the feature. Standardization adjusts the features to have a mean of 0 and a standard deviation of 1. It ensures the numerical features in the NSL-KDD dataset are appropriately scaled, preventing any single feature from dominating others during model training.

The data processing and data cleaning exercises were performed to refine the dataset that was used in the training of the various models as well as in the evaluation of their performance. Firstly, the collected dataset was prepared for analysis by cleaning data and, if necessary, handling missing and isolating extraneous records. About missing values, records that contained such values whereby some or all of the variables in a record were missing, these records were either deleted or had the missing values imputed statistically. It was important to do this to preserve the dataset and proceeding quantification and analysis with complete data.

After the cleaning step normalization was done using the min-max normalization technique to normalize all the features. Normalized scaling transformed all the features within the range of 0 to 1 making it possible to enhance the performance of machine learning algorithms that are vulnerable to a range of input features. By scaling the feature values, the diverging variations in features were reduced and the convergence of the model was promoted as well as the effects from features to learning became better balanced. Apart from enhancing the performance of the model, this step was especially helpful for achieving the computational reproducibility of the obtained outcomes because it maintained the data consistency of the pre-processed data across the experiments carried out.

3.3 Feature selection using principal component analysis (PCA)

To increase the performance of the model, and to bring down the dimensionality, Feature selection

using PCA was incorporated. The PCA is a statistical method that provides a new set of variables, called principal variables or simply principal components which are linear combinations of the original variables and that are uncorrelated between them and whose variances are maximized. In other words, through the use of PCA, we selected several attributes that lay in the hyperplane of the principal components that have most of the variation in it to reduce the number of attributes significantly but in equal measure retain all key information. This approach helped also to clear the model and avoid many problems of multicollinearity between the features.

In our experiments the possible attributes or features were chosen to be as follows, duration, protocol type, service, flag, and various features of the network traffic. The importance of these features as well as selecting the most significant principal components was done through PCA. As a result of dimensionality reduction, we were able to focus on the most relevant features, which had the most influence on threat detection, thus subsequently decreasing computational cost and increasing the efficiency of the considered model. This process of selection and transformation guaranteed that the type of model that was being developed was both optimized and effective, based on the most relevant variables that would thereby enhance its performance in the detection of threats in HWSNs.

3.4 Recognition of threat with gradient boosting algorithm

Gradient boosting is a widely used ensemble learning technique for classification and regression problems. In threat recognition in HWSNs, GB can be leveraged to effectively identify and classify potential threats in HSWN. The fundamental principle of GB involves iteratively constructing an ensemble model, typically comprised of weak learners, such as decision trees and they are vital musical instruments in contemporary security operations because of their capacity to adjust to shifting surroundings and absorb new information. Fig. 2 represents the Architecture of GBM.

The goal is to optimize an objective loss function that captures the discrepancy between predicted outputs and actual targets. It might be the average squared error (MSE) for regression issues and the loss of cross-entropy for classification problems. The goal function will be represented as *A(j), B(i)),* where j stands for the true labels and *B(i)* for the ensemble's current forecast. Determine the objective function's negative gradient about the current prediction at each iteration. The subsequent weak learner who is introduced to the ensemble is guided by this gradient. It is given in Eq. (2),

$$q_a = -\frac{\partial \mathrm{A(j,B(i))}}{\partial\ \mathrm{B(i)}} \tag{2}$$

It instructs a choice branch or other poor-performing learner to suit the negative gradient, q_a. The weak learner is represented by the notation $W_a(i)$. Add the weak learner to the ensemble by multiplying its learning rate (γ) by a reduction parameter to update the ensemble and represented by Eq. (3) [31],

$$B(i) = B(i-1) - \gamma g_n \tag{3}$$

Where $B(i)$ indicates the current model or function at iteration n. $B(i-1)$ represents the model or function at the previous iteration $(i-1)$. γ indicates the step size or learning rate that controls the magnitude of the update. It's often a small positive value, helping control the extent of change from one iteration to the next. gn denotes the gradient or a directional derivative that points in the direction where the objective function decreases most steeply.

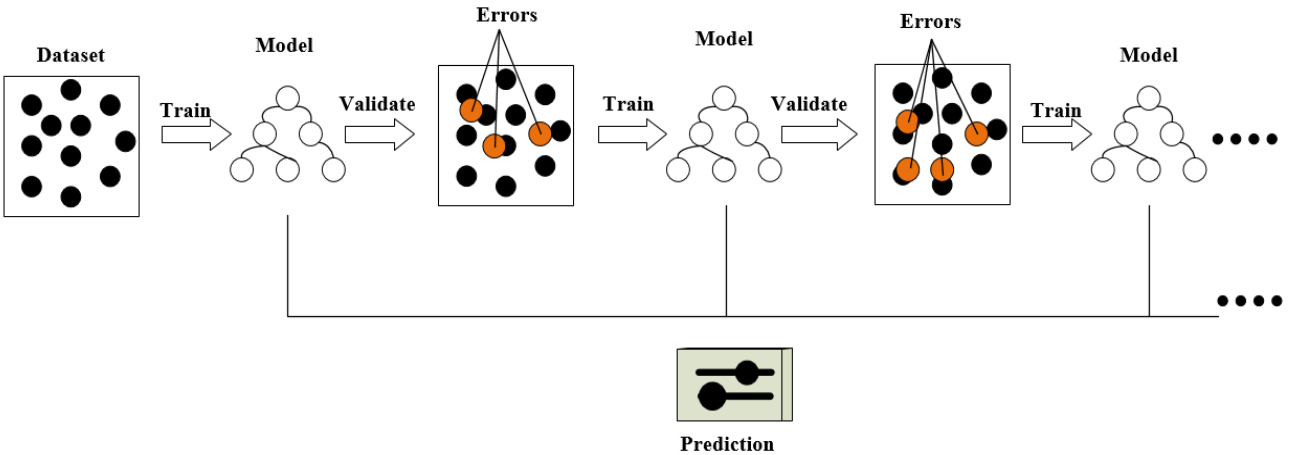

Figure. 2 GBM Architecture

This gradient represents the error or residual of the model at iteration n, guiding the model update.

