Network Security by Identifying Malicious Activity by Optimizing Features and Machine Learning Model

Apoorva Patil[1], Dr. Ankit Pandit[2], Dr. Sanjeev Kumar Gupta[3], Dr. Laxmi Singh[4]
[1]*Research scholar, RNTU Bhopal (M.P)*
[2]*Associate Professor, RNTU Bhopal (M.P)*
[3]*Dean, Faculty of Engineering, RNTU Bhopal (M.P)*
[4]*HoD, ECE, RNTU Bhopal (M.P)*
doi.org/10.64643/IJIRTV12I6-187373-459

***Abstract*—Wireless Sensor Networks (WSNs) are widely used in military and civilian applications for remote monitoring through interconnected sensors. Their open deployment and wireless communication make them highly prone to security vulnerabilities. As a result, WSNs are especially susceptible to various routing attacks. This paper presents a Genetic Algorithm and Feature-Optimized Nonnegative Matrix Factorization System (GAFNS) for efficient network intrusion detection. The proposed model performs dataset preprocessing, feature selection using a Genetic Algorithm, and feature transformation through Nonnegative Matrix Factorization (NMF). An ensemble of Long Short-Term Memory is then used for classification.**

***Index Terms*—Genetic Algorithm, Intrusion Detection, Feature Optimization, Machine Learning.**

I. INTRODUCTION

The recent wave of industrial innovation has been driven by the growing demand for smarter and more efficient systems. In Cyber-Physical Systems (CPS), physical devices are interconnected with the cyber world through wireless communication primarily via the Internet and can be monitored or controlled remotely [1]. Most CPS applications function autonomously, utilizing sensors to observe and interpret environmental conditions. These sensor technologies have become integral across multiple domains, such as healthcare, smart agriculture, and intelligent transportation systems [2]. The collected sensor data is transmitted through wireless communication channels, as illustrated in Figure 1, which depicts the basic mechanism of wireless sensor communication.

However, this interconnectivity also introduces security vulnerabilities, making such systems potential targets for malicious attacks. Cyber adversaries can infiltrate sensor networks, threaten their integrity and cause significant harm to CPS operations. Since human intervention in mitigating these attacks is often minimal, there is a growing need for automated security mechanisms to ensure the protection of sensor networks.

The detection of malicious nodes has long been a key research focus in wireless sensor networks (WSNs). Various detection techniques have been proposed by researchers to identify and isolate compromised nodes. While many existing approaches concentrate on identifying malicious nodes within a single communication path, the increasing use of multipath data transmission in WSNs presents new challenges for reliable detection and secure data delivery [3].

An Intrusion Detection System (IDS) serves as a critical network security mechanism that monitors and analyses network traffic to identify suspicious behaviour and protect system resources [4]. It plays a vital role in ensuring overall network security [5]. In WSNs, intrusions typically fall into two main categories: anomalies and misuse. Anomaly detection applies mathematical models to compare observed network features with reference patterns, identifying deviations that indicate potential threats [6]. Misuse detection, in contrast, recognizes attacks based on known malicious signatures or activity patterns. Common cyber threats in WSNs include Denial of Service (DoS) and web-based attacks, which may be active involving actions like packet manipulation or passive, such as unauthorized eavesdropping and data interception [7].

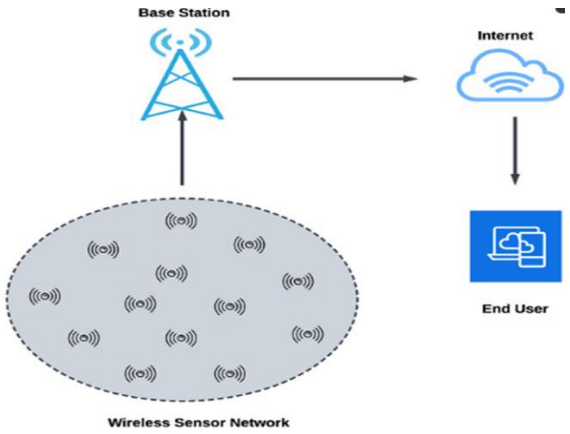

Fig. 1 Block diagram of proposed model.

