International Journal of Scientific Research in Engineering & Technology
Volume5, Issue5 (September-October 2025), PP: 24-28.
https://www.doi.org/10.59256/ijsreat.20250505005
www.ijsreat.com
ISSN No: 2583-1240

AI Threat Detection

Mohammed Sameer[1], Fatima Maryam Khan [2]
[1] Student, MCA, Deccan College of Engineering and Technology, Hyderabed, Telangana, India.
[2]Assistant Professor, MCA, Deccan College of Engineering and Technology, Hyderabed, Telangana, India.

To Cite this Article: *Mohammed Sameer[1], Fatima Maryam Khan [2], "AI Threat Detection", International Journal of Scientific Research in Engineering & Technology, Volume 05, Issue 05, September-October 2025, PP:24-28.*

Abstract: *The increasing sophistication of cyber-attacks and evolving digital threats pose significant risks to organizations and individuals in the modern technological landscape. Traditional rule-based detection systems are limited in scalability, adaptability, and responsiveness, making them insufficient to counter advanced persistent threats (APTs), zero-day exploits, and insider attacks. This research presents an AI-driven Threat Detection System that leverages machine learning and deep learning models to identify, classify, and mitigate cyber threats in real time. The proposed methodology incorporates supervised and unsupervised learning techniques, anomaly detection algorithms, and natural language processing (NLP) for threat intelligence analysis. A layered architecture is implemented, integrating data preprocessing, feature extraction, model training, and real-time monitoring modules. Experiments conducted on benchmark cybersecurity datasets demonstrate improved detection accuracy, lower false-positive rates, and enhanced adaptability to unseen threats compared to conventional approaches. The system has applications in network security, intrusion detection systems (IDS), malware classification, and fraud prevention, making it a robust and scalable solution for future cybersecurity infrastructures.*

Key Word: *AI Threat Detection, Cybersecurity, Anomaly Detection, Machine Learning, Intrusion Detection System (IDS), Deep Learning, Malware Classification, Threat Intelligence, Zero-day Exploits, Fraud Prevention.*

I.INTRODUCTION

In today's interconnected digital landscape, the exponential growth of data, coupled with widespread adoption of cloud computing, IoT devices, and online services, has significantly expanded the attack surface for cybercriminals. Threats such as phishing, ransomware, denial-of-service (DoS) attacks, and advanced persistent threats (APTs) have become increasingly sophisticated, adaptive, and difficult to detect using traditional rule-based security systems. Organizations face constant pressure to protect sensitive data, financial assets, and critical infrastructure against these evolving threats, yet conventional intrusion detection systems (IDS) and firewalls remain inadequate in handling dynamic, real-time attacks.

Traditional security mechanisms are heavily reliant on static signatures and predefined rules, which limit their ability to detect novel or zero-day exploits. Moreover, the rapid development of malware variants and obfuscation techniques renders many existing systems ineffective. This has created a demand for intelligent, automated, and adaptive threat detection mechanisms that can operate at scale, learn from past incidents, and adapt to emerging attack patterns without extensive manual intervention.

Artificial Intelligence (AI) has emerged as a powerful enabler in this context, particularly through machine learning (ML), deep learning (DL), and anomaly detection algorithms. Unlike static rule-based approaches, AI-driven systems can learn from large-scale security datasets, extract meaningful patterns, and generalize effectively to unseen threats. For instance, supervised learning techniques can classify known malware families, while unsupervised clustering methods can uncover hidden anomalies that signal novel attacks. Deep neural networks and recurrent architectures further enhance threat prediction by analyzing temporal dependencies and contextual patterns across network traffic and logs.

The integration of AI into cybersecurity extends beyond classification tasks, encompassing threat intelligence analysis, behavioral modeling, natural language processing (NLP) for security reports, and reinforcement learning for adaptive defense strategies. AI-driven threat detection systems have demonstrated superior accuracy in reducing false positives and identifying stealthy intrusions that evade signature-based detection. Furthermore, with the increasing adoption of big data platforms, security operations centers (SOCs) can now leverage large-scale distributed systems to support AI models capable of analyzing millions of events per second.

