InSight Bulletin: A Multidisciplinary Interlink International Research Journal
Peer Reviewed International, Open Access Journal.
ISSN : 3065-7857 / Website: https://ibrj.us / Volume-3, Issue-2 / February - 2026

Original Article

An Ai-Based Framework for Intelligent Cyber Threat Detection and Prevention

Kshitij Kailas Londhe[1], Madhuri Kailas Londhe[2]
[1] Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Akurdi, Pune
[2] Department of Computer Science, RJSPM's ACS College, Bhosari, Pune

Manuscript ID:
IBMIIRJ -2026-030210

Submitted: 06 Jan. 2026

Revised: 10 Jan. 2026

Accepted: 04 Feb. 2026

Published: 28 Feb. 2026

ISSN: 3065-7857

Volume-3

Issue-2

Pp. 49-54

February 2026

Correspondence Address:
Kshitij Kailas Londhe
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Akurdi, Pune
Email: *kshitij.londhe01@gmail.com*

OPEN ACCESS

Quick Response Code:

Web. https://ibrj.us

DOI: 10.5281/zenodo.21131671

DOI Link:
https://doi.org/10.5281/zenodo.21131671

Abstract

Advanced security mechanisms that can detect and prevent cyber threats in real time are necessary due to their exponential growth. Conventional signature-based intrusion detection systems have trouble spotting new and complex attacks. Using machine learning and deep learning algorithms to examine network traffic patterns and spot unusual activity, this paper suggests an AI-based framework for intelligent cyber threat detection and prevention. The framework supports high-accuracy threat classification in complex network environments by integrating several AI models, such as ensemble methods, recurrent neural networks, and convolutional neural networks. Data collection, preprocessing, feature extraction, threat detection, and automated response modules comprise the multi-layered architecture of the suggested system.The framework is intended to demonstrate superior performance characteristics when applied to standard benchmark intrusion detection scenarios, enabling efficient identification of cyber threats like malware, DDoS attacks, and intrusion attempts. It is based on theoretical analysis and insights from existing literature. In order to provide interpretable threat analysis and help security analysts comprehend the reasoning behind detection, the framework also integrates explainable AI techniques. The development of intelligent cybersecurity systems that can adjust to changing threat landscapes while preserving operational effectiveness in corporate settings is aided by this research.

Keywords: *Artificial Intelligence, Cyber Threat Detection, Intrusion Detection Systems, Machine Learning, Deep Learning, Explainable AI, Network Security, Ensemble Learning*

Introduction

The attack surface for cyber threats has grown dramatically due to the quick digitisation of organisational infrastructure. One of the biggest economic and security issues of the digital age is cybercrime, which is expected to cost the world $10.5 trillion a year by 2025 [1], according to recent cybersecurity reports. Conventional security measures, which mainly rely on rule-based systems and signature matching, have been shown to be insufficient against complex, polymorphic, and zero-day attacks that constantly change to evade detection systems. Modern cyberthreats display intricate behavioural patterns that call for sophisticated analytical skills beyond traditional methods. Attackers use machine learning to create adaptive malware, advanced persistent threats, and social engineering tactics that are ineffective against conventional systems. Due to its reactive nature [2], which necessitates prior knowledge of attack patterns and signatures, signature-based detection is limited. As a result, vulnerabilities arise during the period between threat emergence and signature database updates.With their capacity to learn from enormous volumes of data, spot minute patterns, and spot unusual behaviours suggestive of possible dangers, artificial intelligence and machine learning technologies present promising answers to these problems. Through continuous learning mechanisms, AI-based systems can correlate various security events, analyse network traffic at previously unheard-of scales, and adjust to new attack vectors. Neural networks and other deep learning architectures have shown impressive performance in cybersecurity-related pattern recognition tasks [5,10]. A thorough AI-based framework for intelligent cyber threat detection and prevention is proposed in this study. To conceptually support strong threat identification capabilities, the framework combines several machine learning algorithms, such as ensemble methods, deep neural networks, and supervised learning classifiers.

