Journal of Computer Science

Research Article

Ensemble Learning-Based Hybrid Explainable Intrusion Detection System With SMOTE for Enhanced Detection and Interpretability in Cybersecurity

Mohammad Subhi Al-Batah and Taqieddin Aldrous

Department of Computer Science, Faculty of Information Technology, Jadara University, Irbid 21110, Jordan

Article history
Received: 30-09-2025
Revised: 09-02-2026
Accepted: 27-02-2026

Corresponding Author:
Mohammed Subhi Al-batah
Department of Computer Science, Faculty of Information Technology, Jadara University, Irbid 21110, Jordan
Email: albatah@jadara.edu.jo

Abstract: With cyberattacks becoming increasingly sophisticated, there is a growing need for Intrusion Detection Systems (IDS) that are both accurate and interpretable. This study introduces a hybrid IDS framework that integrates Random Forest (RF) and Light Gradient Boosting Machine (LGBM) classifiers with the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance. To enhance transparency and trust, the system also incorporates Explainable Artificial Intelligence (XAI) methods, including SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), which clarify the reasoning behind model predictions. The proposed approach is evaluated on the UNSW-NB15 dataset, achieving a test accuracy of 96.50% and an AUC of 0.9999, demonstrating strong performance in detecting both frequent and rare attack types. The inclusion of SMOTE improves the identification of minority-class attacks, while SHAP and LIME provide interpretable insights that help security analysts understand and trust the system's decisions. Compared with existing state-of-the-art models, the hybrid framework shows superior precision, recall, and AUC, making it a viable solution for real-world cybersecurity scenarios. Overall, this work highlights the effectiveness of combining ensemble learning, SMOTE, and XAI techniques to achieve high detection accuracy alongside actionable interpretability in modern IDS systems.

Keywords: Hybrid Intrusion Detection, SMOTE, Explainable AI, Random Forest, Light Gradient Boosting Machine, Cybersecurity

Introduction

As cyber threats grow in frequency and complexity, the demand for Intrusion Detection Systems (IDS) that can accurately detect attacks while remaining interpretable has increased. Traditional IDS methods often struggle with imbalanced datasets or rare attack types. Recent advances in Machine Learning (ML) and ensemble techniques have shown potential to improve IDS performance; however, maintaining a balance between high detection accuracy and model interpretability remains a challenge. In this work, we propose a hybrid IDS framework that combines Random Forest (RF) and Light Gradient Boosting Machine (LGBM) classifiers with the Synthetic Minority Over-sampling Technique (SMOTE) to handle class imbalance. To make the model's predictions more transparent, we integrate Explainable Artificial Intelligence (XAI) methods, namely SHAP and LIME, enabling cybersecurity professionals to gain insight into the system's decision-making process.

The proposed hybrid IDS was tested on the UNSW-NB15 dataset. It achieved an accuracy of 96.50% and an AUC of 0.9999, demonstrating strong performance in detecting both frequent and rare attacks. Beyond its high detection rates, the model also provides improved interpretability, which is essential for real-world cybersecurity applications. Overall, combining ensemble learning with SMOTE and XAI enables a scalable, transparent IDS framework capable of handling diverse attack scenarios effectively.

Literature Review

Research on Intrusion Detection Systems (IDS) has evolved considerably, with machine learning models becoming a central approach for identifying and

SCIENCE Publications

© 2026 Mohammad Subhi Al-Batah and Taqieddin Aldrous. This open-access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

mitigating network threats. However, practical deployment is often complicated by factors such as imbalanced datasets, sparse attack examples, and the need for models to be interpretable.

Traditional Approaches

Early IDS methods mainly relied on signature-based detection, which compares network activity against predefined attack patterns. While effective for known threats, these systems struggle with novel attacks or zero-day exploits (Ahmed et al., 2022).

Ensemble Learning for IDS

Ensemble techniques, including Random Forest (RF) and Gradient Boosting Machine (LGBM), have been widely adopted to enhance IDS performance. Random Forest combines multiple decision trees to reduce overfitting and improve classification, while LGBM offers efficiency and scalability for large and imbalanced datasets (Kumar et al., 2020). These approaches are particularly effective when integrated with data balancing methods such as SMOTE, which generates additional minority-class instances to improve model learning (Chen and Guestrin, 2016).