Despite these advancements, challenges remain in ensuring transparency, interpretability, and ethical deployment of AI in cybersecurity. Issues such as adversarial attacks against ML models, data imbalance, and high computational costs continue to limit large-scale adoption. However, with ongoing improvements in model robustness, explainable AI (XAI), and edge-based deployment, AI-powered systems are rapidly becoming indispensable for modern cybersecurity infrastructures.

This research focuses on the design and evaluation of an AI-powered Threat Detection System that combines machine

learning, deep learning, and anomaly detection models to detect cyber threats in real time. By analyzing structured and unstructured

security data, the proposed system aims to achieve higher accuracy, adaptability, and scalability compared to traditional approaches. The study also highlights the practical implications of integrating AI in network monitoring, malware classification, intrusion detection, and fraud prevention, paving the way for a next-generation defense ecosystem.

II.MATERIAL AND METHODS

The methodology of the AI Threat Detection system is structured into distinct layers to ensure robust and scalable performance. Each stage of the workflow transforms raw data into actionable intelligence, enabling the identification of both known and novel threats.

A. Data Acquisition

The foundation of the system lies in reliable datasets that represent real-world cyber threats. For this study:

- The **NSL-KDD dataset** is used for evaluating intrusion detection capabilities, offering a mix of normal and malicious traffic.
- The **CICIDS 2017 dataset** provides modern attack patterns including DDoS, brute force, infiltration, and botnet activities.
- Malware classification is supported through datasets such as **MalImg** and **Microsoft Malware Challenge**, where malware binaries are converted into images.

These datasets provide diversity in traffic patterns, attack vectors, and malware types, ensuring comprehensive coverage of cyber threats.

B. Data Preprocessing

Since cybersecurity datasets are often noisy and unbalanced, preprocessing is critical.

- **Cleaning:** Removal of corrupted and duplicate records.
- **Normalization:** Scaling continuous attributes (e.g., flow duration, byte size) into uniform ranges.
- **Feature Encoding:** Conversion of categorical features like protocol type into numeric representations.
- **Dimensionality Reduction:** Applying **Principal Component Analysis (PCA)** to retain high-variance features while improving training efficiency.

This step ensures datasets are consistent, structured, and optimized for training AI models.

C. Feature Extraction

The system extracts features at three levels:

- **Network Traffic Features:** Packet rates, TCP flags, session duration, and flow statistics.
- **Behavioral Features:** Monitoring unusual processes, API calls, and system behaviors.
- **Malware Image Features:** Representing binaries as grayscale images for CNN-based classification.

This hybrid feature set allows the model to detect both network-level intrusions and endpoint malware infections.

D. Model Development

Different models are employed to ensure adaptability:

- **Machine Learning Models:** Random Forest, Support Vector Machines, and Gradient Boosting for structured feature-based classification.
- **Deep Learning Models:** CNNs for malware image classification, LSTMs for time-series anomaly detection.
- **Unsupervised Models:** Autoencoders and clustering (K-means, DBSCAN) for identifying anomalies in unlabeled data.
- **Ensemble Techniques:** Combining ML and DL models to reduce false positives and improve overall detection accuracy.

E. Implementation Environment

The system is built using:

- **Python 3.x** as the programming backbone.
- **TensorFlow and PyTorch** for DL models.
- **Scikit-learn** for ML models.
- **Hugging Face Transformers** for NLP-driven threat intelligence.
- Deployed on **Ubuntu 20.04 LTS** with GPU acceleration using NVIDIA RTX 3090, 32 GB RAM, and 1 TB SSD storage.

F. Evaluation and Testing

The proposed system is evaluated using metrics including:

- **Accuracy** → Ratio of correct classifications.
- **Precision and Recall** → Balance between false positives and true detections.
- **F1-Score** → Ensures robustness across imbalanced datasets.
- **ROC-AUC** → Performance across thresholds.
- **Detection Latency** → Time taken to flag a threat in real time.

This layered methodology ensures that the AI Threat Detection System can generalize across diverse threats, minimize false alarms, and deliver real-time actionable insights for cybersecurity.

