Dijlah Journal of Engineering Science (DJES) Vol. 3, No. 3, August, 2026, pp. 337-345
ISSN: Printed: **3078-9656, Online: 3078-9664**, paper ID: 209

Cyber Attack Detection Based on Anomaly Intrusion Detection System for IoT Network

Haifa Hazim al-Tai
Department of Arabic, College of Education for Humanities, University of Mosul, Mosul, Iraq

Article Info

Article history:

Received May,15,2026
Revised Jul.,05,2026
Accepted Aug.,22,2026

Keywords:

First keyword
Second keyword
Third keyword
Fourth keyword
Fifth keyword

ABSTRACT

The rapid expansion of Internet of Things (IoT) technologies has created major opportunities for smart communication systems and digital services. At the same time, this development has introduced serious cybersecurity challenges because IoT devices are often characterized by limited memory capacity, low computational power, and insufficient built-in security mechanisms. Consequently, IoT infrastructures have become vulnerable to a wide range of cyber threats, including Distributed Denial of Service (DDoS) attacks, spoofing attacks, malware injection, and unauthorized access attempts.

To strengthen the security of modern IoT environments, this study proposes an anomaly-based Intrusion Detection System (IDS) utilizing the Random Forest machine learning algorithm. The proposed framework was implemented and evaluated using the CICIDS2017 benchmark dataset, which contains realistic network traffic records and diverse attack categories. In addition, the performance of the proposed model was compared with several machine learning and deep learning techniques, including Convolutional Neural Networks (CNN), Logistic Regression (LR), Support Vector Machine (SVM), and Gradient Boosting (GB).

A preprocessing stage was conducted before model training to improve data quality and learning efficiency. The preprocessing operations included removing missing values, converting categorical variables into numerical form, normalizing dataset features, and balancing minority attack classes using the Synthetic Minority Over-sampling Technique (SMOTE).

Experimental findings demonstrated that the Random Forest model achieved superior performance compared with the other investigated approaches. The model produced high values of accuracy, precision, recall, and F1-score while maintaining low false positive rates. The results also confirmed that integrating SMOTE with ensemble learning techniques improves intrusion detection capability and enhances the identification of rare and sophisticated cyberattacks in IoT communication environments.

Corresponding Author:

Haifa Hazim al-Tai
Department of Arabic, College of Education for Humanities, University of Mosul, Mosul, Iraq
Email: Haifahzim123@uomosul.edu.iq

1. Introduction

The widespread integration of smart digital devices into daily and industrial activities has significantly changed modern communication infrastructures. Internet of Things technologies are now embedded in transportation systems, healthcare applications, industrial monitoring platforms, and smart city services. Despite the operational advantages offered by IoT systems, their rapid expansion has also increased exposure to cybersecurity threats. [1].

Unlike conventional computing devices, many IoT nodes operate with restricted processing capabilities, small storage capacity, and limited security support. These limitations make connected devices attractive targets for cybercriminals seeking to exploit vulnerabilities within communication networks. Cyberattacks directed toward IoT

environments may interrupt services, compromise confidential information, and reduce the reliability of digital infrastructures.

Traditional intrusion detection approaches mainly depend on predefined signatures and static attack patterns. Although such methods are capable of identifying previously known threats, they often experience difficulties when dealing with newly emerging attacks or rapidly changing malicious behaviors. Because cyber threats continue to evolve, intelligent security mechanisms capable of adaptive learning have become increasingly necessary. [**2**].

Machine learning and artificial intelligence techniques provide an effective alternative for modern cybersecurity applications. These techniques can analyze traffic behavior, learn hidden relationships among network attributes, and distinguish between normal and abnormal activities with greater flexibility. As a result, intelligent intrusion detection systems have become one of the most important research areas in cybersecurity. [**3**].

Large cybersecurity datasets also play a central role in evaluating intrusion detection performance. Benchmark datasets such as CICIDS2017 contain realistic communication scenarios and multiple categories of attack traffic, making them suitable for experimental analysis and model comparison [**4**].