This paper has identified few issues:

- Most of work only identify attack and non-attack class of network session.
- Further calculation of most of work is complex and time taken.
- Improves the true alarm rate with intrusion class efficiency.

The rest of this paper is organized as follows: Section 2 presents a review of related studies. Section 3 discusses the preliminaries and outlines the proposed system model. The performance evaluation and experimental results are detailed in Section 4. Finally, Section 5 concludes the paper and highlights possible directions for future research.

II. RELATED WORK

Chuan Long Yin [8] (2017) proposed an intrusion detection framework based on deep learning principles, introducing a Recurrent Neural Network-based Intrusion Detection System (RNN-IDS). The study evaluated the model's performance for both binary and multiclass classification tasks and analyzed how the number of neurons and the learning rate influence its accuracy. The results were compared with traditional machine learning models such as J48, Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM), demonstrating improved detection efficiency on standard benchmark datasets.

Kabir et al. [9] (2018) developed the Optimum Allocation Least Square Support Vector Machine (OALSSVM) model, which selectively chooses optimal sessions from training and testing datasets to train the SVM. This selective data allocation enhances the precision of intrusion detection by ensuring that only the most informative samples are used for training, thereby increasing model accuracy.

Khraisat and Gondal [10] (2019) conducted a comprehensive review of Intrusion Detection Systems (IDS), covering various techniques, benchmark datasets, and the major challenges in the field. Their work provided a holistic understanding of existing methodologies and highlighted future research directions for improving IDS effectiveness.

Kaiyuan et al. [11] (2020) proposed a hybrid network intrusion detection algorithm that integrates hybrid sampling and a deep hierarchical neural network. The approach first applies One-Side Selection (OSS) to remove noise from majority classes and Synthetic Minority Oversampling Technique (SMOTE) to balance the dataset. The model then employs Convolutional Neural Networks (CNN) to extract spatial features and Bidirectional LSTM (BiLSTM) to capture temporal dependencies, significantly improving learning efficiency and detection accuracy.

Yi Lu, Menghan Liu, and Jie Zhou [12] (2021) introduced a hybrid intrusion detection model that combines adaptive clonal genetic algorithms with backpropagation neural networks. This method effectively enhances intrusion detection by adaptively optimizing feature selection and neural network learning. Their study demonstrated that the proposed system performs well in real-world cybersecurity environments, offering a robust and adaptive detection mechanism.

T. Kim et al. [13] (2022) presented a GAN-assisted LSTM-DNN-based intrusion detection system that focuses on identifying misclassified packets. Misclassified samples from the original training dataset are used to train a Generative Adversarial Network (GAN), which helps determine whether an incoming packet can be accurately classified. If uncertain, detection is deferred until further data is available. This method allows real-time intrusion detection without session delays, achieving high precision and adaptability.

C. Park et al. [14] (2023) proposed an AI-driven Network Intrusion Detection System (NIDS) that addresses the issue of data imbalance using Generative Adversarial Networks (GANs). Their approach employs a reconstruction error and Wasserstein distance-based GAN architecture to generate synthetic

minority attack samples. By integrating the generative model with anomaly detection, the proposed system achieved superior classification performance compared to existing methods, demonstrating improved balance and reliability in intrusion detection.

III. PROPOSED MODEL

This section provides detailed information on methodology of working and process flow of the proposed system which includes the basic operations. Figure 1 illustrates its core stages, including dataset preprocessing, dimensionality reduction, feature transformation, and training. The preprocessing phase ensures data consistency, while dimensionality reduction removes redundant features to improve efficiency [12, 14]. Feature transformation refines data attributes for better classification, and the training stage employs machine learning techniques for accurate intrusion detection. Each of these components is discussed in detail under separate headings in this section, explaining their significance in optimizing GAFNS for enhanced network security and performance.

