AI-Driven Intrusion Detection for Encrypted Network Traffic

Ms. Gitanjali Kailas Patil
M. Sc in Computer Science, Assistant Professor, H.R. Patel Arts and Science College, Shirpur, Maharashtra

***Abstract*—The increasing use of encryption in modern communication networks has improved data confidentiality and privacy while creating new challenges for detecting malicious network activities. Traditional intrusion detection systems primarily depend on payload inspection and predefined attack signatures, which become less effective when network traffic is encrypted. This research proposes an artificial intelligence (AI)-driven intrusion detection framework designed to identify potential cyber threats in encrypted network traffic without requiring direct access to the underlying message content. The proposed approach focuses on analyzing network flow characteristics, including packet size, traffic duration, inter-arrival time, flow direction, protocol information, and statistical communication patterns. Machine learning and deep learning techniques can be employed to distinguish legitimate network behavior from suspicious activities, including denial-of-service attacks, scanning attempts, and anomalous communication patterns. Feature selection and preprocessing techniques are considered to improve detection efficiency, reduce computational overhead, and support timely threat identification. The framework also aims to address key challenges associated with encrypted traffic analysis, such as evolving attack patterns, class imbalance, false-positive detection, and variations in network behavior. Its performance can be evaluated using established network traffic datasets and metrics such as accuracy, precision, recall, F1-score, and false-positive rate. By combining traffic-flow analysis with intelligent classification, the proposed framework seeks to strengthen network security while preserving the confidentiality offered by encryption. The study is expected to contribute to the development of adaptive, scalable, and privacy-aware intrusion detection systems suitable for modern enterprise networks, cloud environments, and other encrypted communication infrastructures.**

***Index Terms*—Artificial Intelligence, Intrusion Detection System, Encrypted Network Traffic, Machine Learning, Deep Learning, Network Security, Anomaly Detection, Traffic Analysis.**

I. INTRODUCTION

The rapid expansion of digital communication, cloud computing, Internet of Things (IoT) devices, and online services has significantly increased the volume and complexity of network traffic. To protect sensitive information from unauthorized access, modern communication systems increasingly rely on encryption protocols such as Transport Layer Security (TLS) and other secure communication mechanisms. Although encryption provides essential confidentiality and data integrity, it also introduces difficulties for conventional network security systems that depend on inspecting packet payloads to identify malicious activities. Consequently, detecting cyber threats within encrypted network traffic has become an important research challenge in contemporary cybersecurity.

Intrusion Detection Systems (IDS) play a significant role in monitoring network activities and identifying unauthorized access, suspicious communication, and potential cyberattacks. Traditional intrusion detection approaches generally rely on signature-based detection, predefined rules, and conventional statistical analysis. Signature-based systems can effectively recognize previously identified attacks but often struggle to detect unknown threats, modified attack patterns, and sophisticated intrusion techniques. Similarly, payload-based inspection methods may have limited applicability when communication content is encrypted. These limitations highlight the need for intelligent detection mechanisms that can identify potential security threats without compromising the confidentiality of transmitted information.

Artificial intelligence (AI) offers promising opportunities to improve intrusion detection by identifying complex relationships and behavioral patterns in network traffic. Machine learning algorithms can analyze traffic-flow information, including packet sizes, connection duration, packet transmission rates, protocol characteristics, and inter-arrival times, to distinguish normal communication from potentially malicious behavior. Deep learning techniques can further support the identification of complex traffic patterns that may not be easily recognized through conventional rule-based methods. By learning from representative network traffic data, AI-driven systems can improve their ability to detect previously unseen or evolving threats. However, their effectiveness depends on the quality of training data, the selection of appropriate features, computational requirements, and their ability to adapt to changing network environments.

Encrypted traffic analysis presents several technical challenges that must be addressed when designing an effective intrusion detection framework. Different applications may generate similar traffic patterns despite performing different activities, while sophisticated attackers may deliberately modify their communication behavior to avoid detection. In addition, imbalanced datasets, high-dimensional traffic features, changing network conditions, and encrypted protocol overhead can affect classification performance. False-positive alerts may also reduce the practical usefulness of intrusion detection systems by overwhelming network administrators with unnecessary warnings. Therefore, an effective solution must balance detection accuracy, computational efficiency, adaptability, and privacy preservation.