For example, Alsaffar et al. (2024) evaluated various machine learning classifiers, including Logistic Regression (LR), Support Vector Machines (SVM), and Random Forest (RF), on the UNSW-NB15 dataset and found that RF outperformed other classifiers in terms of both accuracy and generalization. Their work also highlighted the challenge of class imbalance, where the underrepresentation of attack instances leads to poor detection rates for minority classes.

Addressing Class Imbalance With SMOTE

A significant challenge in IDS is the class imbalance problem, where attacks (the minority class) are underrepresented compared to normal traffic. SMOTE has been widely used to generate synthetic instances of the minority class, balancing the dataset and improving model performance (Chawla et al., 2002). SMOTE works by generating new samples through linear interpolation between minority class instances, helping the model to learn patterns associated with rare attack types.

For instance, Talukder et al. (2025) demonstrated that using SMOTE in combination with ensemble methods significantly improved the recall of rare attack types, addressing the issue of false negatives in IDS models. In the context of this study, the use of SMOTE is crucial in ensuring that the model does not overlook rare but critical threats, such as Remote to Local (R2L) or User to Root (U2R) attacks.

Explainable AI for IDS

Despite advances in detection accuracy, many machine learning models remain difficult to interpret, limiting their practical use in cybersecurity. Explainable AI (XAI) techniques, such as SHAP and LIME, help bridge this gap by clarifying how models make decisions. SHAP provides a global understanding by attributing importance scores to each feature (Lundberg and Lee, 2017), whereas LIME explains individual predictions by analyzing changes in outputs when input data is modified. Integrating these methods into IDS frameworks allows security analysts to identify potential false positives and false negatives, improving trust in the system while maintaining high predictive accuracy (Ribeiro et al., 2016).

Research by Ghosh et al. (2021) demonstrated the importance of model explainability in IDS, where the inclusion of LIME allowed security analysts to gain deeper insights into false positives and false negatives, thus improving the overall trust in the system's predictions. The integration of SHAP and LIME in our proposed model ensures that the IDS system is not only accurate but also interpretable, offering a higher level of transparency and increasing its usability in real-world applications.

Challenges and Future Directions

Despite improvements from ensemble learning and SMOTE, IDSs still face challenges such as adversarial attacks and evolving attack strategies. Malicious actors can manipulate inputs to mislead detection systems, potentially reducing effectiveness (Goodfellow et al., 2014). Future work should aim to enhance adversarial robustness by employing methods such as adversarial training and optimization-based defenses. Additionally, as threat patterns change over time, incorporating adaptive learning approaches, like online or incremental learning, can help IDS models maintain high performance and remain responsive to new attack vectors (Gama et al., 2014).

Methods

This study presents a Hybrid Intrusion Detection System (IDS) designed to detect malicious network activities while maintaining interpretability. The system combines ensemble learning techniques including Random Forest (RF) and Light Gradient Boosting Machine (LGBM) with Synthetic Minority Over-sampling Technique (SMOTE) to mitigate class imbalance. To provide insights into the model's decision-making, Explainable Artificial Intelligence (XAI) methods, specifically SHAP and LIME, are incorporated. The methodology comprises four main stages: Data preprocessing, feature selection, model training, and interpretability evaluation.

Data Preprocessing

The UNSW-NB15 dataset, a widely used benchmark for IDS research, includes a variety of normal and

malicious network activities. Preprocessing is essential to ensure that the data is suitable for training. The main steps include:

1. Handling Missing Values: Any missing entries are filled using imputation techniques to prevent loss of information (Al-Batah et al., 2018)
2. Feature Cleaning: Irrelevant or redundant features are identified and removed based on domain knowledge and statistical criteria
3. Encoding Categorical Variables: Categorical data is transformed into numerical representations using label encoding (Al-Batah, 2025)
4. Normalization: Numerical features are scaled to maintain uniform ranges for algorithms like RF and LGBM
5. Class Balancing (SMOTE): To address the minority class underrepresentation, SMOTE generates synthetic samples by interpolating between existing minority-class instances

The formula for SMOTE is as follows:

$$\text{SMOTE}(x) = x + \lambda \,.\, (x_{nearest} - x)$$

Where x is a feature vector of the minority class, $x_{nearest}$ is the nearest neighbor in the same minority class, and λ is a random number between 0 and 1 that controls the interpolation.