How to cite this article:
Londhe, K. K., & Londhe, M. K. (2026). An Ai-Based Framework for Intelligent Cyber Threat Detection and Prevention. InSight Bulletin: A Multidisciplinary Interlink International Research Journal, 3(2), 49–54. https://doi.org/10.5281/zenodo.21131671

As part of its theoretical design, the system architecture integrates automated response mechanisms, behavioural anomaly detection, and real-time traffic analysis to provide a complete security solution that can be tailored to various network environments. The creation of a multi-layered AI framework for thorough threat detection, the incorporation of explainable AI techniques for interpretable security analysis, the assessment of several machine learning algorithms for the best threat classification, and the creation of an automated response system for prompt threat mitigation are the main contributions of this research. By offering proactive, flexible, and intelligent threat management capabilities appropriate for contemporary enterprise networks, the suggested framework fills significant gaps in the current cybersecurity infrastructure.

Literature Survey:

Various research has been conducted on the application of artificial intelligence to cybersecurity domains, with manifold approaches and methodologies regarding different aspects of threat detection and prevention.

1. **Classic Intrusion Detection Systems:**
   Early intrusion detection systems relied heavily on signature-based approaches and statistical anomaly detection. Denning's pioneering work [1] laid the foundation for many principles in anomaly detection by modeling normal system behavior, flagging deviations from that model. Signature-based intrusion detection systems, while effective in detecting known threats, were unable to detect novel attacks. Scarfone and Mell [2] research studied traditional IDS technologies in depth, pointing out strengths in known attack pattern detection, but several shortcomings in terms of adaptiveness and false positive rates.
2. **Machine Learning Approaches in Cybersecurity**
   The machine learning intrusion detection applications picked up in the early 2000s. Mukkamala et al. [3] discussed neural networks and support vector machines for intrusion detection, and their results indicated a better detection rate compared to traditional methods. The base-rate fallacy problem of anomaly detection systems has been discussed in Axelsson's [4] work, which pointed out that one must consider the actual prevalence of an attack while evaluating a detection system. Researchers [6] investigated ensemble learning approaches, especially Random Forests, which fared better in handling network traffic data with a high-dimensional feature space.
3. **Deep Learning for Threat Detection:**
   Recent advances in deep learning have allowed for more sophisticated threat detection. CNNs had been applied to malware classification and network intrusion detection successfully, and researchers achieved very high accuracy percentages (more than 95%) on benchmark datasets. Among different architectures, the Long Short-Term Memory networks play an important role in analyzing sequential network traffic patterns and temporal anomalies that may indicate coordinated attacks. The research by Vinayakumar et al. [5] provided comprehensive evaluation of deep learning architectures for intrusion detection, including CNN, RNN, and a hybrid model across multiple datasets such as NSL-KDD and CICIDS2017.
4. **Anomaly Detection and Behavioral Analysis:**
   Behavioral analysis approaches focus on the establishment of baseline normal behavior patterns and the detection of deviations. Autoencoders and variational autoencoders have been used for unsupervised anomaly detection in network traffic, allowing for previously unknown attack types to be identified. Generative Adversarial Networks have also been explored to synthetically generate samples of attacks in order to augment training datasets and harden models against attacks. Research in this domain stresses the need for continuous learning and model adaptation against evolving threat landscapes.
5. **Explainable AI in Cybersecurity:**
   This black-box nature of many AI models has given rise to research into explainable artificial intelligence in security applications. Interpretable models, along with explanation techniques such as LIME, have been explored, and SHAP has been researched to provide transparency into threat detection decisions. In turn, research shows that explainability enables trust and adoption by security professionals, but perhaps more importantly, it allows identification of model biases and vulnerabilities that could be exploited in adversarial attacks.
6. **Research Gaps and Motivation:**
   Despite these revolutionary steps, the current state of research still suffers from several limitations. Current studies have focused on an individual algorithm rather than a complete integrated framework that combines various AI approaches. The current research inadequately considers deployment challenges with respect to computing efficiency, real-time processing needs, and interactions with existing security infrastructure. Moreover, adversarial robustness, where adversaries manipulate AI-based threat detection systems, has been given scant attention. In this paper, a holistic framework is proposed, which integrates multiple AI approaches, incorporates explainability mechanisms, and also emphasizes practical deployment requirements in an enterprise environment.