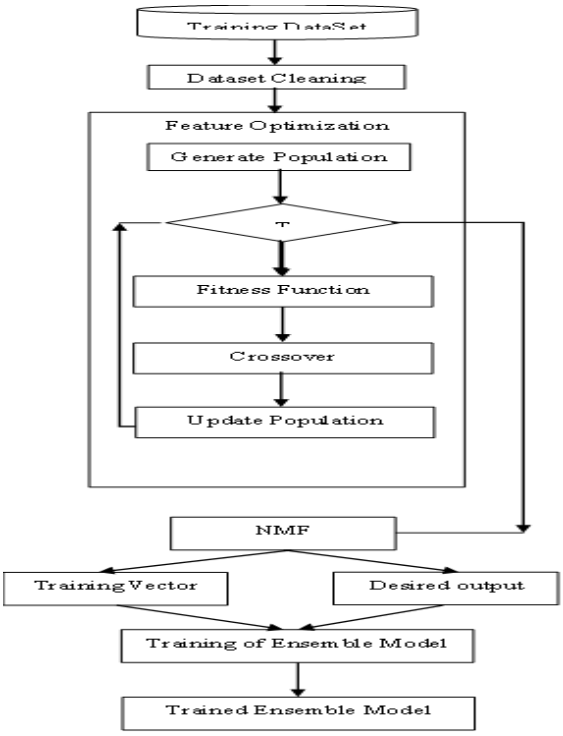

Fig. 1 Block diagram of GAFNS network intrusion detection.

Dataset Processing

The input dataset used for malicious session detection contains numerous features, each with varying levels of significance. In this initial stage, the dataset undergoes a cleaning process to remove irrelevant or redundant information, ensuring that only meaningful data is used for further analysis. For instance, the dataset employed in this research contains several fields, but the first few feature values were deemed unnecessary for detecting malicious sessions [15]. Examples of such fields include session IDs, connection types, and transferring protocols, which do not contribute significantly to identifying malicious activity. By eliminating these features, the dataset becomes more focused and efficient, allowing the model to concentrate on the most critical features.

PNSD←DataProcessing(NSD) ----------Eq. 1

In Equation 1, the Network Session Dataset (NSD) represents the raw input data, while the Processed NSD (PNSD) is the refined version after initial preprocessing. The processed dataset is structured in a matrix format, where each row corresponds to a network session, and each column represents a specific feature associated with that session.

Feature Optimization Following the data preprocessing phase, the PNSD matrix undergoes feature optimization using a Genetic Algorithm (GA). This stage aims to minimize the number of input features in the training vector while improving classification accuracy. By retaining only the most significant features, GA enhances both the computational efficiency and detection capability of the intrusion detection system.

Genetic Algorithm

The Genetic Algorithm is an evolutionary computation method inspired by the natural selection process. In this approach, each chromosome represents a candidate solution (feature subset), which evolves through operations such as selection, crossover, and mutation [16]. The fittest individuals are preserved and recombined across generations to yield progressively better solutions.

In this context, each chromosome corresponds to a possible subset of features, and the GA iteratively refines these subsets to identify the most relevant features for intrusion detection.

Generating Student Population, the initial population consists of several chromosomes (feature sets), where

each chromosome is a binary vector of length n, representing the number of features in the PNSD matrix:

1 → the feature is selected for training

0 → the feature is excluded

To ensure diversity, s chromosomes are randomly generated, forming a population matrix (P) of size s × n. Random binary initialization helps explore a wide range of feature combinations during optimization.

SP←GeneratePopulation(s, n, f)----------Eq. 2

Fitness Function Each chromosome's fitness (SF) is determined based on how effectively its feature subset improves the detection accuracy. The chromosome is evaluated using a learning model trained on the PNSD matrix, and the resulting accuracy represents its fitness value.

SF←learningAccuracy(SP,PNSD, DOC)

The higher the accuracy, the fitter the chromosome. This step ensures that only the most relevant features contributing to better intrusion detection performance are retained.