Once the Gradient Boosting model is trained, it can be used to detect and classify potential threats in real-time data streams. The model uses network activity or incoming sensor measurements to find abnormalities or suspicious trends that could point to the existence of a danger. New data may be added to the gradient-boosting model continually. When threat scenarios change, the model uses adaptive learning approaches to update its predictions and modify its parameters.

3.5 Harris hawks optimization in hawks-GBM system

An optimization technique called Harris Hawks Optimization employs an array of hawks to search for responses inner an examination area. These hawks span a spectrum of ability answers and are placed randomly across the seek space. Hawks use exploitation and exploration techniques to both decorate and explore their modern regions as well as discover unknown territories. The related solution's fitness or closing feature fee determines which hawk is on the pinnacle, or the one in command. Then, relying on some of the variables, which include the leader's position and predetermined criteria for investigation and exploitation, the Hawks adjust their locations. Efficient navigation and convergence toward the most effective solutions are ensured via this equilibrium. Hawks adapt their postures and explore/take advantage of the search space often as part of an iterative optimization procedure. The goal of every iteration is to enhance capacity solutions and hone the population closer to advanced alternatives.

Establish an assortment of hawks in the initial ranges, each of which stands for a capacity answer inside the Hawk-GBM model's hyperparameter space. The hawk's exploration vector, which depicts haphazardly exploring the hunt space and is calculated in Eq. (4),

$$H_{exploration} = p_{rand} - p_{hawk} \tag{4}$$

Where P_{rand} is a solution vector that was created at random. Determine every hawk's exploitation vector, which denotes the exploitation of favourable areas and is determined in Eq. (5),

$$H_{exploitation} = p_{best} - p_{hawk} \tag{5}$$

Where p_{best} is the position of the best solution vector. Eq. (6) represents a model where the probability of a "hawk" (an agent or entity choosing an aggressive or risk-taking strategy) is influenced by both exploration and exploitation components. Here, p_{hawk} denotes the initial probability of a hawk strategy, while $H_{exploration}$ and $H_{exploitation}$ represent adjustments based on exploratory and exploitative behaviours, respectively. In this context, exploration encourages trying new or diverse actions, which can reveal unknown opportunities, while exploitation focuses on maximizing known, advantageous outcomes. This balance helps agents adaptively update their probability of aggressive (hawk-like) behaviour, enhancing decision-making by blending both risk-seeking (exploration) and risk-averse (exploitation) strategies,

$$p_{hawk} = p_{hawk} + H_{exploration} + H_{exploitation} \tag{6}$$

The locations of the hawks in the hyperparameter space are used to update the Hawk-GBM model's hyperparameters or settings, facilitating the detection of anomalies and attacks. Utilizing the improved hyperparameters that were acquired during the HHO optimization procedure, train the Hawk-GBM model and apply the Hawk-GBM model that has been trained to the ensemble model.

The Hawk-GBM model's optimization procedure is done by iteratively repeating a process for a predefined number of iterations or until convergence conditions are satisfied. By using an iterative process, the hyperparameter space may be continuously explored and exploited, leading the optimization process towards optimum or nearly ideal solutions to detect anomalies and attacks effectively. When the convergence conditions are met for instance, by reaching a maximum number of iterations or witnessing minimal progress in further iterations, the optimization process comes to an end. The process continues with the visualization of the output of the Hawk-GBM model after optimization where the generation of the best solution is again improved and then analyzed using evaluation metrics and appraisal methods to evaluate if it gave a good result. The optimized Hawk-GBM model is a tool, therefore it can make better identification of abnormalities and assaults in the real world. By using HHO (Hawk Genetic Algorithm) based on nature, this model brings to the table a new optimization method of hyperparameters which helps to improve the model performance in hazard prediction tasks.

Algorithm for Hawks-GBM System

Input: NSL-KDD dataset

Output: Recognition of Threat in HSWN Using Hawks-GBM System

Start

Load the input image

Perform preprocessing operations using Data Cleaning and Feature Scaling

Initialize Hawks-GBM Model

Perform HHO to explore and exploit the search space for optimal solutions
Update Hawk positions based on explorations and exploitations

Implement GBM to construct an ensemble model of weak learners in threat detection

Optimize an Objective loss function to minimize prediction errors

Iterate the optimization process until convergence conditions are met

Update hyperparameters based on the positions of Hawks in the hyperparameter space

Refine the final solution obtained by the Hawks-GBM Model

Detection of threat

End

4. Results and discussion

The study results on "Recognition of Threats in Hybrid Wireless Sensor Networks by Integrating Harris Hawks with Gradient Boosting Algorithm" show that the proposed model has better performance than various data sets. By combining the exploration-exploitation features of Harris Hawks Optimization with the robust classification characteristic of Gradient Boosting, good accuracy, precision, recall, and F1 score are attained for the tested datasets of NSL-KDD, CIDDS-001, and WSN-DS. It performed better than other approaches in recognizing threats within hybrid wireless sensor networks and stable robustness for many metrics. The results are given below.

