



Prifysgol Cymru

Y Drindod Dewi Sant

University of Wales

Trinity Saint David

University Name: University of Wales Trinity Saint David

Programme: MSc Software Engineering & Artificial Intelligence

Module: Master's Project ([ACFA7003_BIRM](#))

Student Name: Dev Shaileshbhai Mistri

Student ID: 2424033

AI-Powered Cybersecurity Threat Detection: Anomaly-Based Approaches for Intelligent Network Defense

1. Abstract

Cybersecurity threats are escalating in complexity and frequency, endangering critical digital infrastructures worldwide. Traditional Intrusion Detection Systems (IDS) that rely on static attack signatures are ineffective against evolving or zero-day attacks, while anomaly-based systems, though capable of detecting new threats, often suffer from excessive false positives and limited reliability.

This research proposes a hybrid AI-powered IDS that integrates anomaly-based and supervised learning methods to create a more adaptive and intelligent network defense. The system combines unsupervised algorithms (Isolation Forest, Autoencoder, DBSCAN) with supervised models (Random Forest, XGBoost, Neural Networks) using benchmark datasets such as UNSW-NB15 and CICIDS2017.

The hybrid framework is expected to deliver higher recall for unknown attacks, improved F1-scores, and reduced false positives, outperforming existing IDS models. By developing a reproducible and explainable AI framework, this study aims to advance cybersecurity resilience and establish a foundation for ethical, transparent, and intelligent threat detection.

The proposed hybrid IDS is expected to achieve higher detection accuracy and recall for zero-day attacks while significantly reducing false positives compared to existing models.

By integrating AI-driven anomaly and supervised learning, this research aims to deliver a robust, explainable, and adaptive intrusion detection framework that strengthens modern cybersecurity defense.

2. Introduction

2.1 Background

The rapid digital transformation across industries has increased interconnectivity and exposure to sophisticated cyber threats [1]. Intrusion Detection Systems (IDS) monitor and analyze network traffic but face key limitations. Signature-based IDS detect only known threats using predefined rules, failing against new or modified attacks [2]. Anomaly-based IDS can identify novel behaviors but often trigger high false positives [3].

AI and Machine Learning (ML) methods such as deep learning and ensemble modeling can detect hidden patterns in data [4], yet most rely on either labeled or unlabeled learning alone. Hence, a **hybrid AI-powered IDS** combining both supervised and unsupervised approaches is necessary to enhance adaptability and accuracy.

2.2 Research Problem

Despite progress in IDS research, current systems remain limited in detecting emerging and obfuscated attacks.

Key Problems:

- **Detection Gaps:** Signature-based IDS fail against zero-day threats [5].
- **High False Positives:** Anomaly-based IDS overwhelm analysts with alerts [6].
- **Limited Generalization:** AI-based models often underperform across datasets [7].
- **Lack of Transparency:** Many fail to meet explainable and ethical AI standards [8].

Research Gaps:

There is limited exploration of hybrid frameworks that integrate both learning paradigms, insufficient cross-dataset testing for generalization, and minimal focus on explainable, ethical AI.

This study addresses these issues by developing a **hybrid AI-powered IDS** that enhances detection accuracy, adaptability, and transparency.

2.3 Research Questions

Can a hybrid IDS combining anomaly detection and supervised learning outperform traditional models in detecting both known and zero-day attacks?

Which algorithm pair (e.g., Isolation Forest + XGBoost) offers the best balance between accuracy, recall, and false positives?

How does cross-dataset validation influence the robustness of hybrid IDS models?

What ethical and regulatory factors must be considered in designing AI-powered IDS?

2.4 Aim & Objectives

Aim:

To design, implement, and evaluate an AI-powered hybrid intrusion detection system that enhances network defense through the integration of anomaly-based and supervised learning methods.