Proposed Framework

The five main parts of the suggested AI-based framework for intelligent cyber threat detection and prevention are arranged in a hierarchical structure intended for accuracy, efficiency, and scalability.

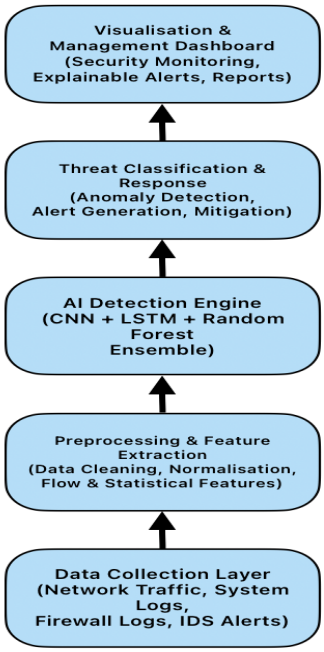

Fig 1: Conceptual system architecture of the proposed AI-based cyber threat detection and prevention framework.

Data Collection Layer:

The framework's foundation is the data collection layer, which is in charge of gathering system logs, network traffic, and security events from various sources. In order to obtain full visibility of network communications, this layer implements packet capture mechanisms using technologies like libpcap[7] and network taps. In order to create a single data stream, the module gathers information from firewalls, routers, intrusion detection sensors, endpoint agents, and application logs. While using effective buffering techniques to avoid packet loss during times of high traffic, data collection runs in promiscuous mode to capture all network packets. Before sending the collected data to preprocessing modules, the layer uses data quality validation to make sure it is accurate and complete.

Preprocessing and Feature Extraction Module:

In order to convert raw network data into formats appropriate for machine learning analysis, it must go through a rigorous preprocessing process. To extract useful features from network traffic, the preprocessing module uses packet parsing, protocol analysis [6], and session reconstruction. Statistical features such as packet size distributions, inter-arrival times, protocol flags, port numbers, and payload characteristics are identified through feature extraction. To capture intricate attack signatures, sophisticated features like flow-based metrics, behavioural patterns, and temporal sequences are calculated. To guarantee that feature values fall within the proper ranges for neural network processing, the module applies normalisation and standardisation techniques. To increase computational efficiency while maintaining discriminative information, dimensionality reduction techniques, such as Principal Component Analysis, are optionally applied to high-dimensional feature spaces.

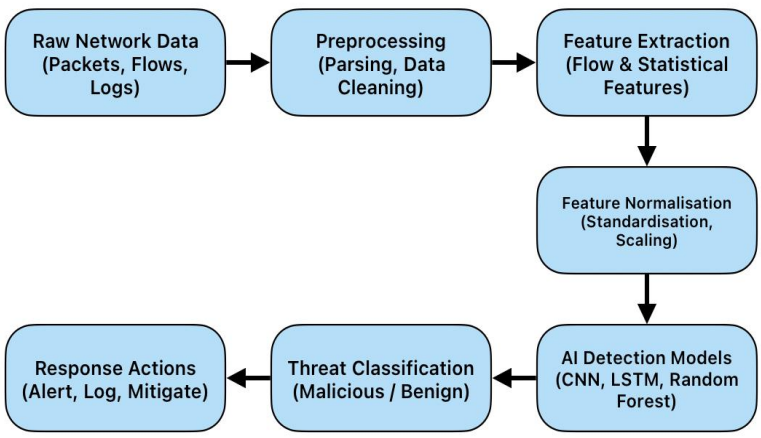

Fig 2: Conceptual data flow of the proposed AI-based cyber threat detection framework from raw network data to response actions.