Selection Phase In the selection phase, chromosomes with higher fitness values are given greater chances to reproduce. Common strategies such as roulette wheel selection, tournament selection, or rank-based selection are used to choose parent chromosomes [17]. This step ensures that strong candidates have a higher probability of passing their genes (features) to the next generation.

Crossover Phase Selected parent chromosomes are paired to perform crossover the exchange of feature positions to create new offspring. This process promotes diversity and helps combine beneficial traits from different parents. For example, a single-point or uniform crossover may be applied to merge two feature sets into new combinations, possibly leading to improved detection performance.

SP←Crossover(SP, SF, PNSD, DOC)

Iteration and Convergence After crossover and mutation, the new generation is evaluated using the fitness function. The population is updated by replacing weaker chromosomes with stronger offspring. This evolutionary cycle continues for a predefined number of iterations or until the population converges, i.e., no significant improvement in fitness is observed.

Feature Selection Once the GA process converges, the best-performing chromosome (Pbest) is selected as the optimal feature subset. Features corresponding to "1" in the chromosome vector are retained for training, while those marked as "0" are discarded. The desired output matrix is then updated accordingly to match the reduced feature set. This feature optimization approach ensures that the final model operates on the most informative attributes, thereby enhancing both the efficiency and accuracy of intrusion detection within cloud or network environments.

FSD←Filter(PSD,SPb)

Nonnegative Matrix Factorization (NMF) Feature optimization is a vital component of machine learning and data analytics, as it enables dimensionality reduction thereby improving both the performance and interpretability of models [18]. Nonnegative Matrix Factorization (NMF) is one such widely used feature extraction technique that decomposes a nonnegative data matrix into two smaller nonnegative matrices, effectively capturing the underlying structure of the data in a lower-dimensional space.

NMF is particularly advantageous in scenarios where data values are inherently nonnegative, such as in image processing, text mining, medical imaging, and bioinformatics. Unlike other matrix decomposition techniques, NMF enforces nonnegativity constraints, ensuring that each data point is represented as an additive combination of basic, interpretable components. This property makes the results of NMF easy to understand and directly associated with physical or conceptual parts of the data [19].

By reducing redundancy and focusing on the most informative patterns, NMF enhances the quality of feature representation. Consequently, it plays a key role in feature selection and dimensionality reduction, helping improve the efficiency, stability, and overall accuracy of subsequent classification or prediction models.

Ensemble Model An ensemble learning model integrates the predictions of multiple learning algorithms to produce a stronger and more robust predictive system. In this framework, two different models are trained independently using the same set of optimized input features, and their outputs are combined to generate the final classification result.

In this study, an ensemble of Long Short-Term Memory (LSTM) and a Feedforward Neural Network (FNN) is employed for intrusion detection. Both models are trained using the matrices H (NMF output coefficient matrix) and DOC (Desired Output Class) as inputs. The ensemble approach leverages the

temporal learning capability of LSTM [20] and the non-linear mapping power of the neural network, thereby improving the system's adaptability and detection accuracy.

IV. EXPERIMENT AND RESULTS

The proposed GAFNS model and the comparative model were implemented using MATLAB software. The experiments were conducted on a system equipped with a 6th generation Intel Core i3 processor and 4 GB of RAM. The dataset used for evaluation was sourced from [21]. To assess the performance of GAFNS, its results were compared with those of the cloud malicious session detection model proposed in [22].

Results

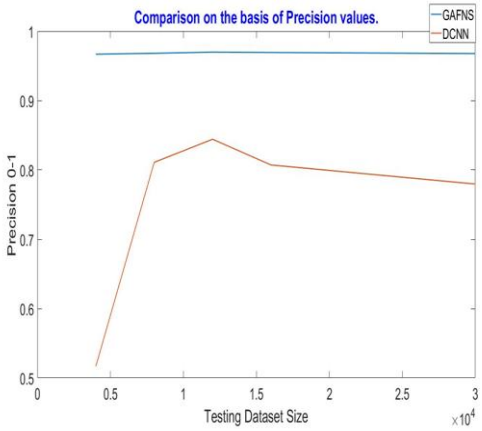

Fig. 2 Precision values based WSN security models comparison.