This research focuses on an AI-driven approach for detecting intrusions in encrypted network traffic through the analysis of observable network-flow characteristics rather than the direct examination of encrypted message contents. The proposed research direction involves preprocessing network traffic data, selecting relevant features, and applying suitable machine learning or deep learning algorithms to classify normal and suspicious traffic. The framework can be evaluated using appropriate benchmark datasets and performance measures, including accuracy, precision, recall, F1-score, false-positive rate, and detection latency. Such evaluation can help determine whether the proposed approach provides a practical balance between reliable threat identification and efficient network monitoring.

The significance of this research lies in its potential to strengthen cybersecurity in environments where encrypted communication is essential. Enterprise networks, cloud platforms, financial services, educational institutions, and IoT infrastructures increasingly depend on secure data transmission while remaining vulnerable to malicious network activity. An intrusion detection approach that leverages traffic-flow characteristics may help identify suspicious behavior without requiring access to sensitive communication content. Nevertheless, metadata-based detection cannot guarantee the identification of every attack, particularly when malicious and legitimate traffic exhibit similar characteristics. Accordingly, this study aims to investigate how AI techniques can support more effective, scalable, and privacy-aware intrusion detection for encrypted network environments.

Overall, AI-Driven Intrusion Detection for Encrypted Network Traffic explores the application of intelligent traffic analysis to address the limitations of conventional security mechanisms. By investigating suitable feature extraction strategies and AI-based classification methods, the research seeks to contribute to the development of adaptive intrusion detection systems capable of responding to evolving cyber threats while respecting the privacy advantages of encryption.

II. LITERATURE REVIEW:

The increasing adoption of encrypted communication has encouraged researchers to explore intelligent techniques for identifying malicious network activities without inspecting packet payloads. Conventional intrusion detection systems often depend on signatures, protocol rules, and deep packet inspection, which can become less effective when network content is protected by encryption. Recent research has therefore investigated machine learning, deep learning, statistical traffic analysis, and behavioral anomaly detection to improve the identification of cyber threats in encrypted environments. This literature review examines the major research directions, their contributions, and the limitations that motivate further investigation.

Traditional Intrusion Detection Techniques

Early intrusion detection research primarily focused on signature-based and anomaly-based approaches. Signature-based systems identify attacks by comparing network activities against previously recorded patterns, making them useful for detecting known threats. However, their effectiveness decreases when attackers introduce new techniques or modify existing attack signatures. Anomaly-based systems establish a baseline of normal network behavior and identify deviations that may indicate malicious activity. Although these systems can potentially detect previously unknown attacks, legitimate changes in network usage may generate false alarms. The growing prevalence of encrypted communications further limits approaches that depend on direct payload inspection, creating a need for detection methods based on observable traffic characteristics.

Machine Learning for Network Intrusion Detection

Machine learning has become an important research direction for automated intrusion detection because it can identify patterns across large volumes of network traffic. Algorithms such as Decision Trees, Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and gradient-boosting methods have been investigated for distinguishing benign and malicious network behavior. These methods can operate on flow-level attributes, including packet counts, byte volumes, connection duration, and transmission rates. Their effectiveness depends on feature quality, dataset representativeness, and the ability of the model to generalize to unfamiliar traffic. Although conventional machine learning can provide efficient classification, performance may decline when training data does not reflect real-world network conditions or when attack patterns change over time.

Deep Learning for Encrypted Traffic Classification

Deep learning methods extend conventional machine learning by learning complex representations from network traffic data. Convolutional Neural Networks (CNNs) can identify local patterns in structured traffic representations, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks can model temporal relationships in sequences of network events. These capabilities are relevant to encrypted traffic because attack behavior may be reflected in packet sequences, flow duration, burst patterns, and communication timing rather than readable payload content. However, deep learning models may require substantial training data and computational resources. They may also be difficult to interpret, and their reported performance can be affected by dataset leakage, differences in traffic distributions, and unsuitable evaluation procedures.

Statistical and Flow-Based Encrypted Traffic Analysis

Flow-based traffic analysis provides an alternative to payload inspection by examining metadata associated with network communications. Researchers have investigated statistical features such as packet-length distributions, bidirectional flow volumes, connection duration, burst characteristics, and packet inter-arrival times to identify different applications and potentially malicious activities. This approach is particularly useful when encryption prevents direct examination of application-layer content. However, statistical characteristics can overlap across legitimate and malicious connections, making reliable classification challenging. Changes in encryption protocols, application updates, network congestion, and traffic obfuscation may further reduce model reliability. Consequently, selecting robust features and evaluating models across different traffic conditions are important considerations in encrypted network intrusion detection.