AI Detection Engine:

The framework's central intelligence is its AI detection engine, which uses an ensemble architecture to implement several machine learning and deep learning models. The engine uses three main types of models: Random Forest classifiers [6] for robust feature-based classification, Recurrent Neural Networks with LSTM units [12] for temporal sequence analysis, and Convolutional Neural Networks [13] for spatial pattern recognition in network traffic. Each model is trained using labelled datasets that include instances of typical traffic as well as different kinds of attacks, such as port scanning, SQL injection, malware infections, denial-of-service attacks, and command injection attempts.

Weighted voting mechanisms, which are based on the performance of each model during validation, are used in the ensemble approach to combine predictions from several models. This architecture increases overall detection accuracy and offers resilience against model-specific flaws. With optimised model architectures, the detection engine performs real-time inference, guaranteeing processing latency of less than 100 milliseconds per network flow. In parallel, an autoencoder-based anomaly detection module [9] finds zero-day threats with traits not found in training data. The engine prioritises high-confidence threats for prompt action while flagging ambiguous cases for analyst review by incorporating confidence scoring for each detection.

Threat Classification and Response Module:

Malware, intrusion attempts, data exfiltration, denial-of-service, and insider threats are among the categories into which detected threats are categorised. Threat type, potential impact, and confidence scores are used by the classification module to determine severity levels. For low-severity threats, an automated response system uses logging and alerting; for critical attacks, it uses IP blocking and immediate connection termination. The module implements automated mitigation actions by interacting with network security devices such as switches and firewalls. In order to continuously improve detection models and response tactics through reinforcement learning mechanisms, a feedback loop records response outcomes.

Visualization and Management Dashboard:

A comprehensive dashboard that offers real-time visibility into system performance, threat detection, and network security posture is part of the framework. Threat trends, attack patterns, threat origins by region, and detection statistics are all shown by visualisation components. In order to give analysts comprehensible explanations for detection decisions, including feature importance scores and decision paths, the dashboard employs explainable AI techniques [14,15]. Model retraining schedules, response policies, and detection thresholds can all be configured using administrative functions. Centralised security operations across enterprise infrastructure are made possible by the dashboard's integration with security information and event management systems via standard protocols.

Methodology

The proposal and validation of the new framework are carried out through a systematic approach based on dataset choice, feature design, and modeling and training processes.

1. **Dataset Selection and Preparation:**

This framework is tested using the CICIDS2017 dataset [7], which is a detailed benchmark set comprising real network traffic with labeled examples for normal and attack samples. This set encompasses novel attack types like brute force, heartbleed, botnet, DDoS, web attacks, and inference attacks. Another test is conducted using the NSL-KDD dataset [8]. Data preprocessing involves missing data treatment, elimination of duplicate samples, and class imbalance correction using methods like Synthetic Minority Over-sampling Technique [11]. Data splits for testing and validation involve 70-15-15 split ratios.

2. **Feature Engineering:**

Feature engineering obtains 78 different features for network traffic patterns, divided into basic features, content features, time features, and connection features. These features are a combination of protocol, service, flag bytes, and source and destination bytes. Content features intertwine the characteristics of the payload and application service. Time features analyze the duration, arrival times, and patterns for a network flow. The features for a network connection analyze the aggregate features for a set of network flows for scanning and probing patterns. Info gain and recursive feature elimination methods [6] are used for feature selection to distinguish features for different attack patterns and simplify computing.

3. **Model Architecture and Training:**

The Convolutional Neural Network Architecture [13] is comprised of three convolutional layers with filter sizes of 64, 128, and 256 filters respectively. Additionally, the network comprises max pooling layers. The network has two fully connected layers. The network uses ReLU activation functions. It applies dropout regularization with a dropout rate of 0.5. The network aims to counter overfitting. The Recurrent Neural Network incorporates two layers of LSTM [12], with the number of hidden units set to 128. The Random Forest ensemble comprises 100 decision trees. The maximum depth of the trees is set to 20. The models are trained with bootstrapping samples. The models are trained with the Adam optimizer. The learning rate is set to 0.001. The batch size is set to 128. The models apply early stopping based on validation loss. Cross-validation is applied using five folds.