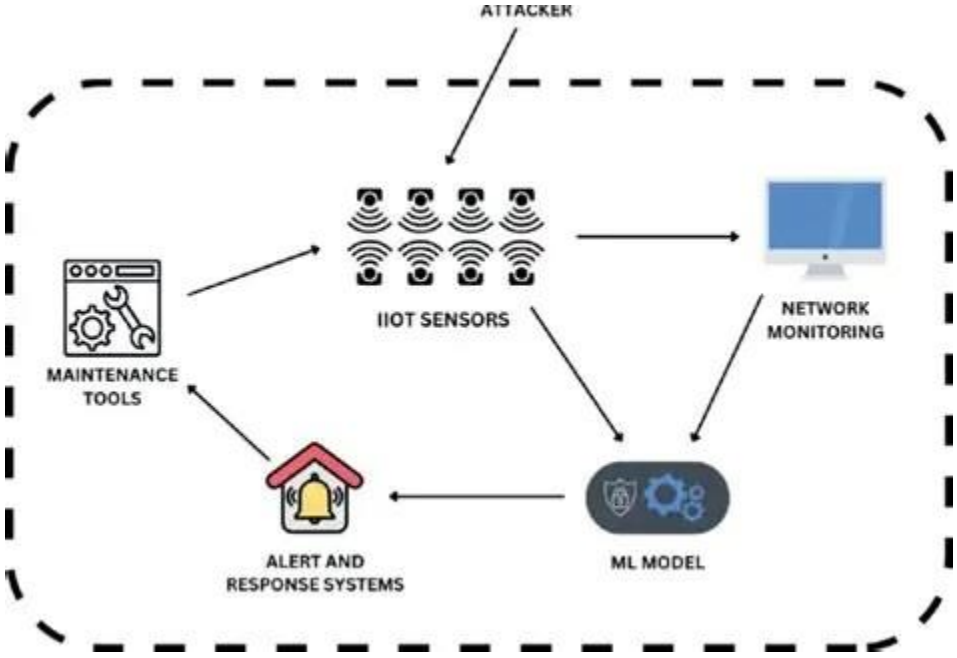

Figure 1 Cyberattack Detection Framework in IIoT Networks

This figure illustrates a cybersecurity attack detection framework in Industrial Internet of Things (IIoT) networks using machine learning techniques. The diagram demonstrates how an attacker targets IIoT sensors and attempts to disrupt network operations through malicious activities. The collected network traffic is continuously analyzed by the network monitoring system and processed by a machine learning (ML) model to identify abnormal

Another challenge commonly observed in intrusion detection datasets is the imbalance between normal records and malicious samples. In many situations, attack instances represent only a small portion of the total traffic data, which may negatively affect the learning process. Oversampling techniques such as SMOTE are therefore widely used to improve class balance and increase attack detection sensitivity**.**

Accordingly, this research proposes an anomaly-based intrusion detection framework that combines Random Forest ensemble learning with SMOTE preprocessing. The proposed model is compared with CNN, Logistic Regression, SVM, and Gradient Boosting techniques to evaluate its effectiveness in identifying cyber threats within IoT communication systems. [**5**].

2. **Previous Studies and Critical Analysis**

Recent years have witnessed substantial progress in the development of intelligent intrusion detection systems for IoT environments. Researchers have explored different machine learning and deep learning approaches to improve attack detection performance and strengthen cybersecurity protection. [**6**].

One major research direction focuses on deep learning architectures such as Convolutional Neural Networks (CNNs) [2]. These models are capable of automatically extracting complex features from large-scale datasets and identifying hidden patterns within network traffic. Several studies reported high classification accuracy when CNN-based models were applied to intrusion detection tasks. However, deep learning approaches generally require high computational power, large memory resources, and extended training times. Such requirements reduce their practicality in resource-constrained IoT environments. [**7**].