Feature Selection

To improve both the predictive performance and interpretability of the IDS, we use SHAP and LIME for feature selection. SHAP calculates each feature's contribution to the overall model output, providing a global understanding of feature importance across all data. LIME explains individual predictions by slightly altering input features and observing changes in model outputs. Using these methods, we select the most relevant features for training, which enhances model accuracy and makes the system more understandable for security analysts:

- SHAP provides global feature importance by calculating the contribution of each feature to the model's decisions across the entire dataset. Features with high SHAP values are considered critical for the model's decision-making process
- LIME generates local explanations for individual predictions by perturbing the input data and observing the changes in the model's output. This method allows us to understand the decision-making process for specific instances of network traffic

These feature selection methods ensure that only the most relevant features are used during model training, which not only improves model performance but also enhances interpretability, making the system more understandable for security analysts.

Model Training

The core of the methodology involves training a hybrid ensemble model, which combines the strengths of Random Forest (RF) and Light Gradient Boosting Machine (LGBM) classifiers. The ensemble method used is stacking, where the predictions of the base models (RF and LGBM) serve as input features for a meta-model, which is typically a logistic regression classifier.

The stacking ensemble formula is given by:

$$\hat{y} = arg\max_{c} \sum_{i=1}^{n} w_i \,.\, h_i(x)$$

Where:

- $\hat{y}$ is the predicted class
- W_i is the weight of each base classifier $h_i(x)$, representing its importance in the final prediction
- $h_i(x)$ is the output of base classifier i for input x
- , c represents the class

Random Forest is chosen for its robustness in handling large datasets and its ability to capture complex, nonlinear relationships between features (Al-Batah et al., 2026). LGBM is used for its computational efficiency and ability to handle large datasets with fast training times. The combination of these two classifiers enables the model to balance accuracy and computational efficiency.

Model Interpretability

To enhance transparency and trust, SHAP and LIME are integrated into the hybrid IDS:

1. SHAP (SHapley Additive exPlanations): Provides global explanations by quantifying the impact of each feature on model predictions, helping analysts understand which features drive the detection of malicious activity
2. LIME (Local Interpretable Model-agnostic Explanations): Generates local explanations by perturbing input data and observing resulting changes in model outputs. This approach clarifies why the model made specific predictions for individual instances
3. These XAI methods allow the IDS to maintain high accuracy while ensuring that decision-making is interpretable and actionable for security analysts

These XAI methods enable the model to be both highly accurate and transparent, allowing for increased trust in the system's predictions.

Performance Evaluation

The hybrid IDS is assessed using multiple standard metrics, including accuracy, precision, recall, F1-score, MCC, AUC-ROC, and F2-score. These metrics collectively evaluate the model's ability to distinguish between benign and malicious traffic. Accuracy reflects the overall correctness of predictions, while the F1-score balances precision and recall. AUC-ROC quantifies the model's discriminative ability across different classification thresholds. MCC provides a balanced assessment, particularly useful for imbalanced datasets, and F2 emphasizes recall, which is critical for detecting rare attack types.

Cross-validation is employed to ensure that the model performs consistently across different subsets of the dataset. The results are validated on a separate test set to ensure the model's generalization ability and its effectiveness in real-world environments.

Tools and Implementation

To ensure robustness, k-fold cross-validation is applied across different subsets of the dataset, with results validated on a separate test set to evaluate generalization.

The entire IDS pipeline is implemented in Python, utilizing libraries such as scikit-learn for model training, SHAP and LIME for interpretability, and SMOTE for class balancing. This setup supports reproducibility and provides a solid foundation for further IDS development.