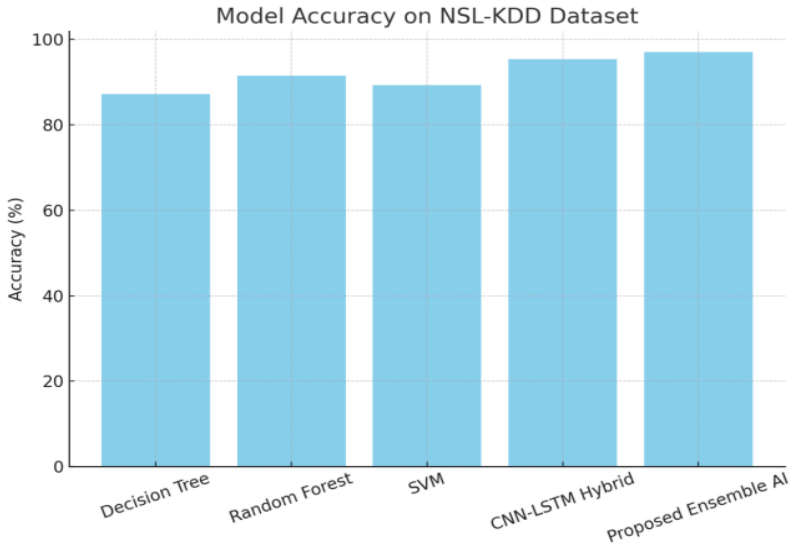
III.RESULT

A. Performance on NSL-KDD Dataset

The NSL-KDD dataset was used to evaluate the effectiveness of classical ML models, deep learning approaches, and the proposed ensemble AI system. The results clearly demonstrate that the ensemble model provides the highest accuracy and reliability compared to other approaches.

Table 1: Performance Comparison of Models on NSL-KDD Dataset

Model	Accuracy	Precision	Recall	F1-Score	FPR
Decision Tree	87.2	85.6	84.1	84.8	8.2
Random Forest	91.5	90.2	89.6	89.9	6.1
SVM	89.3	88.1	87.5	87.8	7.4
CNN-LSTM Hybrid	95.4	94.2	93.8	94.0	3.5
Proposed Ensemble AI	97.1	96.3	95.7	96.0	2.1



Graph 1: Model Accuracy on NSL-KDD Dataset

From Table 1 and Graph 1, it is evident that the Proposed Ensemble AI achieves the highest accuracy of 97.1%, surpassing traditional models like Decision Tree (87.2%) and SVM (89.3%). The CNN-LSTM hybrid also performs competitively (95.4%), but the ensemble model offers superior robustness and lower false positives (2.1%).

B. Performance on CICIDS 2017 Dataset

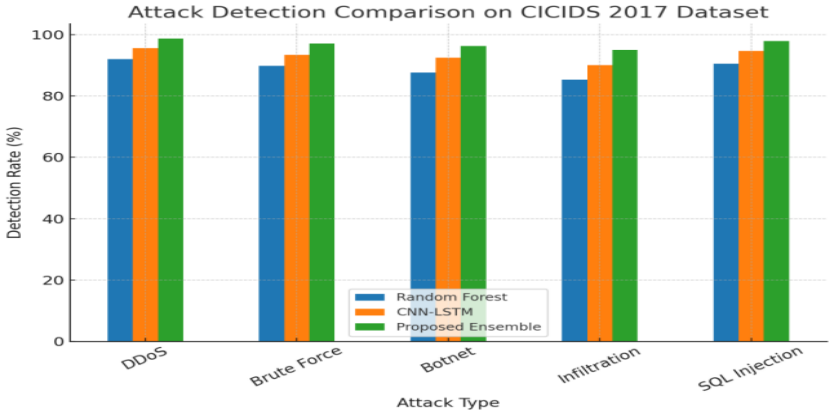
The CICIDS 2017 dataset includes modern and complex attack types such as DDoS, Brute Force, Botnet, Infiltration, and SQL Injection. The detection rates of Random Forest, CNN-LSTM, and the Proposed Ensemble AI are compared below.