Fig. 3 shows the accuracy curve of the proposed Hawkes-GBM model in representing the combined performance of the HHO and GBM models, respectively. Training accuracy measures how the model performs on the training dataset **NSL-KDD**, meaning that the proposed Hawks-GBM is learned from the given data Validation accuracy provides an estimate of the predicted performance of the model under normal circumstances work in Evaluating the model's generalization performance on unseen data, usually a separate test dataset.

The loss curve for the proposed Hawks-GBM version in Fig. 4 shows the evolution of the loss characteristics of the version in successive iterations or epochs throughout the optimization process This curve provides valuable insights value in the learning efficiency and convergence behavior of HHO and GBM added versions NSL-KDD. Understanding the relationship between optimization iterations and loss reduction can better test over-parameters, change learning costs, and improve educational design to provide the prediction accuracy and reliability of the Hawks-GBM version to the sky.

Fig. 5 shows "Training and Validation Accuracy for CIDDS-001, training as well as validation accuracy curves are showing a rising trend; that is, the quality of the model improves at every training step. However, in the validation accuracy curve it remains flat around the 40th epoch, which also means further training might not improve the generalization capability of the model. It might be overfitting the model, which learns too much about the training data and fails to perform well on unseen data. The above might be resolved through regularization, data augmentation, or reducing model complexity.

Fig. 6 shows "Training and Validation Loss for CIDDS-001". The model is learned over 50 epochs, and as can be seen, both training and validation losses go down. In other words, it means that the model has learned well; however, it decreases more drastically for training loss than validation loss and stabilizes after 20 epochs while fluctuating for validation loss before leveling off. This gap between the two curves can indicate that the model is overfitting, the model is over-specialized to fit the training data but does not generalize well to unseen data. Techniques to overcome such issues include regularization, data augmentation, or simply reducing model complexity for better performance and reduction of overfitting.

Fig. 7 shows "Training and Validation Accuracy for WSN-DS," plots accuracy in the model over 50 epochs. Both curves-the training and validation accuracy curve-are increasing. That is, the model does seem to be improving accuracy with training. Yet again, the validation accuracy curve flattened out at the 40th epoch, which indicated further training might not help to improve the generalization ability.