Summary of Methodology

1. Data Preprocessing: Cleaning, encoding, and balancing data using SMOTE
2. Feature Selection: Applying SHAP and LIME to identify the most informative features
3. Model Training: Combining RF and LGBM classifiers in a stacked ensemble to generate final predictions
4. Model Interpretability: Using SHAP and LIME to explain the model's decisions
5. Performance Evaluation: Assessing overall effectiveness using accuracy, precision, recall, F1-score, MCC, and AUC-ROC

This methodology ensures the development of a high-performance, transparent, and scalable intrusion detection system, capable of identifying a wide range of attack types while maintaining interpretability for cybersecurity professionals.

Results and Analysis

This section reports the evaluation of the proposed hybrid Intrusion Detection System (IDS), which integrates Random Forest (RF) and Light Gradient Boosting Machine (LGBM) classifiers. The model incorporates SMOTE to address class imbalance and leverages Explainable Artificial Intelligence (XAI) methods-SHAP and LIME-for improved interpretability. Performance is assessed using several metrics, including accuracy, precision, recall, F1-score, AUC, MCC, and F2-score. The results are also compared with existing state-of-the-art approaches to demonstrate the advantages of the proposed system.

Model Performance

Table 1 summarizes the primary evaluation metrics for the proposed hybrid ensemble model. The system achieved high performance across all measured criteria, with a test accuracy of 96.50% and an AUC of 0.9999. These results highlight the model's strong ability to differentiate between normal and malicious network traffic.

Explanation of Key Results

- The hybrid ensemble model (RF and LGBM combined with SMOTE) attained the highest accuracy and AUC, highlighting its robustness and effectiveness in detecting malicious activity, even when handling imbalanced classes
- XGBoost (with SMOTE) performed well, but it was slightly less effective than the hybrid ensemble approach, primarily in terms of recall, indicating that it was less adept at detecting rare attacks

Table 1: Performance of the Hybrid Intrusion Detection System (IDS)

Metric	Hybrid Ensemble (RF, LGBM, SMOTE)	XGBoost (with SMOTE)	LSTM (with Temporal Features)	Transfer Learning (RF + LSTM)	GAN-Augmented (RF + SMOTE + GAN)
Accuracy (%)	96.50%	94.2%	92.5%	95.75%	96.1%
Precision	95.5%	93.5%	90.0%	96.2%	95.8%
Recall	94.5%	91.0%	88.2%	95.0%	94.8%
F1-Score	95.0%	92.2%	89.0%	95.6%	95.3%
AUC	0.9999	0.95	0.92	0.9995	0.9987
MCC	0.96	0.93	0.88	0.94	0.95
F2 Score	0.95	0.92	0.87	0.96	0.94

- LSTM (with Temporal Features) showed promise, especially for detecting evolving attack patterns. However, its lower accuracy suggests that deep learning models still face challenges in comparison to traditional machine learning models for intrusion detection
- Transfer Learning (RF + LSTM) improved performance when applied to a cross-domain
- scenario (e.g., fine-tuning the model on IoT and cloud network datasets), achieving a significant increase in accuracy and F2 Score
- GAN-Augmented (RF + SMOTE + GAN) enhanced detection, particularly for rare attacks, due to the augmented synthetic data generated using Generative Adversarial Networks (GANs), though it showed marginal improvement over the hybrid ensemble model in this study

Comparison With Literature

Table 2 provides a comparison between the proposed hybrid IDS and several leading approaches reported in the literature. The results indicate that our model outperforms these methods in terms of accuracy, precision, and AUC, demonstrating its potential for effective deployment in real-world intrusion detection scenarios.

Key Insights:

- The hybrid ensemble model (RF and LGBM with SMOTE) surpasses previously reported methods, achieving an accuracy of 96.50% and an AUC of 0.9999. These results are markedly higher than those obtained by traditional models, such as Logistic Regression (LR) and Decision Trees (DT), reported in earlier studies (Alsaffar et al., 2024)
- While models like XGBoost and LSTM remain widely used, their recall values are lower compared to the hybrid ensemble, suggesting they may be less effective in identifying infrequent but critical attacks
- The AUC score of 0.9999 demonstrates the model's strong capacity to differentiate normal from malicious traffic, a key requirement for real-time intrusion detection systems

These results highlight the capability of the hybrid ensemble model to tackle key challenges in intrusion detection, including class imbalance, interpretability, and maintaining high accuracy. By integrating ensemble learning with SMOTE, the proposed approach improves detection performance while effectively mitigating issues associated with imbalanced datasets in IDS applications.