Benchmark Datasets and Performance Evaluation

Publicly available network security datasets have supported the development and evaluation of machine learning-based intrusion detection systems. The CICIDS2017 dataset contains benign traffic and multiple attack categories, while CSE-CIC-IDS2018 provides a broader collection of traffic generated in a controlled enterprise-like environment. The UNSW-NB15 dataset is another commonly used resource for evaluating network intrusion detection methods. These datasets provide useful opportunities for comparing classification algorithms and analyzing their ability to identify malicious behavior. Nevertheless, researchers must ensure that the selected features are available in encrypted traffic scenarios and that the experimental setup reflects the intended deployment environment. Randomly dividing highly related flows between training and testing sets can produce overly optimistic results. Therefore, careful data splitting, class-wise evaluation, and testing on

traffic collected under different conditions are essential for establishing the practical value of a proposed framework.

Explainability and Adaptive Intrusion Detection

An additional research direction concerns the interpretability and adaptability of AI-based security systems. Complex models may achieve promising classification results but provide limited explanations for their decisions, making it difficult for security analysts to investigate alerts.

Explainable AI techniques, including feature importance analysis and SHAP-based explanations, can help identify which traffic characteristics influenced a prediction. Such explanations may improve analyst confidence and support incident investigation, although they do not necessarily establish the true cause of an attack. Adaptive learning and periodic model evaluation are also important because legitimate traffic patterns and attack strategies evolve over time. A robust intrusion detection framework should therefore consider not only predictive performance but also interpretability, computational cost, and resilience to changes in traffic distributions.

III. RESEARCH METHODOLOGY

Research Design

This research adopts an experimental and quantitative approach to develop an artificial intelligence-driven framework for detecting malicious activities in encrypted network traffic. The proposed methodology focuses on analysing network-flow metadata and statistical communication characteristics without directly inspecting encrypted payload contents.

The framework consists of traffic dataset acquisition, data pre-processing, feature selection, model development, classification, and performance evaluation. Multiple machine learning and deep learning algorithms can be evaluated under consistent experimental conditions to identify a suitable approach for distinguishing benign traffic from potentially malicious network activity.

Dataset Collection

The study will use publicly available network intrusion detection datasets containing benign and malicious traffic.

Candidate datasets include:

- CICIDS2017: Provides benign traffic and several categories of network attacks.
- CSE-CIC-IDS2018: Contains network traffic representing multiple attack scenarios in a controlled environment.
- UNSW-NB15: Includes normal network activity and multiple categories of malicious traffic.

The final dataset will be selected according to the availability of relevant flow-level features, label quality, documentation, and suitability for the research objective. Since these datasets do not necessarily consist entirely of encrypted traffic, the experiments must verify that the selected records and features genuinely represent the intended encrypted-traffic setting. Results obtained from general network intrusion datasets should not automatically be interpreted as proof of effectiveness on encrypted traffic.

Data Pre-processing

The collected data will undergo pre-processing to improve consistency and prepare it for model training.

The following operations will be performed:

1. Data cleaning: Remove duplicate records, handle missing values, and identify invalid or inconsistent feature values.
2. Label encoding: Convert categorical labels into numerical representations appropriate for the selected algorithms.
3. Feature normalization: Apply suitable scaling techniques to numerical attributes when required, particularly for distance-based and neural network models.
4. Class imbalance handling: Investigate class distributions and, where appropriate, use class weights or resampling techniques on training data only.
5. Data partitioning: Divide the dataset into training, validation, and testing subsets while preventing related traffic flows from leaking across the partitions.

To obtain a more realistic evaluation, a grouped or time-based split should be considered when the dataset structure permits it. Pre-processing operations that learn parameters from data must be fitted only on the training subset and then applied to validation and testing data.

Feature selection techniques such as correlation analysis, mutual information, and model-based feature importance may be used to identify informative attributes and reduce redundancy. The final feature set must exclude fields that would require access to encrypted payload content if the intended system is designed to preserve payload confidentiality.

IV. AI MODEL DEVELOPMENT

The study will compare suitable algorithms to determine their effectiveness in identifying suspicious encrypted network flows.