Another category of studies investigated traditional machine learning algorithms, including Logistic Regression and Support Vector Machine (SVM) . These methods are usually characterized by lower computational complexity and faster execution speed compared with deep learning architectures. Consequently, they are more suitable for lightweight IoT devices. Nevertheless, experimental evaluations revealed that these algorithms often experience reduced effectiveness when dealing with highly nonlinear attack behaviors and imbalanced datasets.

Several researchers also examined feature selection and dimensionality reduction techniques to improve intrusion detection efficiency. These approaches attempt to remove redundant information and reduce computational overhead. Although feature engineering methods can improve classification performance in some cases, they may also lead to information loss if important network characteristics are removed during preprocessing. [**8**].

Another important area of cybersecurity research concerns dataset imbalance handling techniques. In many intrusion detection datasets, attack samples are significantly fewer than normal traffic records. This imbalance can negatively influence learning performance and increase classification bias toward majority classes. To address this issue, researchers frequently employ oversampling techniques such as SMOTE to generate synthetic attack samples and improve model sensitivity toward rare cyber threats.

Despite the advances achieved by previous studies, several limitations remain unresolved. Some deep learning models provide high detection accuracy but require computational resources unsuitable for real-time IoT applications. On the other hand, lightweight algorithms may fail to recognize sophisticated attacks due to limited representation capabilities. Therefore, there is still a need for an intrusion detection framework capable of achieving a balance between computational efficiency, detection accuracy, and robustness against evolving cyber threats.

This research attempts to address these limitations by integrating Random Forest ensemble learning with SMOTE-based data balancing. The proposed approach aims to enhance attack detection capability while maintaining reasonable computational complexity suitable for IoT communication systems. [**9**].

3. Research Gap and Scientific Contribution

A critical examination of existing intrusion detection studies indicates that current cybersecurity solutions still face several practical and technical challenges. Many previously proposed approaches either focus primarily on maximizing detection accuracy without considering computational efficiency or emphasize lightweight implementation while sacrificing detection reliability. [**10**].

In addition, several existing systems experience difficulties in identifying rare attack categories due to the imbalance between normal and malicious traffic records. This issue frequently causes machine learning models to become biased toward dominant traffic classes, resulting in higher false negative rates and reduced sensitivity toward sophisticated cyberattacks. [**11**].

The present study contributes to the field through two primary scientific aspects:

1. Adoption of Random Forest Ensemble Learning: The proposed framework utilizes the Random Forest algorithm because of its strong capability to process high-dimensional network traffic data while minimizing overfitting problems. The ensemble structure based on multiple decision trees improves classification stability and enhances anomaly detection reliability. [**12**].
2. Integration of SMOTE-Based Data Balancing: To address class imbalance issues, the study incorporates the SMOTE algorithm during preprocessing. This strategy generates synthetic minority attack samples and improves the capability of the detection model to identify uncommon cyber threats. [**13**]

The proposed framework aims to provide a practical and dependable intrusion detection solution suitable for modern IoT communication environments.

4. Research Methodology and Experimental Framework The methodology adopted in this research is based on a systematic experimental framework designed to evaluate the effectiveness of machine learning algorithms in

cyberattack detection. The implementation process included several sequential stages to ensure accurate and reliable experimental outcomes. [**14**].

4.1 Dataset Preparation and Preprocessing The CICIDS2017 dataset was selected as the primary benchmark for experimental evaluation because it contains realistic traffic records and multiple categories of cyberattacks. Before model training, several preprocessing operations were performed.