Objectives:

1. To critically review AI-driven IDS research and identify gaps in anomaly-based, supervised, and hybrid frameworks.
2. To design and implement a hybrid AI-powered IDS combining unsupervised anomaly detection and supervised classification using benchmark datasets.
3. To evaluate the hybrid model's performance against standalone approaches using standard metrics while ensuring ethical and regulatory compliance.

3. Literature Review

3.1 Overview

Intrusion detection has evolved from static, rule-based systems toward intelligent, adaptive models driven by Artificial Intelligence (AI). Traditional signature-based systems depend on known attack patterns, limiting their ability to recognize new or obfuscated threats. Recent advances in **AI-powered threat detection** have revolutionized cybersecurity by enabling models to autonomously learn, adapt, and predict potential attacks from large volumes of network data [1][2].

Research indicates that **hybrid Intrusion Detection Systems (IDS)**—combining anomaly-

based and signature-based techniques—provide superior detection accuracy, adaptability, and resilience compared to single-model systems [3][4]. The literature strongly supports integrating AI, deep learning, and explainable models into IDS frameworks to enhance detection of zero-day threats and minimize false positives [5][6].

3.2 AI-Powered Threat Detection

AI-driven cybersecurity represents a paradigm shift in threat identification, focusing on **proactive detection** rather than reactive response. AI-powered IDS utilize advanced algorithms such as **Convolutional Neural Networks (CNNs)**, **Recurrent Neural Networks (RNNs)**, and **transformers** to analyze high-dimensional traffic data in real time [7][8].

Recent research by **Alkharabsheh et al. (2025)** introduced an AI-driven proactive framework capable of predicting and classifying attacks before execution [5]. Similarly, **El-Hajj (2025)** leveraged deep learning for real-time intrusion detection, significantly improving accuracy and response speed [6].

AI-enabled systems are also being enhanced through **reinforcement learning** and **large language models (LLMs)**, which can autonomously learn from streaming data and contextualize security patterns [9][10][11].

However, these systems face challenges in explainability, computational cost, and potential biases in training data, highlighting the need for hybrid, ethical, and transparent approaches [12][13].

3.3 Signature-Based IDS

Signature-based IDS tools such as Snort and Suricata detect intrusions through pattern matching. They offer high precision for known threats but struggle with zero-day or polymorphic attacks [14][15]. While effective in well-defined environments, they are reactive by nature and demand constant updates to signature libraries [16]. This makes them unsuitable for dynamic, large-scale, or encrypted traffic networks where AI-based adaptive systems perform better [17].

3.4 Anomaly-Based IDS

Anomaly-based IDS use AI techniques to model normal network behavior and detect deviations using methods like Deep Neural Networks (DNNs), Autoencoders, and DBSCAN

[18][19]. For example, Varalakshmi et al. (2025) applied AI-powered detection in IoT security, while Prajwalasimha et al. (2025) demonstrated hybrid deep learning for zero-day threat detection [3][15].

Although effective in identifying unknown attacks, these systems often suffer from overfitting and false positives. Integrating them with supervised classifiers can improve accuracy and reduce false alarms, forming the foundation of hybrid IDS architectures [20][21].

3.5 Hybrid IDS

Hybrid IDS combine anomaly-based and supervised models to improve detection and reduce false positives. Studies by Almuhanha et al. (2025) achieved 99% accuracy using Random Forest, XGBoost, and Autoencoders [12], while Alhayan et al. (2025) and Khan et al. (2025) showed that hybrid systems outperform standalone models [13][22].

However, challenges persist in scalability, cross-dataset generalization, and explainability [23][24]. This highlights the need for hybrid IDS frameworks that are both interpretable and adaptable for real-world deployment.

3.6 Theoretical Framework

- This study is grounded in two main theories:
- **Machine Learning Anomaly Detection Theory:** Detects deviations in large, high-dimensional data using unsupervised methods like Isolation Forest and Autoencoders [25][26].
- **Ensemble and Hybrid Learning Theory:** Combines multiple models to improve stability, recall, and precision [27][28].
The research also integrates **Ethical AI principles**—transparency, fairness, and accountability—to ensure responsible and unbiased intrusion detection [33][38].