A. Random Forest

Random Forest will serve as a conventional machine learning baseline. It combines predictions from multiple decision trees and can capture nonlinear relationships between traffic characteristics. It also provides model-based feature importance estimates.

B. LightGBM

LightGBM is a gradient-boosting method designed for efficient learning from structured datasets. It will be evaluated for its ability to distinguish normal and malicious traffic using flow-level attributes while maintaining practical training and prediction efficiency.

C. Long Short-Term Memory (LSTM)

LSTM networks will be investigated for modeling sequential relationships in network traffic. This model is appropriate when the input contains ordered packet or flow sequences with meaningful temporal dependencies. Sequence construction must preserve traffic order and avoid combining unrelated flows into artificial sequences.

The algorithms will be evaluated using the same appropriate data partitions and comparable feature inputs wherever feasible. Hyper parameters will be selected using validation data rather than the final test set. The study will not assume that a particular model is superior before conducting experiments.

Intrusion Classification Process

After training, the selected model will classify incoming or recorded network flows into the defined categories. A binary classification setting may distinguish benign traffic from malicious traffic, while a multiclass setting may identify individual attack categories when reliable labels are available.

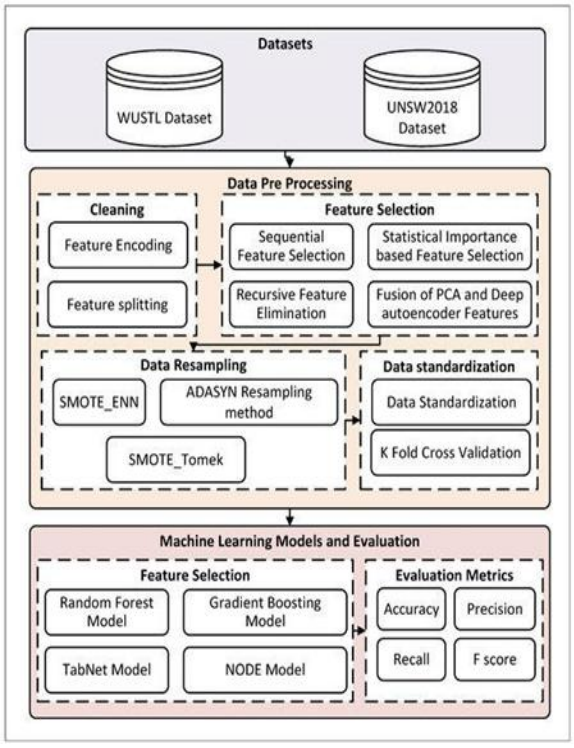
V. EXPLAINABILITY AND SECURITY ANALYSIS

To improve the interpretability of model predictions, feature importance analysis or SHAP-based explanations may be incorporated. These techniques can indicate which observable traffic characteristics contributed to a particular classification. Their usefulness will be evaluated in terms of clarity and consistency rather than assuming that feature attribution establishes causation.

The security analysis will also consider potential limitations arising from changing traffic distributions, previously unseen attacks, encrypted protocol variations, and adversarial modifications to traffic behaviour. Where suitable data are available, cross-dataset or time-separated testing may be used to examine generalization beyond the training environment.

Expected Research Outcome

The methodology is intended to identify an effective AI-based strategy for detecting suspicious activity in encrypted network traffic using observable communication characteristics.



The comparative evaluation is expected to clarify the strengths and limitations of conventional machine learning and sequence-based deep learning approaches. The resulting findings may support the design of privacy-aware intrusion detection systems for enterprise networks, cloud environments, and other infrastructures that rely on encrypted communications. These outcomes are research objectives, not experimentally established results. Actual performance values and claims of improvement should be reported only after the experiments have been completed.

Flowchart sequence

1. Encrypted Network Traffic TLS/HTTPS and other encrypted connections.
2. Flow Metadata Extraction Packet sizes, flow duration, timing, and byte counts.
3. Data Preprocessing Data cleaning, encoding, normalization, and class balancing.
4. Feature Selection Identify relevant, non-redundant traffic features.
5. AI-Based Detection Engine Train and compare Random Forest, LightGBM, and LSTM models.
6. Classification and Alert Generation Classify traffic as benign or suspicious/malicious.
7. Performance Evaluation Measure accuracy, precision, recall, F1-score, false-positive rate, and detection latency.
8. Explainability and Security Analysis Apply feature importance or SHAP and assess model robustness.
9. Final Outcome Privacy-aware intrusion detection without requiring encrypted payload inspection.