Table 2 Network security alarm models comparison on the basis of precision values.

Testing Dataset	GAFNS	DCNN
4000	0.9668	0.5168
8000	0.9681	0.8109
12000	0.9698	0.8442
16000	0.9693	0.8071
30000	0.9677	0.7796

Table 3 Network security alarm models comparison on the basis of recall values.

Testing Dataset	GAFNS	Previous Model
4000	0.9623	0.8598
8000	0.9672	0.8789
12000	0.9674	0.8034
16000	0.9671	0.7476
30000	0.9639	0.8011

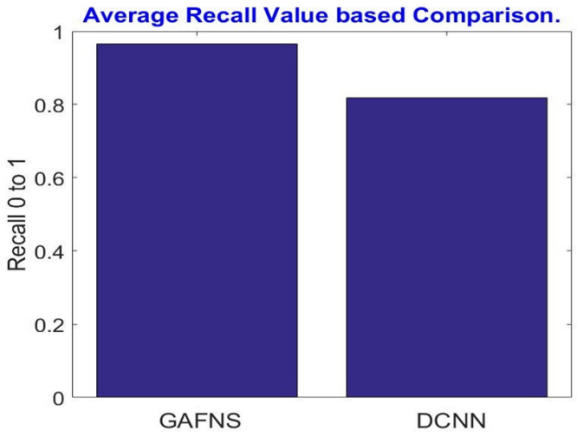

Fig. 3 Average recall values based WSN security models comparison.

Table 4 Network security alarm models comparison on the basis of f-measure values.

Testing Dataset	GAFNS	Previous Model
4000	0.9646	0.6456
8000	0.9676	0.8435
12000	0.9686	0.6837
16000	0.9682	0.7476
30000	0.9658	0.7593

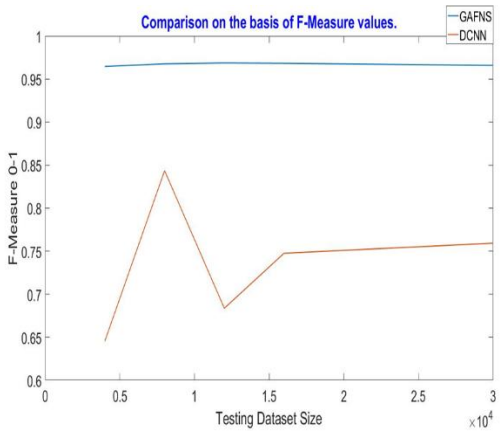

Fig. 4 F-measure values based WSN security model comparison.

The proposed GAFNS model shows significant improvement across all performance metrics compared to the DCNN model [22]. Fig. 2 shows that Precision increased by 22.37%, Fig. 3 shows that recall by 15.26%, and similarly fig. 4 shows that F-measure by 23.89% across different testing datasets. These gains are attributed to the effective use of the modified teacher learning group phase for feature selection and Nonnegative Matrix Factorization

(NMF) for feature transformation, which together enhanced the ensemble model's ability to accurately detect network attacks.

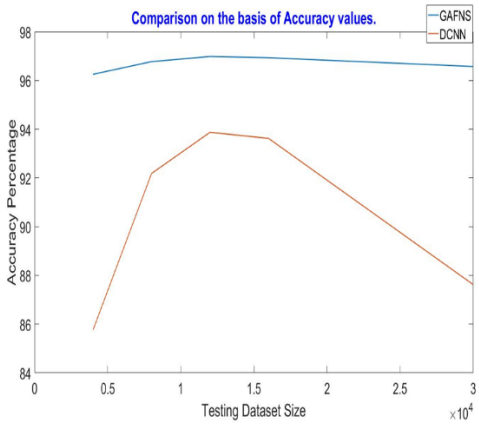

Fig. 2 Accuracy values based WSN security model comparison.

Table 5 Network security alarm models comparison on the basis of accuracy values.