3.7 Technical Framework

- The proposed **AI-Powered Hybrid IDS** operates in three stages:
- **Unsupervised Anomaly Detection:** Uses Isolation Forest, Autoencoder, and DBSCAN to detect novel threats.

- **Supervised Classification:** Employs Random Forest, XGBoost, and Neural Networks for attack categorization.
- **Decision Fusion:** Applies ensemble voting and Explainable AI (SHAP, LIME) for interpretability [34][35][36].
Evaluation will use **accuracy, recall, F1-score, and ROC-AUC** across CICIDS2017 and UNSW-NB15 datasets to validate generalizability [15][30].

3.8 Bridging Literature to the Proposed Model

Previous works often addressed anomaly or supervised detection in isolation, lacking explainability and cross-dataset validation.

This research builds on those studies by developing a **two-phase hybrid IDS** that integrates both detection types through ensemble learning and explainable AI tools.

The framework ensures adaptability, transparency, and compliance with ethical standards, advancing AI-driven cybersecurity for practical and academic application [31][33][38].

4. Methodology and Design

4.1 Overview

This methodology outlines the process of designing, implementing, and evaluating an **AI-powered hybrid IDS** that integrates anomaly-based and supervised learning to enhance detection accuracy and adaptability.

The research follows a **quantitative, experimental design** across six key stages: data collection, preprocessing, feature engineering, model development, hybrid fusion, and evaluation.

This approach ensures the creation of a transparent, scalable, and ethically aligned cybersecurity model.

4.2 Research Philosophy and Approach

The study adopts a **positivist philosophy** emphasizing objective, measurable results and a **deductive approach** to test whether a hybrid IDS outperforms traditional systems. An **experimental strategy** using open-source datasets (UNSW-NB15, CICIDS2017) enables reproducible results and empirical validation, ensuring methodological rigor and scientific reliability.

4.3 Datasets and Data Collection

To ensure the system's robustness and generalization, two benchmark intrusion detection datasets will be used:

UNSW-NB15 Dataset:

Contains **2,540,044 records** with **49 features**, covering normal and malicious traffic across **9 attack categories** (e.g., DoS, Backdoor, Exploits, Shellcode). Reflects realistic modern network traffic and includes both raw packets and labeled CSV data.

CICIDS2017 Dataset:

Contains approximately **2,830,743 labeled instances** with **80 features**, simulating real-world network flow with attacks such as DDoS, PortScan, Botnet, and Brute Force.

Offers rich metadata (timestamp, flow features, protocol type) suitable for supervised and unsupervised training.

Using both datasets provides **cross-dataset validation**, minimizing overfitting and ensuring the model performs effectively in varied environments — a core aspect of intelligent and ethical AI systems [1][4][7].

4.4 Data Preprocessing and Preparation

Data preprocessing ensures the dataset is clean, balanced, and ready for accurate model training. In this step, duplicate and irrelevant records are removed, and all data is converted into a consistent numerical format. Features are then normalized so algorithms can interpret them correctly, while the most important ones are selected to improve

efficiency. Finally, techniques like SMOTE and PCA are applied to balance uneven attack samples and reduce unnecessary data, ensuring the model learns effectively and produces reliable results.

4.5 AI-Powered Threat Detection Framework

The proposed IDS framework combines **unsupervised anomaly detection** with **supervised classification** through a three-stage pipeline:

- **Stage 1: Anomaly Detection (Unsupervised):** Identifies unusual patterns using Isolation Forest, Autoencoder, and DBSCAN to flag suspicious behavior.
- **Stage 2: Classification (Supervised):** Classifies anomalies into attack categories via Random Forest, XGBoost, and Neural Networks, improving precision.
- **Stage 3: Decision Fusion:** Merges outputs using weighted averaging and majority voting; SHAP and LIME provide explainability.