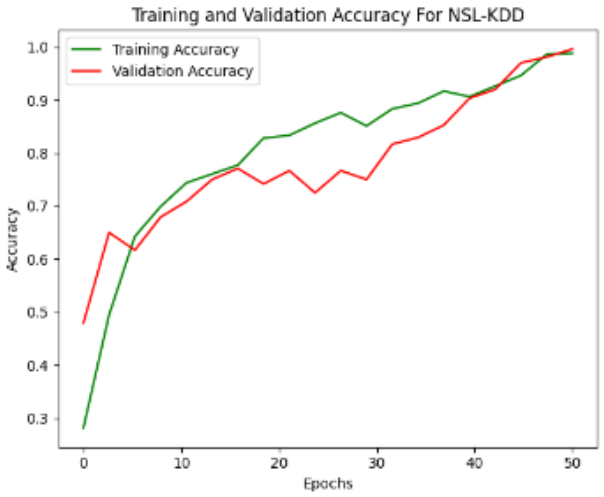

Figure. 3 Accuracy Curve for Proposed Hawks-GBM NSL-KDD Model

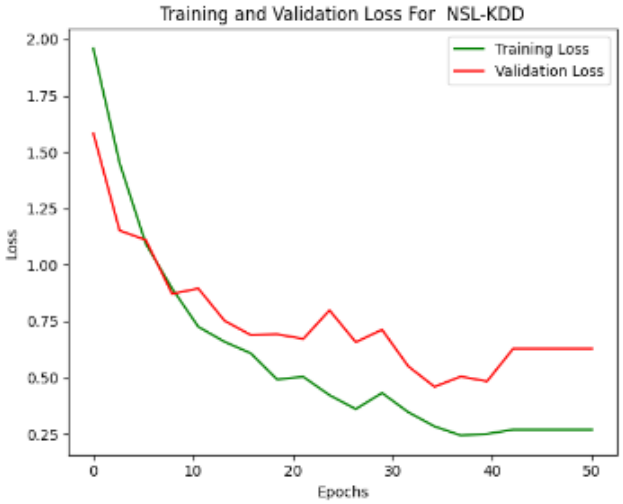

Figure. 4 Loss Curve for Proposed Hawks-GBM NSL-KDD model

Figure. 5 Accuracy Curve for Proposed Hawks-GBM CIDDS-001 Model

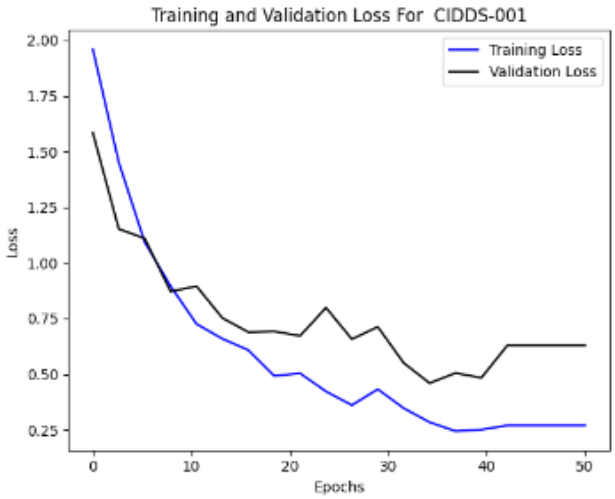

Figure. 6 Loss Curve for Proposed Hawks-GBM CIDDS-001 model

Figure. 7 Accuracy Curve for Proposed Hawks-GBM WSN-DS Model

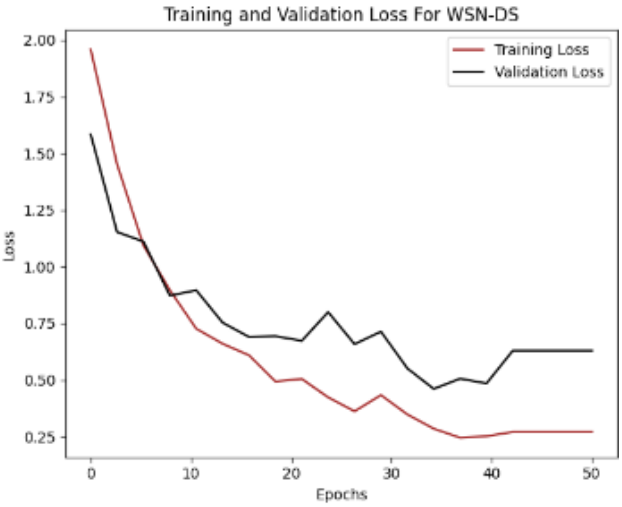

Figure. 8 Loss Curve for Proposed Hawks-GBM WSN-DS model

This most probably could be a case of overfitting, whereby the model is overly specific to the training data, hence it fails to generalise on unseen data. Techniques applied in this include regularization, data augmentation, or simplification of the model in question. Fig. 8 is provided as the losses for the training and the validation sets across 50 epochs. It is interesting to see how both these curves depict an overall fall, which indicates the fact that a model is learning over time; however, in this context, the curve for the validation loss has been flattening out at about the 30th epoch, showing the signal of overfitting. This means that the model is overfitting: it learns the noise and specific patterns in the training data too well, failing to generalize well to unseen data: the validation set. Solutions include techniques such as regularization, data augmentation, or reducing model complexity.

4.1 Performance metrics

Performance evaluation of the HawksGBM model using various performance metrics that enable quantitative analysis of its effectiveness in threat detection tasks in hybrid wireless sensor networks (HWSNs).

Accuracy: It provides a general evaluation of the HawksGBM's model accuracy in detection tasks by calculating the proportion of properly identified cases compared to the total number of incidents and given by Eq. (7),

$$Acc = \frac{Tp\prime+Tn\prime}{Tp\prime+Tn\prime+Fp\prime+Fn\prime} \tag{7}$$

Precision: It determines the extent to which the HawksGBM model can prevent false positives by calculating the percentage of actual positive predictions among all the positive predictions of the model and it is given by Eq. (8),

$$P = \frac{Tp\prime}{Tp\prime+Fp\prime} \tag{8}$$

Recall: It determines the percentage of real positive predictions across all real positive occurrences in the dataset, which quantifies the ability of the model to capture all positive cases which is given by Eq. (9),

$$R = \frac{Tp\prime}{Tp\prime+Fn\prime} \tag{9}$$

Where Tp' represents the number of cases that the model accurately identified as positive, Tn' represents the number of cases that the model accurately identified as negative, Fp' represents the number of cases that the model inaccurately identified as positive and Fn' represents the number of cases that the model inaccurately identified as negative.

F1 score: A harmonic mean of recall and accuracy that provides an accurate assessment of the strategy output and is especially useful in situations when class distributions are unbalanced and represented by Eq. (10),

$$F1\ score = 2 * \frac{\mathrm{P} * R}{\mathrm{P} + \mathrm{R}} \tag{10}$$

Table 2 below displays the performance metrics of a proposed model, HawksGBM. The three datasets used include NSL-KDD, CIDDS-001, and WSN-DS. The metrics are accuracy, precision, recall, and F1-score, which are indicators of classification performance. HawksGBM performed exceptionally high accuracy at 99.6%, precision at 99.7%, recall at 99.5%, and F1-score at 99.6% with the NSL-KDD dataset, meaning excellent detection of both positive and negative cases with minimal misclassifications. The model still has good performance on the CIDDS-001 dataset with an accuracy of 99.1% and, importantly, a recall of 99.8%, indicating that it can classify nearly all true positives correctly.

Table 2. Performance Metrics of the Proposed HawksGBM Model on Various Datasets

	Proposed HawksGBM			
Dataset	Accuracy	Precision	Recall	F1-Score
NSL-KDD	99.6	99.7	99.5	99.6
CIDDS-001	99.1	99.6	99.8	99.3
WSN-DS	98.9	98.8	98.9	98.5

Table 3 Comparison of Performance Metrics for Different Methods on the NSL-KDD Dataset"

Methods	Data Collection	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
ANN [1]	**NSL-KDD**	97.5%	99%	96.7	95.7
RBM [2]	**NSL-KDD**	73%	62%	68%	75 %
MCNN [3]	**NSL-KDD**	69%	84%	69%	
proposed	**NSL-KDD**	99.6	99.7	99.5	99.6

Table 4 Comparison of Performance Metrics for Different Methods on the CIDDS-001 Dataset

Methods	Data Collection	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CNN-LSTM [4]	CIDDS-001	98.88	98.41	99.5	99.91
DCNN [5]	CIDDS-001	87.5	89.7	97.5	78.8
KNN [6]	CIDDS-001	90.8	98.6	98.5	97.5
Proposed	CIDDS-001	99.1	99.6	99.8	99.3

Results on the WSN-DS dataset are slightly lower but still impressive, with accuracy (98.9%), precision (98.8%), recall (98.9%), and F1-score (98.5%), indicating consistent and robust performance across diverse datasets.