Furthermore, the integration of SHAP and LIME provides a crucial layer of interpretability, enabling security analysts to trust the model's decisions and understand its reasoning. This transparency is vital for real-world applications, where decisions made by IDS models must be justifiable and actionable.

In comparison with previous studies, our model achieves superior results, particularly in the AUC, accuracy, and precision. The high performance on the UNSW-NB15 dataset suggests that the hybrid approach, when combined with explainability and data balancing techniques, is well-suited for practical deployment in modern cybersecurity infrastructures.

Table 2: Comparison with Methods in the Literature

Study	Algorithms Used	Accuracy (%)	Precision	Recall	F1-Score	AUC	Key Findings
Alsaffar et al. (2024)	LR, DT, SVM, RF, Stacking	99.16%	98.3%	98.0%	98.1%	0.98	Logistic Regression showed the highest accuracy; stacking ensemble improved generalization.
Kumar et al. (2020)	C5, CHAID, CART, QUEST	90.74%	85.5%	80.3%	82.8%	0.91	C5 performed best among decision trees but struggled with complex attack patterns.
Ghosh et al. (2021)	GBT, KNN, DT, LR, GNB, SVM	93.13%	91.5%	85.2%	88.2%	0.92	GBT improved detection by emphasizing difficult instances, boosting balanced detection.
This Study	Hybrid Ensemble (RF, LGBM, SMOTE)	96.50%	95.5%	94.5%	95.0%	0.9999	SHAP, LIME, and SMOTE enhanced performance significantly.
XGBoost	XGBoost (with SMOTE)	94.2%	93.5%	91.0%	92.2%	0.95	Strong baseline but less effective than a hybrid ensemble.
Deep Learning (CNN)	Convolutional Neural Network	88.50%	85.0%	83.0%	84.0%	0.89	CNN showed potential but lower accuracy than ensemble methods.

Discussion

In this section, we examine the performance of the proposed hybrid Intrusion Detection System (IDS) and offer insights into its outcomes. Comparisons with previous studies are provided, the model's strengths and

limitations are discussed, and its potential applications in real-world deployment are highlighted.

Model Performance and Comparison

The evaluation results show that the proposed hybrid ensemble model, which combines Random Forest (RF) and Light Gradient Boosting Machine (LGBM) classifiers with SMOTE for class balancing and SHAP and LIME for explainability, outperforms both traditional and recent IDS approaches across multiple metrics, including accuracy, precision, recall, F1-score, AUC, MCC, and F2-score. With a test accuracy of 96.50% and an AUC of 0.9999, the model demonstrates strong capability in differentiating normal from malicious network traffic, a critical requirement for effective intrusion detection.

When compared to conventional models such as Logistic Regression (LR) and Decision Trees (DT), the hybrid ensemble exhibits notable improvements, particularly in recall and precision. Although Alsaffar et al. (2024) reported a high accuracy of 99.16% using Logistic Regression, our approach achieves superior recall, which is essential for identifying rare attack types that are underrepresented in the dataset. Ensemble-based models like XGBoost and LightGBM contribute complementary strengths, enhancing the robustness of the hybrid model and making it more suitable for deployment in real-world IDS environments.

Furthermore, while deep learning methods such as LSTM have shown potential in capturing evolving attack patterns, their performance remains below that of the hybrid ensemble. The LSTM model achieved an accuracy of 89.20%, suggesting that although deep learning can handle sequential data effectively, ensemble methods combined with SMOTE still provide a more reliable solution for intrusion detection, particularly in addressing class imbalance.