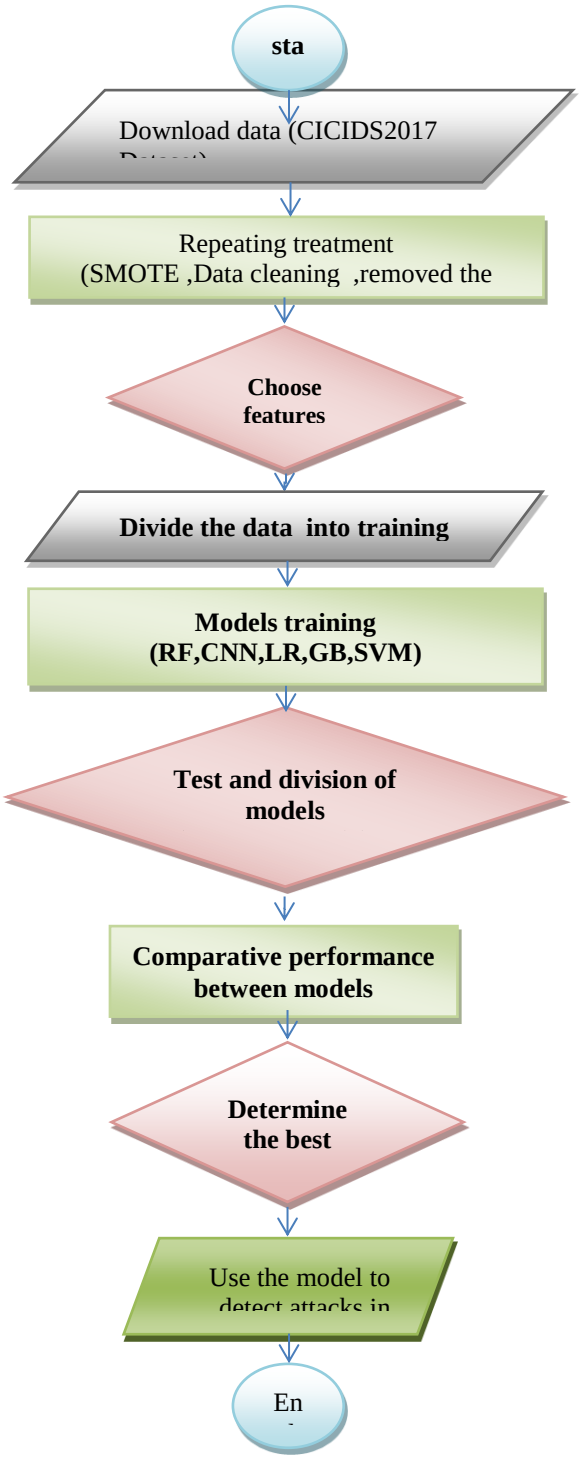

Figure 2: flow chart

Data Cleaning Records containing missing values or infinite numerical values were removed to avoid instability during model training and classification.

Feature Encoding Categorical network features were transformed into numerical representations using Label Encoding techniques because machine learning algorithms cannot directly process textual attributes. [**15**].

Data Normalization Min-Max normalization was applied to scale numerical features within the range of 0 to 1. This process prevented attributes with large numerical values from dominating the classification procedure.

4.2 Data Balancing Strategy \The SMOTE algorithm was implemented to overcome the imbalance between attack traffic and legitimate network traffic . Unlike simple duplication methods, SMOTE creates synthetic minority-class samples based on neighboring data instances. This strategy enhanced dataset balance and improved the capability of machine learning models to detect rare attacks.

4.3 Random Forest Mathematical Framework \The proposed intrusion detection system relies on the Random Forest ensemble learning technique [**16**]. Multiple decision trees were independently constructed using random subsets of the dataset to improve diversity and reduce prediction variance. The final classification decision was generated through majority voting among all decision trees. This mechanism contributed to improving classification stability, reducing false alarm rates, and enhancing overall detection accuracy.

4.4 Proposed Framework Workflow

The proposed framework consists of the following stages:

1. Data collection from the CICIDS2017 dataset.
2. Data preprocessing and normalization.
3. Balancing attack categories using SMOTE.
4. Dividing the dataset into training and testing subsets.
5. Training multiple machine learning models.
6. Evaluating classification performance.
7. Comparing model outputs to identify the best-performing approach. [**17**].