Table 3 shows the comparison of different methods on the NSL-KDD dataset in terms of accuracy, precision, recall, and F1-score. The ANN method is good in terms of precision at 99% and has overall solid performance at 97.5% accuracy. On the other hand, RBM is way worse since it provides very low accuracy at 73% and precision at 62%. MCNN gives moderate precision at 84% but lower accuracy at 69%. The proposed model, HawksGBM, exhibited the best performance in all other models, with highest accuracy, precision, recall, and F1-score set at 99.6%. Table 4 will compare different approaches- CNN-LSTM, DCNN, KNN and the developed approach-regarding CIDDS 001 data. Comparing all tables, it can clearly be seen that this new approach of developing Hawks G-BM offers the finest accuracy (99.1%), precision as well as recall results, which are 99.6, and 99.8 respectively. Other competitive methods are of CNNLSTM, this performed well, comparing results from it like 98.88% was accurate which was obtained in result percentage at 99.5%. KNN is developed good precision, result rates of 98.6 percent and has a score of recall, which are at a percentage of 98.5 percent. Nevertheless, the case of having smaller accuracy scores is reported with DCNN, i.e., which is at only 87.5%, while F-score is obtained, which is about 78.8%. 9 Performance Metrics of proposed HawksGBM for different Dataset is decipted in Fig. 9.

Fig. 10 depicts the Fitness Improvement Graph of HHO. The fitness improvement graph for HHO visually signifies the enhancement in fitness or optimization objective achieved over successive iterations or generations. For learning the optimization procedure and for making wise decisions about algorithm design and parameter adjustments it is a useful tool.