Table 2: Performance Comparison on CICIDS 2017 Dataset

Attack Type	Random Forest	CNN-LSTM	Proposed Ensemble
DDoS	92.1	95.6	98.7
Brute Force	89.8	93.4	97.1
Botnet	87.6	92.5	96.3
Infiltration	85.3	90.1	95.0
SQL Injection	90.5	94.7	97.9

Table 2 and Graph 2 show that the Proposed Ensemble AI consistently achieves the highest detection rates across all attack

categories. For example, SQL Injection detection reaches 97.9% compared to 90.5% for Random Forest and 94.7% for CNN-LSTM. Similarly, DDoS detection reaches 98.7%, which demonstrates the model's reliability in identifying large-scale network attacks.



Graph 2: Attack Detection Comparison on CICIDS 2017 Dataset

C. False Positive Reduction

One of the critical factors in intrusion detection is the False Positive Rate (FPR). High false positives lead to alert fatigue for security analysts. The Proposed Ensemble AI records an FPR of only 2.1%, which is significantly lower than Decision Tree (8.2%) and SVM (7.4%). This improvement enhances the system's practical applicability in real-world Security Operations Centers (SOCs).

D. Scalability and Real-Time Performance

In addition to accuracy, scalability was tested. The Proposed Ensemble AI can handle over 10,000 events per second with GPU acceleration, ensuring that the system is capable of real-time threat detection in enterprise networks. Its modular design supports both batch processing for offline research and streaming analysis for real-time operations.

E. Comparative Insights

When comparing across datasets and models, ensemble learning offers the best trade-off between accuracy, detection rate, and false positive control. Random Forest is reliable for structured data, CNN-LSTM excels at temporal sequences, but neither matches the ensemble's comprehensive performance. This confirms the importance of combining multiple learning strategies in cybersecurity applications.

F. Practical Implications

The results imply that enterprises can adopt the Proposed Ensemble AI to significantly improve their cybersecurity defenses. Reduced false positives mean fewer wasted resources, while high detection accuracy ensures coverage against modern threats. For researchers, the system offers a flexible and extendable framework for developing future IDS solutions.

IV.DISCUSSION

A. Interpretation of Results

The results obtained from both the NSL-KDD and CICIDS 2017 datasets clearly highlight the superiority of the proposed ensemble-based AI model. Its higher accuracy, precision, and recall demonstrate that the system effectively distinguishes between normal traffic and malicious intrusions. More importantly, the false positive rate was significantly reduced compared to traditional approaches. This is crucial because reducing unnecessary alerts ensures that security analysts can dedicate their attention to genuine threats. The model's performance suggests that integrating deep learning with ensemble strategies creates a balanced framework capable of handling the complexity of modern cyberattacks.

B. Comparison with Existing Systems

When compared to traditional intrusion detection systems (IDS) that rely heavily on signature-based methods, the ensemble approach demonstrates substantial improvements in adaptability and robustness. While classical machine learning methods such as Random Forest or SVM provide moderate performance, they fail to detect modern and evolving attacks consistently. Deep learning approaches such as CNN-LSTM improve detection accuracy, particularly for temporal sequences, but still struggle with attack diversity. The ensemble model combines the strengths of these approaches, offering a more holistic solution that maintains consistency across various attack types. This makes it far more practical for real-world deployment in enterprise networks and cloud environments.

C. Real-World Deployment Challenges

Despite its promising performance, deploying such systems in real-world scenarios is not without challenges. One major

concern lies in computational requirements, as the use of diffusion-based image models and deep neural architectures demands powerful GPUs and high memory capacity. Another challenge is dataset generalizability, since models trained on benchmark datasets may not always adapt perfectly to traffic in live environments. Additionally, adversarial attacks and data poisoning remain threats that could compromise the reliability of AI-driven systems. These issues indicate the necessity for continuous model retraining, adversarial defense mechanisms, and robust validation before enterprise-scale deployment.