4.5 Comparative Machine Learning Models

To provide a comprehensive comparison, several machine learning and deep learning algorithms were evaluated: [**18**]

- Convolutional Neural Networks (CNN)
- Logistic Regression (LR)
- Random Forest (RF)
- Support Vector Machine (SVM)
- Gradient Boosting (GB)

Each model was trained and tested under similar conditions to ensure fair experimental evaluation.

4.6 Experimental Environment

All simulations were conducted using the Python programming language within the Anaconda development environment [**19**]. Several scientific libraries, including NumPy, Pandas, and Scikit-learn, were employed for preprocessing, model training, and performance evaluation.

The CICIDS2017 dataset contains approximately 2.8 million traffic records collected over several days and includes multiple communication activities such as HTTP, SSH, and Email traffic. The dataset also contains different attack categories including DDoS attacks, brute-force intrusions, botnet activities, and infiltration attacks.

5. Implementation

The implementation stage aimed to evaluate the capability of multiple machine learning approaches to identify malicious activities within IoT network traffic.

The process began with preparing the CICIDS2017 dataset through a sequence of preprocessing operations designed to improve data quality and learning stability. Missing records and irrelevant values were removed, while feature normalization techniques were applied to standardize numerical attributes.

Because cybersecurity datasets often suffer from class imbalance problems, SMOTE was utilized to generate balanced attack samples. This preprocessing step reduced classification bias toward normal traffic classes and improved the capability of the algorithms to identify rare attacks. [20].

After preprocessing and balancing, the dataset was divided into training and testing subsets. The training data was used to construct the machine learning models, whereas the testing subset was reserved for performance evaluation using previously unseen traffic records. [21].

Five different machine learning and deep learning techniques were evaluated, including Random Forest, CNN, Logistic Regression, SVM, and Gradient Boosting. The performance of each model was measured using several evaluation metrics derived from the confusion matrix, including accuracy, precision, recall, and F1-score.

The overall implementation framework enabled a systematic comparison between the selected algorithms and assisted in identifying the most efficient approach for securing IoT-enabled communication systems.

6. Results

The experimental evaluation revealed substantial differences in the performance of the investigated machine learning and deep learning models when applied to intrusion detection tasks within the CICIDS2017 environment.

The Random Forest classifier achieved the strongest overall performance among all evaluated approaches, which is consistent with findings reported in previous cybersecurity studies [22]. The model produced an accuracy of 99.42%, precision of 99.31%, recall of 99.27%, and an F1-score of 99.29%. These findings indicate that ensemble learning techniques provide strong classification capability for detecting malicious activities within complex IoT network traffic.

Gradient Boosting also achieved competitive results, although its performance remained slightly lower than Random Forest. Support Vector Machine and CNN models demonstrated acceptable performance levels but faced limitations related to computational complexity and classification stability. [24-23].

On the other hand, Logistic Regression recorded the weakest performance among the evaluated algorithms. The relatively low results can be attributed to the linear structure of Logistic Regression, which limits its capability to represent the complex and nonlinear relationships associated with modern cyberattack behaviors.

The confusion matrix analysis further confirmed the effectiveness of the Random Forest framework. The model successfully distinguished between legitimate and malicious traffic records while maintaining low false positive and false negative rates. [25].

Table 1. Performance Evaluation of the Proposed Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
RF	99.42	99.31	99.27	99.29
GB	98.84	98.72	98.65	98.68
SVM	97.95	97.61	97.44	97.52
CNN	97.38	97.12	96.88	97.00
LR	92.64	91.87	91.23	91.55

The results presented in Table 1 reveal that the Random Forest (RF) model achieved the highest performance among all tested models, recording an accuracy of 99.42%, precision of 99.31%, recall of 99.27%, and F1-score of 99.29%. This superior performance confirms the effectiveness of ensemble learning approaches in detecting sophisticated cyber threats within high-dimensional IoT environments.