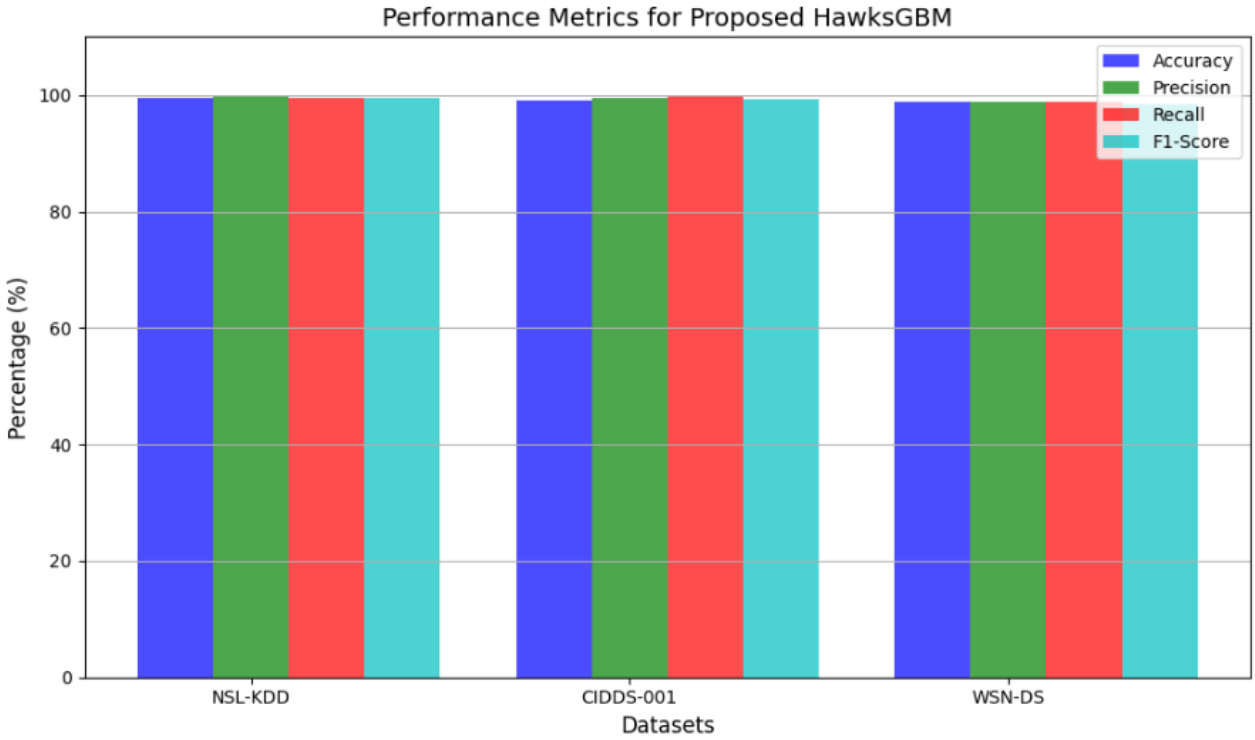

Figure. 9 Performance Metrics of proposed HawksGBM for different Dataset

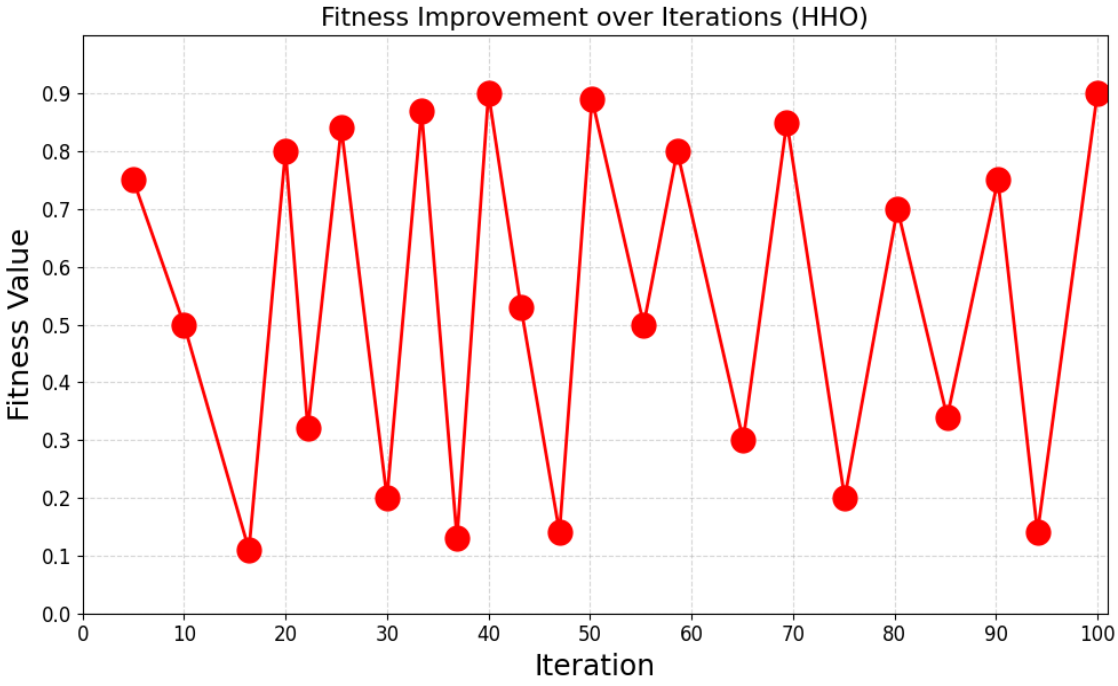

Figure. 10 Fitness Improvement Graph of HHO

4.2 Discussion

The incorporated model combines the predictive powers of GBM for precise danger class with the blended collective intelligence of HHO for worldwide search area exploration and exploitation. Traditional techniques for hazard reputation in HWSNs often depend upon guide rule-primarily based systems or simplistic machine learning models, which may additionally warfare to effectively seize the complexity and dynamics of real-world chance situations [36]. The proposed approach gives a sturdy and efficient solution that overcomes the drawbacks of traditional techniques, marking a great development within the area of chance recognition in HWSNs. The combination of superior system learning algorithms with optimization techniques inspired by nature provides the course to more reliable and effective threat detection systems in wireless sensor networks.

As given in the results section, the HHO-GBM model reaches a mean accuracy of 99.6 % in threat detection within HWSNs. The high accuracy of this model points to something quite important: the effectiveness of threat detection here is significantly higher than in earlier approaches. It is, therefore, very accurate with a precision level of 99.7% and the recall of 99.5% can further testify to the fact that the presented model has the potential to identify the majority of threats while at the same time providing as few false alerts as possible. High precision means that most of the time when the model is signalling that there is a threat, it is pretty much guaranteed that it is the genuine article, which is very important for critical infrastructure defence where alerts that are not real can cause a lot of unnecessary attention. On the same note, the high recall rate shows that the model can easily identify almost all of the actual threats, which is essential for an all-rounded security system.

As one can see from the subsequent comparison with other methods, the newly proposed HHO-GBM model shows better results than SG-IDS and ANN-GA. The combination of HHO with GBM has a positive interaction that improves performance. HHO enhances optimization by increasing the speed of convergence which in turn helps to achieve the optimal value of the hyperparameters search space. However, GBM employs this optimized search to construct a very accurate and very strong predictive model. This combination leads to a model that provides better accuracy than the earlier models and also promises enhanced performance in the process of classification between normal and malicious behaviours in comparison to the conventional techniques.

The practical concern of the HHO-GBM model is relevant to network management and information security specialists. Mainly due to its high accuracy and efficiency, the developed classifier can be viewed as one of the effective tools that help protect HWSNs from new and further cyber threats. Performance

diagrams in terms of the accuracy curve and the fitness improvement graph are depicted to establish the model's stability and learning speed. However, the limitations of the proposed approach may also arise in the problem of validating the proposed method in different network environments and with other sets of data. The following studies might consider augmenting the present optimization approaches or utilizing a combination of the above-mentioned models to improve threat identification and guarantee the model's versatility in several problem environments. The use of these insights will offer a detailed discussion to show their implementation and how it will be used to explain the value of the model as well as the research direction needed.

5. Conclusion and future works

The proposed approach is a primary improvement inside the vicinity of wireless sensor community chance detection and type. It has evolved into an exceedingly solid and powerful device that can detect and mitigate safety dangers in actual time by merging HHO with GBM. In this regard, the proposed Hawks-GBM system provides a comprehensive solution to dealing with dynamic and complex threat detection problem in network environment by integrating the optimization mechanism of Harris Hawks Optimization (HHO) and gradient boosting ability of Gradient Boosting. By exploiting the exploration-exploitation functionality of HHO to maneuver solution spaces with better efficiency and the GB technique for producing high-accuracy classification outcomes, this method reveals a higher, more specific level of detection accuracy across different attack types in the 3 different datasets. By decreasing false positives and negative consequences and growing hazard detection accuracy, the aggregate of HHO with GBM raises the system's overall reliability. Along with identifying and categorizing threats, the method also permits prompt response and remediation measures to address detected risks. To mitigate the effect of security incidents, this will involve separating infected devices, restricting malicious community traffic, and changing safety guidelines. The version's functions increase past early danger identification and reaction to encompass continuous network safety tracking and enhancement. The model determines rising risks and modifies its detection and mitigation techniques in response by keeping a test on community operations and performance metrics.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Mustafa and Ali; methodology, Ali; Software, Mustafa Hamid; validation, Hussein, Ali; formal analysis, Mustafa; investigation, Hussein; resources, Mustafa; data curation, Ali and Mustafa; writing—original draft preparation, Ali; writing—review and editing, Hussein; visualization, Mustafa Hamid; supervision, Ali; project administration, Mustafa, and Ali; funding acquisition, Ali". All authors have read and approved the final manuscript.