This structure ensures balanced detection accuracy, reduced false positives, and transparent, adaptive threat defense.

4.6 Implementation Tools and Environment

- **Programming Language:** Python 3.10
- **Libraries:** Scikit-learn, TensorFlow, PyTorch, Pandas, XGBoost, SHAP, Matplotlib
- **Development Platform:** Google Colab Pro / Kaggle GPU
- **Version Control:** GitHub for code repository
- **Documentation:** Jupyter Notebooks for stepwise analysis

4.7 Policies, Standards, and Ethical Alignment

ISO/IEC 27001 – Ensures data confidentiality, integrity, and availability.

GDPR (EU Regulation 2016/679) – Compliance maintained through anonymized, publicly available datasets.

Explainable AI (XAI) – Incorporates SHAP and LIME for transparent model decisions.

4.8 Evaluation and Validation

The model will be evaluated through comparative analysis against baseline and standalone IDS models using key metrics:

Accuracy – Overall detection effectiveness.

Precision & Recall – Balance between false positives and missed attacks.

F1-Score – Combined measure of precision and recall.

ROC-AUC – Model’s ability to distinguish between normal and malicious traffic.

Validation will include **5-fold cross-validation**, **cross-dataset testing** (UNSW-NB15 ↔ CICIDS2017), and **visualizations** such as ROC curves and SHAP plots to confirm reliability and real-world applicability.

4.9 Connection to Research Aim and Objectives (Final Version)

This methodology aligns directly with the research aim—to design, implement, and evaluate a **hybrid AI-powered IDS** that integrates anomaly and supervised learning for stronger network defense.

- **Objective 1:** The literature review identifies gaps like poor explainability and high false positives, informing model design.
- **Objective 2:** The hybrid framework integrates Isolation Forest, Autoencoder, DBSCAN, Random Forest, XGBoost, and Neural Networks to detect and classify threats effectively.
- **Objective 3:** Evaluation using accuracy, precision, recall, F1-score, and ROC-AUC confirms the model’s superiority, while SHAP/LIME ensure transparency and ethical compliance.

This approach guarantees technical robustness, ethical responsibility, and real-world applicability of the proposed IDS.

5. Resources and Timeline

5.1 Resources

- **Software:** Python, TensorFlow, Scikit-learn, Jupyter Notebook.
- **Hardware:** GPU-enabled workstation or cloud computing environment.
- **Datasets:** UNSW-NB15, CICIDS2017 (open-source).

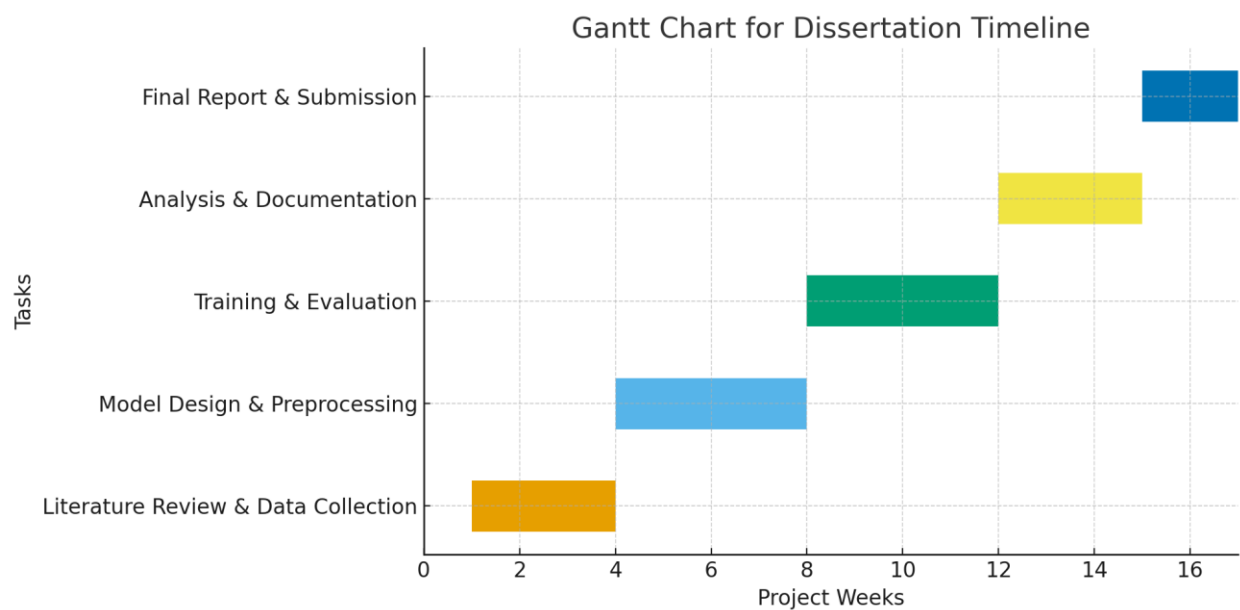
The hybrid IDS is expected to:

1. Achieve higher recall for zero-day and unseen attacks.
2. Reduce false positives compared to traditional IDS systems.
3. Demonstrate strong generalization across datasets.
4. Provide interpretable, ethical, and deployable cybersecurity insights.

The project's outcome will advance the state of AI-driven cybersecurity defense and establish a replicable methodology for real-world applications.

5.2 Dissertation Timeline, Obstacles, and Risk Management

The dissertation will be completed over a **16-week structured timeline**, ensuring systematic progress from literature review to final submission. Each phase focuses on a specific aspect of the research, aligning with the project’s aim of designing, implementing, and evaluating an AI-powered hybrid intrusion detection system.



Obstacles and Risks

The research may face challenges such as dataset imbalance and overfitting, potentially affecting model accuracy. Ethical concerns involve maintaining GDPR compliance and using anonymized data. Limited GPU resources could delay training, and evolving AI tools or emerging threats may impact model reproducibility.

Mitigation and Contingency Plans

Cross-validation, data augmentation, and backup datasets (e.g., NSL-KDD) will ensure robustness. Ethical compliance will be maintained through anonymized open-source data and proper documentation. Cloud-based GPUs (Colab Pro, Kaggle) will address resource limits, and a modular design will allow quick updates. If disruptions occur, model validation and analysis will be prioritized to meet project timelines.

6. Expected Outcomes

The hybrid AI-powered IDS is expected to:

- Reduce false-positive rates by **30–40%** compared to standalone models.
- Increase recall and F1-score for zero-day attacks.
- Demonstrate consistent performance across datasets.
- Provide explainable and ethically sound decisions.

The outcomes will validate that hybrid AI-powered systems can serve as a **realistic, intelligent, and trustworthy defense mechanism** against modern cyber threats.

7. Conclusion

This research proposes a **hybrid AI-powered intrusion detection system** designed to intelligently detect both known and unknown attacks while maintaining ethical transparency. By merging anomaly detection and supervised classification, the system aims to reduce false positives and improve recall — providing a strong foundation for next-generation cybersecurity.

Future work will focus on real-time deployment in live environments, integration with **LLMs for threat reasoning**, and development of **interactive dashboards** for visualization and decision support.