The obtained results also demonstrated the importance of SMOTE during preprocessing. The generated synthetic minority samples significantly improved attack detection sensitivity and enhanced the capability of the models to identify rare and sophisticated cyber threats.

Overall, the findings indicate that combining Random Forest with SMOTE provides a reliable and efficient cybersecurity framework for protecting IoT communication infrastructures.

7. Discussion

The obtained results confirm that intelligent learning algorithms can significantly enhance intrusion detection capability within IoT communication systems. Among the investigated models, the Random Forest classifier demonstrated the most balanced performance in terms of stability, classification accuracy, and attack recognition capability.

The strong performance of the Random Forest model is mainly related to its ensemble-based architecture. Instead of relying on a single decision structure, the algorithm combines multiple decision trees to generate the final prediction. This strategy reduces instability during classification and improves resistance against noisy or highly variable network traffic conditions.

Another important observation concerns the role of SMOTE during preprocessing. The balanced distribution of attack samples improved the capability of the learning models to recognize minority attack categories that are usually difficult to identify in imbalanced cybersecurity datasets [26]. Consequently, recall values improved while false negative predictions decreased.

The comparative analysis also revealed several limitations associated with conventional algorithms. Logistic Regression produced lower performance because its mathematical structure depends mainly on linear relationships between features. Such assumptions are not always suitable for representing the complex behavioral characteristics of modern cyberattacks. [27].

Deep learning models such as CNN demonstrated acceptable detection capability; however, their training procedures required greater computational effort and longer execution times. These limitations reduce their practicality for real-time IoT environments where computational resources are restricted. [28].

The findings of this study suggest that combining ensemble learning with efficient preprocessing techniques can provide reliable cybersecurity protection for future communication infrastructures. Nevertheless, additional research is still required to improve scalability, reduce computational overhead, and enhance adaptability against continuously evolving cyber threats. Future investigations may also explore federated learning, blockchain-supported security architectures, and adaptive deep learning frameworks for distributed IoT systems [29]

8. Conclusion

This research introduced an intelligent intrusion detection framework for IoT communication environments based on Random Forest ensemble learning and SMOTE preprocessing techniques.

The proposed framework was experimentally evaluated using the CICIDS2017 benchmark dataset and compared with several machine learning and deep learning algorithms, including CNN, SVM, Logistic Regression, and Gradient Boosting.

Experimental evaluation demonstrated that the Random Forest classifier achieved the most reliable overall performance among the investigated approaches. The model produced high classification accuracy and maintained low false alarm rates while successfully identifying malicious network activities.

The study also confirmed that handling dataset imbalance through SMOTE preprocessing improves the capability of intrusion detection systems to recognize minority attack categories and sophisticated cyber threats.

Overall, the integration of ensemble learning and balanced data preprocessing provides an efficient and dependable cybersecurity solution suitable for modern IoT infrastructures and future smart communication systems.

References

[1] Wang, J., Na, Z., & Liu, X. (2020). Collaborative design of multi-UAV trajectory and resource scheduling for 6G-enabled Internet of Things systems. IEEE Internet of Things Journal, 8(20), 15096–15106.

[2] Khan, E. U., Qureshi, I. M., Aziz, M. A., Cheema, T. A., & Shah, S. B. H. (2020). Intelligent routing protocol for FANET communication systems. IEEE Access, 8, 56371–56378.

[3] Abdullahi, A., & Fathi, M. (2020). Intrusion detection systems for Internet of Things environments. Wireless Personal Communications, 112, 2057–2070.

[4] Khan, I. U., et al. (2021). Intelligent cyberattack detection approaches for aerial networks. Wireless Communications and Mobile Computing, 1–9.

[5] Symantec. (2019). Internet Security Threat Report. Symantec Corporation.

[6] Hussein, A., & Sharma, P. (2019). Signature-based intrusion detection technologies in computer networks. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 12(10), 60–64.