References

[1] P. Joshi and A. S. Raghuvanshi, "Hybrid Approaches to Address Various Challenges in Wireless Sensor Network for IoT Applications: Opportunities and Open Problems", *IJCNA*, Vol. 8, No. 3, p. 151, 2021, doi: 10.22247/ijcna/2021/209186.

[2] B. Bhushan and G. Sahoo, "Requirements, Protocols, and Security Challenges in Wireless Sensor Networks: An Industrial Perspective", In: *Proc. of Handbook of Computer Networks and Cyber Security: Principles and Paradigms*, pp. 683-713, 2020, doi: 10.1007/978-3-030-22277-2_27.

[3] A. Jain and T. Singh, "Security Challenges and Solutions of IoT Ecosystem", In: *Proc. of Information and Communication Technology for Sustainable Development*, Vol. 933, M. Tuba, S. Akashe, and A. Joshi, Eds., in Advances in Intelligent Systems and Computing, Vol. 933, pp. 259-270, 2020, doi: 10.1007/978-981-13-7166-0_25.

[4] Z. Nurlan, T. Zhukabayeva, M. Othman, A. Adamova, and N. Zhakiyev, "Wireless Sensor Network as a Mesh: Vision and Challenges", *IEEE Access*, Vol. 10, pp. 46-67, 2022, doi: 10.1109/ACCESS.2021.3137341.

[5] S. Lata, S. Mehfuz, and S. Urooj, "Secure and Reliable WSN for Internet of Things: Challenges and Enabling Technologies", *IEEE Access*, Vol. 9, pp. 161103-161128, 2021, doi: 10.1109/ACCESS.2021.3131367.

[6] U. S. R. D, S. P, K. Arunkumar, S. R, and M. P, "A HSEERP—Hierarchical secured energy efficient routing protocol for wireless sensor networks", *Peer-to-Peer Netw. Appl.*, Vol. 17, No. 1, pp. 163-175, 2024, doi: 10.1007/s12083-023-01575-w.

[7] U. Inayat, F. Ali, H. M. A. Khan, S. M. Ali, K. Ilyas, and H. Habib, "Wireless Sensor Networks: Security, Threats, and Solutions", In: *Proc. of 2021 International Conference on Innovative Computing (ICIC)*, pp. 1-6, 2021, doi: 10.1109/ICIC53490.2021.9693021.

[8] M. Zeeshan *et al.*, "Protocol-Based Deep Intrusion Detection for DoS and DDoS Attacks Using UNSW-NB15 and Bot-IoT Data-Sets", *IEEE Access*, Vol. 10, pp. 2269-2283, 2022, doi: 10.1109/ACCESS.2021.3137201.

[9] G. Yang, M. A. Jan, A. U. Rehman, M. Babar, M. M. Aimal, and S. Verma, "Interoperability and Data Storage in Internet of Multimedia Things: Investigating Current Trends, Research Challenges and Future Directions", *IEEE Access*, Vol. 8, pp. 124382-124401, 2020, doi: 10.1109/ACCESS.2020.3006036.

[10] T. Aljrees, A. Kumar, K. U. Singh, and T. Singh, "Enhancing IoT Security through a Green and Sustainable Federated Learning Platform: Leveraging Efficient Encryption and the Quondam Signature Algorithm", *Sensors*, Vol. 23, No. 19, 2023, doi: 10.3390/s23198090.

[11] A. K. Gautam and R. Kumar, "A comprehensive study on key management, authentication and trust management techniques in wireless sensor networks", *SN Appl. Sci.*, Vol. 3, No. 1, p. 50, 2021, doi: 10.1007/s42452-020-04089-9.

[12] D. P. Somani, "Network Security in Wireless Sensor Networks: Threats and Countermeasures", *Iconic Research And Engineering Journals*, Vol. 7, No. 2, 2023.

[13] K. Sood, M. R. Nosouhi, N. Kumar, A. Gaddam, B. Feng, and S. Yu, "Accurate Detection of IoT Sensor Behaviors in Legitimate, Faulty and Compromised Scenarios", *IEEE Transactions on Dependable and Secure Computing*, Vol. 20, No. 1, pp. 288-300, 2023, doi: 10.1109/TDSC.2021.3131991.

[14] G. G. Gebremariam, J. Panda, and S. Indu, "Design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks", *Connection Science*, 2023, Accessed: Mar. 01, 2024. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/09540091.2023.2246703

[15] R. A. Muhajjar, N. A. Flayh, and M. Al-Zubaidie, "A Perfect Security Key Management Method for Hierarchical Wireless Sensor Networks in Medical Environments", *Electronics*, Vol. 12, No. 4, 2023.

[16] D. E. Boubiche, S. Athmani, S. Boubiche, and H. Toral-Cruz, "Cybersecurity Issues in Wireless Sensor Networks: Current Challenges and Solutions", *Wireless Pers Commun*, Vol. 117, No. 1, pp. 177-213, 2021, doi: 10.1007/s11277-020-07213-5.