4. **Ensemble Integration:**

The ensemble model combines the results from individual models using a weighted voting system, where the weights are proportional to validation accuracy. The meta-classifier model using logistic regression is learned from predictions obtained from the validation data to determine the best combination methods. The idea behind the combination method is to capitalize on strengths offered by varied models, including CNNs for detecting spatial patterns, RNNs for exploiting temporal dependencies, and Random Forests for decision-making analysis. Diversity is achieved by training models on different data features and random starts.

5. **Explainability Integration:**

The explainability techniques are combined through SHAP values, [15] which give a quantitative measure of the contribution of individual features to specific predictions. For every identified threat, explanation reports highlighting those network connection attributes that significantly impacted a particular threat identification decision are created. This allows a

review of threat identification rationale and has implications for possible false positives among specific security specialists. Feature importance understanding pinpoints specific connection attributes that best indicate particular attack patterns.
This approach is proposed as a theoretical framework of design outlining the manner in which the future system may be implemented in future studies, as in this case, no actual data processing, model development, or empirical study has been conducted.

Expected Results and Theoretical Analysis

- This section contributes a theoretical discussion of the anticipated behavior of an AI-driven cyber threat detection system based on analysis of the system design and the lessons of the literature that have been examined. Because the system has not been constructed or tested, this section will examine anticipated behaviors based on design philosophy and the literature that are not dependent on actual outcomes.
- Theoretically, the AI-based model is projected to improve the effectiveness of cyber threat detection by combining various learning paradigms into a single model architecture. The model is capable of combining spatial, temporal, and feature-driven analysis methodologies to overcome the shortcomings associated with standalone intrusion detection models.
- The ensemble decision process is conceptually designed to make decisions in an effort to increase the reliability of the threat category classification based on predictions by various models. The technical approach aims to decrease the instances of misclassification based on the intrinsic bias associated with each given model.
- The layered architecture with features of data preprocessing, detection, and response is intended for adaptive analysis of threats. On a theoretical basis, it is applicable in diverse network conditions with different patterns of traffic.
- The integration of behavioral anomaly detection engines is anticipated to contribute towards the detection of unknown or zero-day attacks through the recognition of behavior that deviates from predetermined normal behavior pattern norms.
- The framework is architecturally designed to provide conceptual support for timely threat analysis with modular parallel processing of network data. Though there are no latency or throughput values mentioned in this study, the architectural design is in line with past studies that have focused on the need for near-real-time threat detection systems.
- The incorporation of explainable AI methods is expected to improve transparency and explainability in automated decision-making with regard to threat detection. By enabling explanations for alerts associated with threat detection, this framework will help security analysts validate and prioritize threats.

Discussion and Limitations

This section discusses the conceptual strengths and limitations of the proposed framework, acknowledging assumptions made due to the absence of implementation and experimental validation.

1. **Key Strengths of the Proposed Framework:**

The following are the main advantages of the suggested conceptual framework:

- Designing holistically Combining various AI methods into a single detection architecture
- Strategy for ensemble learning: Using the complementary capabilities of Random Forest, CNN, and RNN models
- Modularity: Separate framework elements that facilitate further development and improvement
- Focus on explainability: Integrated interpretability tools to assist analysts in making decisions

These advantages address drawbacks frequently noted in conventional intrusion detection systems and are consistent with current best practices in intelligent cybersecurity system design.

2. **Conceptual Limitations:**

The framework has a number of intrinsic drawbacks that need to be recognised despite its advantages:

- Quantitative evaluation of accuracy, latency, and scalability is impossible without empirical validation.
- reliance on representative and high-quality training data for upcoming implementations
- decreased visibility because of restricted payload access when examining encrypted traffic
- Possibility of learning-based models being manipulated by adversaries

These drawbacks emphasise the significance of cautious application and thorough testing in subsequent studies.