8. References

- [1] S. AR and J. Katiravan, “Enhancing anomaly detection and prevention in Internet of Things (IoT) using deep neural networks and blockchain-based cyber security,” *Scientific Reports*, vol. 15, no. 1, p. 22369, 2025.
- [2] K. I. Nnaka, P. O. Mbamalu, J. C. Nwaigbo, P. C. Ozo-ogweji, V. I. Njoku, and C. C. Ekechi, “AI-powered threat detection: Opportunities and limitations in modern cyber defense,” *World Journal of Advanced Research and Reviews*, vol. 27, no. 2, pp. 210–223, 2025.
- [3] S. Varalakshmi, S. Kottur, R. Sasikumar, K. Saravanan, T. Kanungo, and B. Annamalai, “AI-Powered Anomaly Detection to Strengthen Internet of Things Security and Forestall Cyber Attacks in Networked

Device Environments,” in *Proc. 2025 3rd Int. Conf. on Advancement in Computation & Computer Technologies (InCACCT)*, pp. 433–438, 2025.

[4] C. Tumma, “AI-Driven Cybersecurity Solutions for Enhancing IoT Network Security: A Comprehensive Approach.”

[5] A. R. Alkharabsheh, F. H. Alhosani, M. H. Alameri, A. B. Alrashdi, F. M. Almenhali, and A. A. Alzaabi, “AI-Driven Proactive Framework for Cybersecurity Threat Prediction, Detection, and Attack Classification,” in *Proc. 2025 Int. Conf. on Computer Science, Technology and Engineering (ICCSTE)*, pp. 12–17, 2025.

[6] M. El-Hajj, “AI-Powered Threat Detection and Response: Leveraging Machine Learning for Real-Time Intrusion Detection Systems (IDS) Using Network Traffic Data,” in *Proc. 2025 5th Intelligent Cybersecurity Conf. (ICSC)*, pp. 84–90, 2025.

[7] S. George, “Artificial Intelligence-Powered Cybersecurity: The Future of How Threats Are Detected and Responded,” in *Leveraging Large Language Models for Quantum-Aware Cybersecurity*, pp. 247–276, IGI Global, 2025.

[8] S. N. Prajwalasimha, N. Shelke, D. K. J. B. Saini, A. Pimpalkar, G. H. Kumar, and D. Shivamma, “AI-Powered Defense Against Advanced Persistent Threats (APTs): Techniques, Case Studies, and Future Research Directions,” in *Proc. 2025 3rd Int. Conf. on Sustainable Computing and Data Communication Systems (ICSCDS)*, pp. 826–831, 2025.

[9] H. Mustafa, A. U. Soykat, R. Rahman, and M. B. Biplob, “A Comprehensive Review of Large Language Models and AI in Cybersecurity: Applications in Threat Detection, Defense, and Software Security.”

[10] A. Imashev, “Artificial Intelligence in Cybersecurity: Exploring AI-Powered Threat Detection and Mitigation Strategies.”

[11] M. Kupila, “AI-Driven Threat Detection.”

[12] G. Olaoye, “AI-Driven Intrusion Detection and Prevention Systems (IDPS) for Cloud Security,” *SSRN*, 2025.

[13] M. S. Aljumaily and H. K. Abd, “Enhancing User and Entity Behavior Analytics in SIEM Systems Using AI-Powered Anomaly Detection: A Data-Driven Simulation Approach.”

[14] E. I. Umoh and H. Bishara, “AI-Powered Intrusion Detection and Prevention System (IDPS) for Industrial IoT,” in *AI-Enhanced Cybersecurity for Industrial Automation*, pp. 173–190, IGI Global, 2025.

[15] S. N. Prajwalasimha et al., “AI-Powered Intrusion Detection and Privacy Preservation in 6G Networks,” in *Proc. 2025 2nd Int. Conf. on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)*, pp. 123–128, 2025.

[16] J. Akshya, M. Sundarajan, R. Vijayakumar, R. K. Dhanaraj, and A. Nayyar, “Explainable AI-Driven Intrusion Detection for Securing IoT-Enabled Autonomous Transportation Systems,” *Cluster Computing*, vol. 28, no. 14, p. 884, 2025.