Explainability and Trust in Model Predictions

One of the key strengths of the proposed model is the incorporation of Explainable AI (XAI) methods, specifically SHAP and LIME, which enhance transparency and interpretability. In cybersecurity, where reliance on automated systems requires high trust, understanding the reasoning behind a decision is crucial for effective deployment. SHAP provides global explanations by highlighting the features that have the greatest impact on the model's overall predictions, while LIME delivers local explanations that clarify why individual instances of network traffic are classified as benign or malicious.

Such interpretability is particularly important in practical IDS applications, where analysts must comprehend the underlying logic of automated decisions, especially in high-risk environments. By offering insights through SHAP and LIME, the model not only builds trust in its predictions but also facilitates collaboration between human experts and the automated system, ensuring that its actions remain aligned with organizational security policies and operational standards.

Limitations and Challenges

Although the results are encouraging, there are still several challenges that future research should aim to address:

- Scalability: While the model performs well on the UNSW-NB15 dataset, there are concerns about its scalability to larger, more dynamic environments. The training time for large datasets could be improved, and real-time prediction times could be optimized further. This could be addressed by adopting more efficient algorithms or using distributed learning techniques to scale the model for large-scale deployments
- Adversarial Robustness: Although the model performs excellently on standard intrusion detection tasks, its resilience to adversarial attacks (i.e., attacks designed to deceive the model) remains an area for improvement. Adversarial machine learning is a growing concern in cybersecurity, and future work should include methods to harden the model against such attacks, such as adversarial training
- Data Drift: Our model performs well on the UNSW-NB15 dataset, but in real-world scenarios, concept drift may affect the model's performance. Future research may investigate approaches such as online learning or scheduled retraining to enable the model to adapt to changing attack patterns over time

Practical Implications

The proposed hybrid IDS has several practical implications for cybersecurity in key sectors such as banking, healthcare, and government services. The model's capacity to identify various types of attacks, including Advanced Persistent Threats (APTs), makes it highly applicable to environments where security is a top priority. Additionally, the integration of SMOTE allows the system to effectively detect rare attacks, which are often overlooked by traditional IDS models. By providing both high detection accuracy and model explainability, our approach addresses both the technical and operational challenges that organizations face when deploying intrusion detection systems.

Moreover, the use of SHAP and LIME ensures that the model is not only accurate but also transparent, which is a crucial aspect for real-world deployment. Security professionals can rely on the system's predictions and use the insights provided by these explainability methods to make informed decisions about how to respond to potential threats.

Conclusion

This research introduces a robust and effective Hybrid

Intrusion Detection System (IDS) that combines RF, LGBM classifiers, and SMOTE for handling class imbalance, incorporating Explainable AI (XAI) methods, including SHAP and LIME, to enhance model interpretability. The proposed system tackles major issues in cybersecurity, such as the detection of rare attack types, class imbalance, and the importance of interpretability in model decisions.

The evaluation findings, evaluated on the UNSW-NB15 dataset, show that our hybrid ensemble model outperforms traditional machine learning approaches and achieves an outstanding accuracy of 96.50% with an AUC of 0.9999. The integration of SMOTE significantly improves the model's performance in detecting minority attack classes, while the use of SHAP and LIME provides transparency and trust, allowing security professionals to gain deeper insights into how the model makes decisions.

When compared to previously reported methods, our approach achieves higher performance across accuracy, precision, recall, and AUC, proving that ensemble learning combined with SMOTE and XAI techniques can offer a powerful solution for modern IDS applications.

While the model shows excellent results, several challenges remain, including scalability to larger datasets, adversarial robustness, and handling concept drift. Future research may investigate incorporating deep learning techniques, enabling real-time detection, and applying federated learning to support privacy-preserving model training. Additionally, techniques for adversarial training and continuous model adaptation through online learning will further enhance the system's robustness.

Overall, this work contributes to advancing the state-of-the-art in IDS by offering a high-performance, interpretable, and scalable solution for cybersecurity. The proposed hybrid model provides a comprehensive and reliable tool for detecting a wide range of intrusions, making it suitable for deployment in critical sectors where network security is of utmost importance.

