


# Chapter 2

## Supervised Machine Learning Methods for Cyber Threat Detection Using Genetic Algorithm


**Daniel K. Gasu**

*University of Ghana, Ghana*

**Winfred Yaokumah**

 <https://orcid.org/0000-0001-7756-1832>  
*University of Ghana, Ghana*

**Justice Kwame Appati**

 <https://orcid.org/0000-0003-2798-4524>  
*University of Ghana, Ghana*

### ABSTRACT

*Security threats continue to pose enormous challenges to network and applications security, particularly with the emerging IoT technologies and cloud computing services. Current intrusion and threat detection schemes still experience low detection rates and high rates of false alarms. In this study, an optimal set of features were extracted from CSE-CIC-IDS2018 using genetic algorithm. Machine learning algorithms, including random forest, support vector machines, logistic regression, gradient boosting, and naïve bayes were employed for classification and the results compared. Evaluation of the performance of the proposed cyber security threat detection models found random forest as the highest attacks detection with 99.99% accuracy. K-nearest neighbor achieved 99.96% while a detection accuracy of 97.39% was obtained by support vector machines. The model which used gradient boosting obtained an accuracy of 99.97%, and the logistic regression model achieved a 94.94% accuracy. The lowest accuracy rate was obtained by the naïve bayes model with a detection accuracy of 68.84%.*

DOI: 10.4018/978-1-6684-7702-1.ch002

## **INTRODUCTION**

Cybersecurity measures contain processes, tools, and technologies with various cyber defense systems designed to protect information systems infrastructure (Wazid et al., 2022) to ensure the confidentiality, integrity, and availability of information assets (Almasoudy, Al-Yaseen, & Idrees, 2020; Dua & Xian, 2011; Thakkar & Lohiya, 2020). The cyber threats landscape keeps evolving and threat actors have become more sophisticated with advanced persistent threats (APTs) (Colorossi, 2015). Also, social engineering, ransomware, and fraud are committed through digital identity theft (Wazid et al., 2022). Network intrusion, malware attacks, phishing, unauthorized modification of information, and denial of service attacks negatively impact information systems (Arabo, Dijoux, Poulain, & Chevalier, 2020; Vani & Krishnamurthy, 2018). Though detection and prevention systems exist, attackers strive to evade or adapt to detection schemes to actively exploit vulnerabilities in systems. However, anomalies or sudden changes in systems and user behaviors can be detected if network systems are effectively monitored by intrusion detection systems and appropriate actions are taken (Muller et al., 2018).

Intrusion detection systems leveraging Machine learning (ML) techniques become crucial in detecting malicious activities (Singh et al., 2022), particularly zero-day attacks (Siddique, Akhtar, Lee, Kim, & Kim, 2017). Machine learning algorithms give computers the ability to learn directly from examples and experiences in the form of data points (The Royal Society, 2017; Wazid et al., 2022). They are capable of processing large amounts of data and then making predictions or decisions (Zou, Cui, Huang, & Zhang, 2008). Despite the recent deployments of ML in cyber threat detection, the development of robust and efficient intrusion detection systems remains an ongoing research problem (Gauthama Raman et al., 2017a). Machine learning methods may be employed in cybersecurity to select features from large datasets to improve intrusion detection rates and adaptability (Kunhare et al., 2022). However, most of the techniques suffer some setbacks including dependency on domain knowledge, big data issues resulting in insufficient learning capability, and the apparent lack of modularity and transferability (Wang, 2018). Moreover, intrusion detection systems may encounter high error rates, low true positive rates, and low accuracy. These challenges limit the effectiveness of intrusion detection systems in cybersecurity operations.

Improving the detection of network intrusion depends largely on the comprehensiveness of the dataset used in the training of the ML models and the feature set selected from the dataset (Saibene & Gasparini, 2023). Feature selection is a complex NP-hard problem, nonetheless, it is a determinant of the performance of classification models (Abdulhussien et al., 2023; Vijayanand et al., 2018). Whereas several feature selection methods have been applied to improve ML classifiers, the

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/supervised-machine-learning-methods-for-cyber-threat-detection-using-genetic-algorithm/328930](http://www.igi-global.com/chapter/supervised-machine-learning-methods-for-cyber-threat-detection-using-genetic-algorithm/328930)

## Related Content

---

### A Lightweight Authentication and Encryption Protocol for Secure Communications Between Resource-Limited Devices Without Hardware Modification: Resource-Limited Device Authentication

Piotr Ksiazak, William Farrelly and Kevin Curran (2021). *Research Anthology on Artificial Intelligence Applications in Security* (pp. 586-630).

[www.irma-international.org/chapter/a-lightweight-authentication-and-encryption-protocol-for-secure-communications-between-resource-limited-devices-without-hardware-modification/270617](http://www.irma-international.org/chapter/a-lightweight-authentication-and-encryption-protocol-for-secure-communications-between-resource-limited-devices-without-hardware-modification/270617)

### Is It the End of Undergraduate Dissertations?: Exploring the Advantages and Challenges of Generative AI Models in Education

Benjamin Kenwright (2024). *Generative AI in Teaching and Learning* (pp. 46-65).

[www.irma-international.org/chapter/is-it-the-end-of-undergraduate-dissertations/334772](http://www.irma-international.org/chapter/is-it-the-end-of-undergraduate-dissertations/334772)

### Configuration Pathways to Enhance Green Total Factor Productivity: A Fuzzy Set Qualitative Comparative Analysis

Yani Guo and Yunjian Zheng (2023). *International Journal of Fuzzy System Applications* (pp. 1-16).

[www.irma-international.org/article/configuration-pathways-to-enhance-green-total-factor-productivity/326798](http://www.irma-international.org/article/configuration-pathways-to-enhance-green-total-factor-productivity/326798)

### A Comparative Study for Position Regulation and Anti-Swing Control of Highly Non-Linear Double Inverted Pendulum (DIP) System Using Different Soft Com

Ashwani Kharola and Pravin P. Patil (2017). *International Journal of Fuzzy System Applications* (pp. 59-81).

[www.irma-international.org/article/a-comparative-study-for-position-regulation-and-anti-swing-control-of-highly-non-linear-double-inverted-pendulum-dip-system-using-different-soft-com/179321](http://www.irma-international.org/article/a-comparative-study-for-position-regulation-and-anti-swing-control-of-highly-non-linear-double-inverted-pendulum-dip-system-using-different-soft-com/179321)

## Multi-Agent Systems Integration in Enterprise Environments Using Web Services

Eduardo H. Ramirez and Ramón F. Brena (2006). *International Journal of Intelligent Information Technologies* (pp. 72-88).

[www.irma-international.org/article/multi-agent-systems-integration-enterprise/2406](http://www.irma-international.org/article/multi-agent-systems-integration-enterprise/2406)