Testing Dataset	GAFNS	Previous Model
4000	96.25	85.76
8000	96.77	92.18
12000	96.98	9387
16000	96.93	93.62
30000	96.57	87.62

Table 5 presents the session-wise attack detection accuracy values. The results reveal that the proposed GAFNS model consistently enhances detection accuracy across all experimental datasets. Fig. 5 shows that the integration of a genetic-based feature selection approach with Nonnegative Matrix Factorization (NMF) has effectively strengthened the ensemble learning process. The combined use of the LSTM and Neural Network models in the proposed framework led to a 6.29% improvement in detection accuracy compared to the DCNN model reported in [22].

Table 6 Network security alarm models comparison on the basis of DOS attacks detection accuracy values

Testing Dataset	GAFNS	Previous Model
4000	96.684	95.0517
8000	96.1538	81.7073
12000	96.2705	97.6744
16000	96.22	96.6566
30000	97.77	97.1811

Table 7 Network security alarm models comparison on the basis of Probe attacks detection accuracy values

Testing Dataset	GAFNS	Previous Model
4000	95.8895	84.85
8000	85.2691	84.7007
12000	84.9265	86.9611
16000	85.376	75.33
30000	95.8519	82.3045

Table 8 Network security alarm models comparison on the basis of U2R attacks detection accuracy values

Testing Dataset	GAFNS	Previous Model
4000	84.6361	46.0118
8000	70.7692	50
12000	70.8333	58.1134
16000	70.8661	49.963
30000	84.6939	70.3189

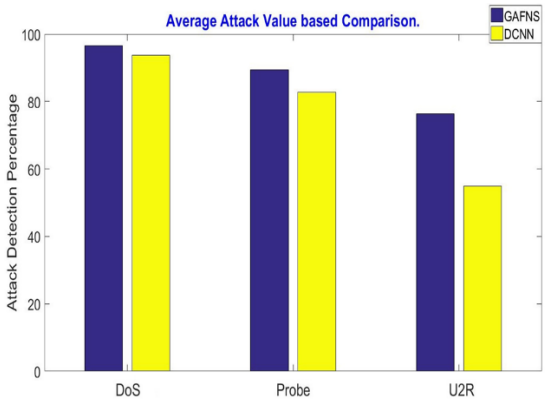

Fig. 6 Attack class detection accuracy values based WSN security models comparison

Tables 6, 7, and 8 present the detection accuracy values for various attack classes across all experimental setups. The results indicate that the proposed model achieves the highest accuracy in detecting DoS (Denial of Service) attacks, while the detection performance for Probe and U2R (User to Root) attacks is comparatively lower. Fig. 6 shows that, the proposed approach demonstrates an improvement in the average attack detection accuracy compared to the DCNN model.

V. CONCLUSION

The proposed GAFNS model effectively enhances network intrusion detection by integrating Genetic

Algorithm-based feature selection, NMF-based feature transformation, and an ensemble of LSTM and neural networks. Experimental results confirm its superior performance over the DCNN model, achieving higher accuracy and reliability across different datasets. The approach provides a scalable and intelligent framework suitable for real-time intrusion detection in complex network environments. Experimental results show that GAFNS achieves higher precision, recall, and overall accuracy compared to the existing DCNN model, with an improvement of 6.29% in detection accuracy. The model effectively detects multiple attack types, including DoS, Probe, and U2R, demonstrating its robustness and efficiency. In future scholar can implement same work in underwater networks.

REFERENCES

[1] Kayan, H.; Nunes, M.; Rana, O.; Burnap, P.; Perera, C. Cybersecurity of industrial cyber-physical systems: A review. ACM Comput. Surv. 2022, 54, 1–35.

[2] Javaid, M.; Haleem, A.; Rab, S.; Singh, R.P.; Suman, R. Sensors for daily life: A review. Sens. Int. 2021, 2, 100121.

[3] Liu A., Zheng Z., Zhang C., Chen Z., and Shen X., Secure and energy-efficient disjoint multipath routing for WSNs, IEEE Transactions on Vehicular Technology. (2012) 61, no. 7, 3255–3265.