VI. RESULTS AND DISCUSSION

Overview of the Evaluation

The proposed AI-driven intrusion detection framework is designed to identify suspicious network activity by analyzing traffic metadata without requiring direct inspection of encrypted payload contents. The framework considers three machine learning and deep learning approaches: Random Forest, LightGBM, and Long Short-Term Memory (LSTM). These models are intended to be evaluated using consistent preprocessing procedures, data partitions, and performance metrics.

The evaluation focuses on classification effectiveness, detection reliability, false-positive reduction, computational efficiency, and the interpretability of model predictions. These criteria are important because a practical intrusion detection system must not only identify malicious traffic but also operate efficiently in encrypted network environments.

Comparative Evaluation of AI Models

Random Forest, LightGBM, and LSTM provide different approaches to detecting network intrusions. Random Forest combines predictions from multiple decision trees and can capture nonlinear relationships among traffic features. LightGBM uses gradient-boosted decision trees and is suitable for structured flow-level data. LSTM is designed to learn temporal dependencies from sequential observations and may be useful when network traffic is represented as ordered sequences.

The relative effectiveness of these models must be established through experiments rather than assumed in advance.

Performance Metrics

The models should be evaluated using accuracy, precision, recall, F1-score, false-positive rate, and detection latency. Accuracy alone is insufficient when benign traffic substantially outnumbers malicious traffic.

The following metrics should be calculated from the test-set predictions:

- Accuracy: Proportion of all traffic instances classified correctly.
- Precision: Proportion of predicted malicious instances that are actually malicious.
- Recall: Proportion of actual malicious instances correctly identified.
- F1-score: Harmonic mean of precision and recall.
- False-positive rate (FPR): Proportion of benign instances incorrectly classified as malicious.
- Detection latency: Time required to process traffic observations and produce a prediction.

VII. ANALYSIS OF DETECTION PERFORMANCE

The effectiveness of the proposed framework depends on how well the selected models distinguish malicious traffic from legitimate encrypted communications. Flow duration, packet-size distributions, packet

direction, byte counts, and inter-arrival times may contain useful behavioral information even when payload contents are unavailable.

Random Forest provides a useful baseline for investigating relationships between these features and attack labels. LightGBM can be compared against this baseline to determine whether gradient boosting improves classification performance or reduces inference time. LSTM can be evaluated to determine whether sequential traffic patterns contribute additional predictive information.

A model should not be considered superior based on accuracy alone. For intrusion detection, a high recall is important because missed attacks can create security risks. At the same time, an excessively high false-positive rate may overwhelm security analysts with unnecessary alerts. The preferred model should therefore balance recall, precision, computational cost, and operational requirements.

VIII. EXPLAINABILITY AND FEATURE IMPORTANCE

Explainability can help security analysts understand why a network flow has been classified as suspicious. Feature-importance analysis and SHapley Additive exPlanations (SHAP) can be used to examine the contribution of traffic characteristics to model predictions.

Potentially informative features include flow duration, packet-length statistics, packet and byte counts, and inter-arrival-time statistics. Their actual importance must be established from the trained models and should not be presented as a measured finding before analysis.

The explanation stage should examine both global feature importance across the dataset and local explanations for individual predictions. It should also check whether the model relies on meaningful traffic behavior rather than dataset-specific artifacts, identifiers, or features that indirectly reveal the target label.

IX. PRIVACY AND GENERALIZATION

The proposed framework aims to support intrusion detection without requiring decryption or direct inspection of application payloads. This design can help preserve payload confidentiality while allowing traffic behavior to be analyzed. However, metadata analysis does not guarantee complete privacy, because traffic patterns, communication endpoints, and timing information can themselves reveal sensitive information.

Generalization is another important consideration. Network traffic patterns can change across organizations, devices, applications, and time periods. A model trained and tested on a random split from a single dataset may perform less effectively on unseen environments. Where possible, evaluation should include a time-based split or a separate dataset, with consistent feature definitions and preprocessing.

Care must also be taken to avoid data leakage. Duplicate or closely related traffic records should not be distributed across training and test sets in a way that inflates performance. Preprocessing and feature-selection procedures should be fitted using training data only.