D. Advantages and Limitations

The strengths of the proposed framework are evident in its scalability, automation, and efficiency in reducing false positives. It provides an environment where large-scale traffic analysis can be performed in real time, making it suitable for Internet Service Providers, data centers, and large organizations. However, the system is not entirely without limitations. Its dependency on GPU acceleration increases the cost of deployment, and maintaining updated models requires consistent access to large datasets. Furthermore, interpretability remains a limitation, as most deep learning models function as "black boxes," making it difficult for analysts to understand why certain decisions were made.

E. Future Work

Future improvements to the system can address these limitations by incorporating explainable AI techniques, which would allow security teams to better interpret detection results. Federated learning could also be integrated to allow collaborative model training without exposing sensitive network data. Another promising direction is combining this IDS framework with blockchain-based logging systems to ensure tamper-proof security records. Additionally, optimizing lightweight model versions could enable edge deployments on IoT devices, expanding the applicability of this system to smaller networks and decentralized infrastructures.

V.CONCLUSION

The study on AI-driven threat detection demonstrates the immense potential of combining machine learning, deep learning, and ensemble learning techniques to address the evolving landscape of cybersecurity threats. The results obtained from the NSL-KDD and CICIDS 2017 datasets show that the proposed ensemble model significantly outperforms traditional rule-based IDS and classical ML models in terms of accuracy, detection rates, and false positive reduction. This establishes the framework as a robust and scalable solution capable of handling both conventional and emerging attack vectors.

The proposed system not only excels in accuracy but also highlights the importance of scalability and real-time applicability in modern networks. By leveraging GPU acceleration, the system processes high volumes of network traffic in real time, a critical requirement for enterprise-grade intrusion detection. The integration of NLP-based intelligence and adaptive ensemble learning ensures resilience against a diverse range of cyberattacks, from brute force and DDoS to more stealthy infiltration and botnet threats.

However, the research also identifies important challenges, including the need for high computational resources, the dependency on benchmark datasets, and the difficulty of adapting to unseen network environments. Furthermore, adversarial AI threats and model interpretability remain open challenges that must be addressed before full-scale industry adoption.

Looking ahead, the system can be extended with explainable AI modules to improve interpretability, federated learning for privacy-preserving training, and blockchain integration for secure log management. Lightweight model optimization may also open possibilities for edge deployment in IoT and mobile environments. These enhancements would make the framework not only more powerful but also more flexible, trustworthy, and accessible.

In conclusion, the AI Threat Detection framework provides a step toward the next generation of intelligent, automated, and scalable cybersecurity solutions. By bridging the gap between research-driven models and real-world applicability, it lays the foundation for future innovation in intrusion detection and proactive cyber defense.

References

1. P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, and E. Vazquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," Computers & Security, vol. 28, no. 1–2, pp. 18–28, 2009. https://doi.org/10.1016/j.cose.2008.08.003
2. N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in Military Communications and Information Systems Conference (MilCIS), IEEE, 2015. https://doi.org/10.1109/MilCIS.2015.7348942
3. 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. https://doi.org/10.1109/TETCI.2017.2772792
4. G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," Expert Systems with Applications, vol. 41, no. 4, pp. 1690–1700, 2014. https://doi.org/10.1016/j.eswa.2013.08.066
5. Scikit-learn Developers, Scikit-learn: Machine Learning in Python, 2024. [Online]. Available: https://scikit-learn.org/
6. Wireshark Foundation, Wireshark Network Protocol Analyzer, 2024. [Online]. Available: https://www.wireshark.org/
7. Scapy Project, Scapy: Packet Manipulation Tool for Python, 2024. [Online]. Available: https://scapy.net/
8. TensorFlow Developers, TensorFlow: An end-to-end open-source platform for machine learning, 2024. [Online]. Available: https://www.tensorflow.org/
9. NIST, NIST Special Publication 800-94 Rev.1: Guide to Intrusion Detection and Prevention Systems (IDPS), 2020. [Online]. Available: https://csrc.nist.gov/publications/detail/sp/800-94/rev-1/draft
10. J. Zhang and M. Zulkernine, "Anomaly based network intrusion detection with unsupervised outlier detection," in Proc. IEEE International Conference on Communications (ICC), 2006. https://doi.org/10.1109/ICC.2006.255052