[7] Rakha, M., et al. (2024). Hybrid intrusion detection models for IoT smart cities. Cybersecurity for Next-Generation Computing Technologies, 159–176.

[8] Ganesan, R., et al. (2017). Scheduling optimization for cybersecurity analysts. ACM Transactions in Intelligent Systems and Technology, 8(4), 1–32.

[9] Wang, Y., et al. (2023). Deep learning techniques for anomaly detection. Sensors, 23(4), 2171.

[4] Dutta, M., & Granjal, J. (2020). Secure Internet of Things defense mechanisms. IEEE Access, 8, 127272–127312.

[11] Lal, N., et al. (2022). Deep learning approaches for anomaly detection in social networking systems. Springer.

[12] Jia, Y., et al. (2019). Deep neural network-based intrusion detection algorithm. IET Information Security, 13(1), 48–53.

[13] Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media.

[14] Awad, M., & Freihat, S. (2023). Feature selection methods for machine learning-based intrusion detection systems. Sensors, 12(5), 67.

[15] Amouri, A., et al. (2020). Machine learning-based intrusion detection for mobile IoT systems. Sensors, 20(2), 461.

[16] Lipman, R. P., et al. (2000). Evaluating intrusion detection systems. DARPA Information Survival Conference and Exposition.

[17] Mustafa, N., & Slay, J. (2015). UNSW-NB15 dataset for network intrusion detection systems. MilCIS.

[18] Ong, Y. L., et al. (2020). Virtual private network intrusion traps and datasets. Industrial Control Systems Security Workshop.

[19] Santos, V. F., et al. (2023). Machine learning evaluation for intrusion detection in cyber systems. Energies, 16(16), 6058.

[20] Ji, M., et al. (2019). Deep learning approaches for IoT intrusion detection. IEEE Pacific Symposium on Dependable Computing.

[21] Kodish, M., et al. (2021). Convolutional neural networks for IoT intrusion detection. PST 2021.

[22] De Cainge, A., et al. (2018). Hybrid classification algorithms based on logistic regression and decision trees. European Journal of Operational Research, 269(2), 760–772.

[23] Rahman, K., et al. (2023). Lightweight intrusion detection systems based on logistic regression for IoT environments. Mobile Information Systems.

[24] Biau, G., & Scornet, E. (2016). A guided tour in Random Forests. TEST, 25, 197–227.

[25] Regatti, S. J. (2017). Random Forest methods in cybersecurity systems. Journal of Insurance Medicine, 47(1), 31–39.

[26] Joseph, J. F. C., et al. (2010). SVM approaches for wireless network anomaly detection. IEEE Transactions on Trusted and Secure Computing, 8(2), 233–245.

[27] Javed, S. H., et al. (2022). Intelligent systems for advanced persistent threat detection in Industrial IoT. Electronics, 11(5), 742.

[28] Sharaf El-Din, I., et al. (2018). Creation of cybersecurity intrusion datasets. ICISSP, 108–116.

[29] Ullah, I., & Mahmoud, Q. H. (2020). Generating datasets for detecting anomalous IoT activities. Canadian Conference on Artificial Intelligence, 508–520.

BIOGRAPHIES OF AUTHORS

Haifa Hazim Al-Tai received her Master's degree in Computer Science from the University of Mosul in 2018, specializing in systems and database design. Her master's thesis, submitted to the College of Computer Science and Mathematics, focused on creating an intranet for human resources for the college system. She worked at the College of Nursing, University of Mosul, as an Associate Computer Operator, and later served as an Assistant Teacher since June 4, 2018. Her research interests include computer systems, database design, and information management systems. She has published several works, including "Creating an HR Intranet for the College System" and "Prevalence of Carpal Tunnel Syndrome among Computer Users at the University of Mosul." In addition, she has participated in academic seminars and served as a member of the examinations committee and the person responsible for the examination system. She can be contacted at email: Haifahzim123@uomosul.edu.iq.