[17] M. Dener, S. Al, and A. Orman, "STLGBM-DDS: An Efficient Data Balanced DoS Detection System for Wireless Sensor Networks on Big Data Environment", *IEEE Access*, Vol. 10, pp. 92931-92945, 2022, doi: 10.1109/ACCESS.2022.3202807.

[18] M. Ragab, S. M. Alshammari, L. A. Maghrabi, D. Alsalman, T. Althaqafi, and A. A.-M. AL-Ghamdi, "Robust DDoS Attack Detection Using Piecewise Harris Hawks Optimizer with Deep Learning for a Secure Internet of Things Environment", *Mathematics*, Vol. 11, No. 21, 2023, doi: 10.3390/math11214448.

[19] Simon, "Hybrid intrusion detection system for wireless IoT networks using deep learning algorithm", *Computers and Electrical Engineering*, Vol. 102, p. 108190, 2022, doi: 10.1016/j.compeleceng.2022.108190.

[20] Y. Li, J. Zhang, Y. Yan, Y. Lei and C. Yin, "Enhancing Network Intrusion Detection Through the Application of the Dung Beetle Optimized Fusion Model", *IEEE Access*, Vol. 12, pp. 9483-9496, 2024, doi: 10.1109/ACCESS.2024.3353488.

[21] L. Almuqren, K. Mahmood, S. S. Aljameel, A. S. Salama, G. P. Mohammed, and A. A. Alneil, "Blockchain-Assisted Secure Smart Home Network Using Gradient-Based Optimizer With Hybrid Deep Learning Model", *IEEE Access*, Vol. 11, pp. 86999-87008, 2023, doi: 10.1109/ACCESS.2023.3303087.

[22] M. H. Ali *et al.*, "Threat Analysis and Distributed Denial of Service (DDoS) Attack Recognition in the Internet of Things (IoT)", *Electronics*, Vol. 11, No. 3, 2022, doi: 10.3390/electronics11030494.

[23] E. Elhariri, N. El-Bendary, and S. A. Taie, "Using Hybrid Filter-Wrapper Feature Selection With Multi-Objective Improved-Salp Optimization for Crack Severity Recognition", *IEEE Access*, Vol. 8, pp. 84290-84315, 2020, doi: 10.1109/ACCESS.2020.2991968.

[24] K. Cengiz, S. Lipsa, R. K. Dash, N. Ivković, and M. Konecki, "A Novel Intrusion Detection System Based on Artificial Neural Network and Genetic Algorithm With a New Dimensionality Reduction Technique for UAV Communication", *IEEE Access*, Vol. 12, pp. 4925-4937, 2024, doi: 10.1109/ACCESS.2024.3349469.

[25] Y. Zhang, R. Liu, X. Wang, H. Chen, and C. Li, "Boosted binary Harris hawks optimizer and feature selection", *Engineering with Computers*, Vol. 37, No. 4, pp. 3741-3770, 2021, doi: 10.1007/s00366-020-01028-5.

[26] J. Jose and J. E. Judith, "Unveiling the IoT's dark corners: anomaly detection enhanced by ensemble modelling", *Journal for Control, Measurement, Electronics, Computing and Communications*, Vol. 65, No. 2024.

[27] "NSL-KDD." Accessed: Mar. 04, 2024. [Online]. Available: https://www.kaggle.com/datasets/hassan06/nslkdd

[28] G. G. Gebremariam, J. Panda, and S. Indu, "Design of advanced intrusion detection systems based on hybrid machine learning techniques in hierarchically wireless sensor networks", *Connection Science*, Vol. 35, No. 1, p. 2246703, 2023.

[29] "CIDDS-001", Accessed: Nov. 04, 2024. [Online]. Available: https://www.kaggle.com/datasets/dhoogla/cidds001

[30] "WSN-DS", Accessed: Nov. 04, 2024. [Online]. Available: https://www.kaggle.com/datasets/bassamkasasbeh1/wsnds

[31] J. M. Yousif, "A Comparative Analysis between Various Machine Learning Models and Generalized Linear Models", *PhD Thesis, Stockholm University Stockholm*, Sweden, 2023.

[32] H. M. Saleh, H. Marouane, and A. Fakhfakh, "Stochastic Gradient Descent Intrusions Detection for Wireless Sensor Network Attack Detection System Using Machine Learning", *IEEE Access*, Vol. 12, pp. 3825-3836, 2024, doi: 10.1109/ACCESS.2023.3349248.

[33] B. Mahbooba, R. Sahal, W. Alosaimi, and M. Serrano, "Trust in Intrusion Detection Systems: An Investigation of Performance Analysis for Machine Learning and Deep Learning Models", *Complexity*, Vol. 2021, No. 1, p. 5538896, 2021, doi: 10.1155/2021/5538896.

[34] A. Rashid, M. J. Siddique, and S. M. Ahmed, "Machine and deep learning based comparative analysis using hybrid approaches for intrusion detection system", In: *Proc. of 2020 3rd International Conference on Advancements in Computational Sciences (ICACS)*, pp. 1-9, 2020.

[35] A. Agarwal, P. Sharma, M. Alshehri, A. A. Mohamed, and O. Alfarraj, "Classification model for accuracy and intrusion detection using machine learning approach", *PeerJ Computer Science*, Vol. 7, p. e437, 2021.

[36] A. Shakerimov, T. Alizadeh, and H. A. Varol, "Efficient Sim-to-Real Transfer in Reinforcement Learning Through Domain Randomization and Domain Adaptation", *IEEE Access*, Vol. 11, pp. 136809-136824, 2023, doi: 10.1109/ACCESS.2023.3339568.