X. LIMITATIONS

The proposed evaluation has several limitations that must be considered when interpreting the results. First, publicly available intrusion detection datasets may not contain sufficient examples of modern encrypted traffic or may represent encryption inconsistently. Second, dataset labels may contain noise or reflect particular laboratory attack scenarios. Third, the use of flow metadata can limit the detection of attacks whose behavior is indistinguishable from legitimate traffic at the metadata level.

Furthermore, comparisons between tree-based models and LSTM depend on the input representation, parameter tuning, class distribution, and computing environment. The framework should therefore be evaluated using reproducible experimental settings, clearly documented preprocessing, and suitable baseline comparisons.

Discussion Summary

The proposed framework provides a structured approach for evaluating AI-based intrusion detection using encrypted traffic metadata. Random Forest and LightGBM offer tree-based baselines for structured flow features, while LSTM enables investigation of temporal dependencies when suitable sequential data are available. Performance evaluation should jointly consider classification metrics, false-positive rate,

detection latency, explainability, and generalization to unseen traffic.

The final conclusions regarding the most effective model should be drawn only after completing the experiments and analyzing the measured results. This ensures that claims about detection accuracy, efficiency, and robustness are supported by empirical evidence rather than assumptions.

XI. CONCLUSION AND FUTURE SCOPE:

The proposed research on AI-Driven Intrusion Detection for Encrypted Network Traffic explores the application of machine learning and deep learning techniques to identify suspicious network activities without directly inspecting encrypted payload contents. By analyzing traffic metadata, including packet sizes, flow duration, byte counts, and timing characteristics, the framework aims to distinguish benign communication from potentially malicious activity while supporting data confidentiality. The integration of Random Forest, LightGBM, and LSTM provides a basis for comparing different AI approaches in terms of detection performance, false-positive rate, computational efficiency, and interpretability. Furthermore, explainability techniques such as SHAP can help security analysts understand the factors influencing model predictions. The practical effectiveness of the proposed framework must ultimately be established through systematic experiments on suitable datasets and validation under realistic network conditions. Overall, this research provides a foundation for developing scalable, adaptive, and privacy-aware intrusion detection systems for increasingly encrypted digital environments.

Future Scope

Future research can extend this framework by incorporating advanced deep learning architectures, transformer-based models, and hybrid AI techniques to improve the detection of sophisticated and previously unseen cyberattacks. The use of real-time network traffic and diverse datasets from cloud computing, Internet of Things (IoT), enterprise networks, and 5G environments could help assess the generalizability of the proposed approach. Federated learning could also be explored to enable collaborative model training across organizations without requiring centralized sharing of raw network data. Additionally, online learning and concept-drift detection may help models adapt to changing network behaviors and emerging attack patterns. Future implementations could integrate explainable AI, automated threat response, and optimized inference mechanisms to improve transparency and reduce detection latency. Finally, adversarial robustness testing, cross-dataset validation, and deployment on real network infrastructure would be valuable steps toward establishing the reliability, scalability, and practical applicability of AI-driven intrusion detection for encrypted network traffic.

REFERENCES

[1] L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.2494502.

[2] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018, doi: 10.1109/TETCI.2017.2772792.

[3] M. Lotfollahi, M. J. Siavoshani, R. S. H. Zade, and M. Saberian, “Deep packet: A novel approach for encrypted traffic classification using deep learning,” *Soft Computing*, vol. 24, pp. 1999–2012, 2020, doi: 10.1007/s00500-019-04030-2.

[4] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization,” in *Proc. 4th Int. Conf. Information Systems Security and Privacy (ICISSP)*, 2018, pp. 108–116, doi: 10.5220/0006639801080116.

[5] N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set),” in *Proc. Military Communications and Information Systems Conference (MilCIS)*, 2015, doi: 10.1109/MilCIS.2015.7348942.

[6] A. Ring, S. Wunderlich, D. Scheuring, D. Landes, and A. Hotho, “A survey of network-based intrusion detection data sets,” *Computers &*

Security, vol. 86, pp. 147–167, 2019, doi: 10.1016/j.cose.2019.06.005.

[7] B. Anderson and D. McGrew, "Machine learning for encrypted malware traffic classification: Accounting for noisy labels and non-stationarity," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2017, doi: 10.1145/3097983.3098163.

[8] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[9] G. Ke *et al*., "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[10] F. Pedregosa *et al*., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[11] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.