[4] Zhang, W., Han, D., Li, K. C., & Massetto, F. I. (2020). Wireless sensor network intrusion detection system based on MK-ELM. Soft Computing, 24(16), 12361–12374.

[5] Wang, Y., Ma, J., Sharma, A., Singh, P. K., Gaba, G. S., Masud, M., & Baz, M. (2021). Exhaustive research on the application of intrusion detection technology in computer network security in sensor networks. Journal of Sensors, 2021.

[6] Godala, S., & Vaddella, R. P. V. (2020). A study on intrusion detection system in wireless sensor networks. International Journal of Communication Networks and Information Security, 12(1), 127–141.

[7] Elsaid, S. A., & Albatati, N. S. (2020). An optimized collaborative intrusion detection system for wireless sensor networks. Soft Computing, 24(16), 12553–12567.

[8] ChuanlongYin ,Yuefei Zhu, JinlongFei, And Xinzheng He. “A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks” current version November 7, 2017.

[9] E. Kabir, J. Hu, H. Wang, and G. Zhuo, ‘‘A novel statistical technique for intrusion detection systems,’’ Future Genre. Computer. Syst., vol. 79, pp. 303–318, Feb. 2018.

[10] Khraisat, A., Gondal, I., Vamplew, P. et al. Survey of intrusion detection systems: techniques, datasets and challenges. Cyber security 2, 20 (2019).

[11] KaiyuanJiang ,Wenya Wang , Aili Wang , And Haibin Wu. "Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network". IEEE Access February 24, 2020.

[12] Yi Lu, Menghan Liu, Jie Zhou, Zhigang Li, "Intrusion Detection Method Based on Adaptive Clonal Genetic Algorithm and Back propagation Neural Network", Security and Communication Networks, vol. 2021.

[13] T. Kim and W. Pak, "Early Detection of Network Intrusions Using a GAN-Based One-Class Classifier," in IEEE Access, vol. 10, pp. 119357-119367, 2022

[14] C. Park, J. Lee, Y. Kim, J. -G. Park, H. Kim and D. Hong, "An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks," in IEEE Internet of Things Journal, vol.10,no.3,pp.2330-2345,1Feb.1,2023.

[15] N. Moustafa and J. Slay, ``UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS), Nov. 2015, pp. 16.

[16] W. Li, Y. Fan and Q. Xu, "Teaching-Learning-Based Optimization Enhanced with Multi objective Sorting Based and Cooperative Learning," in IEEE Access, vol. 8, pp. 65923-65937, 2020

[17] Wu, X.; Li, S.; Wu, F.; Jiang, X. Teaching–Learning Optimization Algorithm Based on the Cadre–Mass Relationship with Tutor Mechanism for Solving Complex Optimization Problems. Biomimetics 2023, 8, 462.

[18] D. P. Truong, E. Skau, D. Desantis and B. Alexandrov, "Boolean Matrix Factorization via Nonnegative Auxiliary Optimization," in IEEE Access, vol. 9, pp. 117169-117177, 2021.

[19] W. Zhu and Y. Yan, "Joint Linear Regression and Nonnegative Matrix Factorization Based on Self-Organized Graph for Image Clustering and Classification," in IEEE Access, vol. 6, pp. 38820-38834, 2018.

[20] Kuo, C.-T.; Lin, J.-J.; Jen, K.-K.; Hsu, W.-L.; Wang, F.-C.; Tsao, T.-C.; Yen, J.-Y. Human Posture Transition-Time Detection Based upon Inertial Measurement Unit and Long Short-Term Memory Neural Networks. Biomimetics 2023, 8, 471.

[21] https://github.com/defcom17/NSL_KDD/blob/master/Original%20NSL%20KDD%20Zip.zip

[22] C. Yao, Y. Yang, K. Yin, and J. Yang, ‘‘Traffic anomaly detection in wireless sensor networks based on principal component analysis and deep convolution neural network,’’ IEEE Access, vol. 10, pp. 103136–103149, 2022.