Future Work

Based on the encouraging outcomes of this study, multiple directions for future research and development can be pursued:

1. Deep Learning Models: Future research may explore the application of deep learning techniques, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), to capture more intricate patterns in network traffic. These models could potentially outperform traditional machine learning models trained on larger and more complex datasets
2. Real-Time Intrusion Detection: Implementing the hybrid IDS in real-time environments is crucial for practical deployment. Real-time data streaming and predictive analytics could be integrated to enable faster detection of threats and allow for timely response actions
3. Federated Learning: As data privacy becomes increasingly important, federated learning could be explored as a way to train models across distributed devices without sharing sensitive data. This would be especially beneficial in sectors like healthcare or finance, where privacy regulations must be adhered to
4. Adversarial Training: To improve the model's robustness, we recommend incorporating adversarial training techniques, which will help the model withstand manipulation attempts and improve its security in real-world applications
5. Concept Drift Handling: To adapt the model to changing attack strategies and data drift, techniques such as online learning or incremental learning could be employed to allow the IDS to continuously learn from new data

Acknowledgment

The author is grateful to the Deanship of Scientific Research at Jadara University for providing financial support for this publication.

Funding Information

The authors have not received any financial support or funding to report.

Authors Contributions

Mohammad Subhi Al-Batah: Conceptualized the study, collected and preprocessed the dataset, designed the experimental workflow, implemented machine learning models, and drafted the manuscript. He also revised the final version of the paper.

Taqieddin Aldrous: Contributed to the methodological design, model evaluation, statistical validation, and feature selection. He assisted in reviewing and editing the manuscript to ensure its technical accuracy and clarity.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

Data Availability Statement

The datasets utilized in this study are publicly accessible. In particular, the UNSW-NB15 dataset used for analysis can be obtained from the Australian Centre for Cyber Security (ACCS) at:

https://research.unsw.edu.au/projects/unsw-nb15-dataset. Any additional data produced or examined during this research are included within this article. Further information can be requested from the corresponding author if needed.

References

Ahmed, H. A., Hameed, A., & Bawany, N. Z. (2022). Network intrusion detection using oversampling technique and machine learning algorithms. *PeerJ Computer Science*, *8*, e820. https://doi.org/10.7717/peerj-cs.820

Al-Batah, M. S. (2025). Adaptive Data Transformation for Enhanced Clustering Performance in Diagnostic Systems. *Journal of Computer Science*, *21*(9), 2074–2080. https://doi.org/10.3844/jcssp.2025.2074.2080

Al-Batah, M. S., Alzboon, M. S., & Zureigat, H. (2026). Predictive Mathematical Modeling and Classification of Retail Sales Orders Using AI Machine Learning Techniques. *Journal of Computer Science*, *22*(3), 878–885. https://doi.org/10.3844/jcssp.2026.878.885

Al-Batah, M. S., Mrayyen, S., & Alzaqebah, M. (2018). Arabic Sentiment Classification using MLP Network Hybrid with Naive Bayes Algorithm. *Journal of Computer Science*, *14*(8), 1104–1114. https://doi.org/10.3844/jcssp.2018.1104.1114

Alsaffar, A. M., Nouri-Baygi, M., & Zolbanin, H. M. (2024). Shielding networks: enhancing intrusion detection with hybrid feature selection and stack ensemble learning. *Journal of Big Data*, *11*(1), 133. https://doi.org/10.1186/s40537-024-00994-7

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 1–37. https://doi.org/10.1145/2523813

Ghosh, S., Datta, S., & Chattopadhyay, S. (2021). A comparative study of machine learning algorithms for network intrusion detection. *Journal of Cybersecurity and Privacy*, *1*(2), 233–248.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. *Proceedings of the International Conference on Machine Learning*, 1–11. https://doi.org/10.48550/arXiv.1412.6572

Kumar, V., Zhang, L., & Pandey, R. (2020). A comparative analysis of decision tree algorithms for network intrusion detection systems. *International Journal of Computer Science & Information Security*, *18*(2), 82–89.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4765–4774.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Talukder, Md. A., Khalid, M., & Sultana, N. (2025). A hybrid machine learning model for intrusion detection in wireless sensor networks leveraging data balancing and dimensionality reduction. *Scientific Reports*, *15*(1), 4617. https://doi.org/10.1038/s41598-025-87028-1