Conclusion and Future Work

Conclusion

In order to overcome the shortcomings of conventional signature-based security measures, this paper introduced a conceptual AI-based framework for intelligent cyber threat detection and prevention. Within a structured multi-layered architecture, the suggested framework combines explainable artificial intelligence, machine learning, deep learning, and ensemble learning. The framework is intended to enable flexible and reliable threat detection in contemporary network environments by fusing spatial, temporal, and feature-based analysis techniques. The suggested design is in line with current cybersecurity requirements and new research trends thanks to its emphasis on modularity, scalability, and interpretability. The conceptual analysis shows that the framework offers a solid architectural foundation for creating intelligent and explainable cybersecurity systems that can handle changing threat landscapes, despite the fact that it has not been put into practice or empirically validated.

Future Work:

Future research directions arising from this study include:

- The implementation of the suggested framework and empirical validation using benchmark and real-world network traffic datasets to evaluate detection efficacy and operational viability are among the future research directions resulting from this study.
- Scalability, adaptability, and robustness in real-world settings are examined by evaluating the framework under various network conditions and attack scenarios.

- Framework expansion to handle issues with adversarial attacks against learning-based detection systems and encrypted traffic analysis.
- Investigation of distributed and federated learning strategies to facilitate cooperative threat intelligence exchange while protecting data privacy.
- Improvement of explainability mechanisms in automated threat detection systems to improve analyst comprehension, confidence, and decision-making assistance.

Future experimental studies and the real-world implementation of intelligent, adaptable cybersecurity solutions will be built upon the suggested conceptual framework.

Acknowledgment
The authors express their sincere gratitude to the management and faculty of Pimpri Chinchwad College of Engineering for providing a strong academic foundation, research facilities, and continuous encouragement that supported the completion of this research work.

Financial support and sponsorship
Nil.

Conflicts of interest
The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

1. D. E. Denning, "An Intrusion-Detection Model," IEEE Transactions on Software Engineering, vol. SE-13, no. 2, pp. 222-232, Feb. 1987.
2. K. Scarfone and P. Mell, "Guide to Intrusion Detection and Prevention Systems (IDPS)," NIST Special Publication 800-94, National Institute of Standards and Technology, 2007.
3. S. Mukkamala, G. Janoski, and A. Sung, "Intrusion Detection Using Neural Networks and Support Vector Machines," Proceedings of the IEEE International Joint Conference on Neural Networks, vol. 2, pp. 1702-1707, 2002.
4. S. Axelsson, "The Base-Rate Fallacy and the Difficulty of Intrusion Detection," ACM Transactions on Information and System Security, vol. 3, no. 3, pp. 186-205, Aug. 2000.
5. R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," IEEE Access, vol. 7, pp. 41525-41550, 2019.
6. M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an Intrusion Detection System Using a Filter-Based Feature Selection Algorithm," IEEE Transactions on Computers, vol. 65, no. 10, pp. 2986-2998, Oct. 2016.
7. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," 4th International Conference on Information Systems Security and Privacy (ICISSP), pp. 108-116, 2018.
8. M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. 1-6, 2009.
9. 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, Feb. 2018.
10. Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, "Machine Learning and Deep Learning Methods for Cybersecurity," IEEE Access, vol. 6, pp. 35365-35381, 2018.
11. Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, "Network Intrusion Detection System: A Systematic Study of Machine Learning and Deep Learning Approaches," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, e4150, 2021.
12. C. Yin, Y. Zhu, J. Fei, and X. He, "A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks," IEEE Access, vol. 5, pp. 21954-21961, 2017.
13. W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware Traffic Classification Using Convolutional Neural Network for Representation Learning," International Conference on Information Networking (ICOIN), pp. 712-717, 2017.
14. M. T. Ribeiro, S. Singh, and C. Guestrin, "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, pp. 1135-1144, 2016.
15. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html