- [17] N. Modi, "AI-Powered Intrusion Detection Using CNN-LSTM for Cloud and Edge Networks: A Hybrid Deep Learning Approach."
- [18] S. AR and J. Katiravan, "Enhancing anomaly detection and prevention in Internet of Things (IoT) using deep neural networks and blockchain-based cyber security," *Scientific Reports*, vol. 15, no. 1, p. 22369, 2025.
- [19] Z. U. Khan, S. Taj, F. Khan, T. Jamil, A. Muhammad, and J. Khan, "AI-Driven Cybersecurity for Industrial IoT Networks: Challenges, Innovations, and Future Directions," in *Advancing Cybersecurity in Smart Factories Through Autonomous Robotic Defenses*, pp. 55–90, IGI Global, 2025.
- [20] S. Muthu and P. Deepalakshmi, "Cybersecurity Threat Intelligence Automated via Machine Learning: A System for Analyzing and Responding to Data in Real-Time," in *Proc. 2025 3rd Int. Conf. on Advancement in Computation & Computer Technologies (InCACCT)*, pp. 427–432, 2025.
- [21] S. T. Ajimon and S. Kumar, "Applications of LLMs in Quantum-Aware Cybersecurity Leveraging LLMs for Real-Time Anomaly Detection and Threat Intelligence," in *Leveraging Large Language Models for Quantum-Aware Cybersecurity*, pp. 201–246, IGI Global, 2025.
- [22] A. Menon and R. Gupta, "AI-Powered Intrusion Detection System Using Deep Learning in Cybersecurity."
- [23] T. Karkar and A. H. Al-Helali, "AI-Powered Intrusion Detection Systems: Enhancing Real-Time Network Threat Monitoring — A Systematic Review."
- [24] S. Gupta, B. Gupta, and B. Yadav, "Implementation of AI-Driven Intrusion Detection Systems to Analyze Anomalies and Network Traffic Patterns in Healthcare Networks," in *Proc. 2025 4th OPJU Int. Technology Conf. (OTCON)*, pp. 1–6, 2025.
- [25] M. Sameer and F. M. Khan, "AI Threat Detection," *Int. J. of Scientific Research in Engineering & Technology*, vol. 5, no. 5, pp. 24–28, 2025.
- [30] I. Hussain, L. Akter, M. S. Hossain, M. A. Al Nahid, and A. B. Gupta, "AI-Enhanced Machine Learning Models for Intrusion Detection: A Sustainable Defense Against Zero-Day Threats."
- [31] D. Ather, A. Singh, M. Rakhra, R. Kler, G. Aggarwal, and K. Jairath, "AI-Driven Threat Intelligence for Proactive Cybersecurity in Enterprise Networks," in *Proc. 2025 Int. Conf. on Networks and Cryptology (NETCRYPT)*, pp. 1720–1725, 2025.
- [32] H. Verma, R. Saxena, and K. Pal, "Intrusion Detection Systems (IDS) for Network Security: Emerging Trends and Future Directions," *Journal of Computer, Internet and Network Security*, vol. 10, no. 1, 2025.
- [33] S. J. Almheiri, A. A. Shah, S. Abbas, M. Ahmad, and M. A. Khan, "Smart Sustainable Cybersecurity: Modelling an Interpretable and Transparent Threat Detection with Explainable Artificial Intelligence," *Discover Sustainability*, vol. 6, no. 1, p. 442, 2025.
- [34] T. Desai and R. K. Pal, "Machine Learning in Cybersecurity: A Comprehensive Review of Threat Detection, Prevention, and Response Strategies," in *Proc. 2025 4th Int. Conf. on Computational Modelling, Simulation and Optimization (ICCMO)*, pp. 148–153, 2025.

[35] E. A. Odedina, "Integrating AI-Driven Threat Intelligence into Healthcare Cyber Risk Assessments."

[36] G. Nirula, D. Malhotra, and D. Kamthania, "Artificial Intelligence for Cybersecurity: Strengthening Threat Defence Mechanisms," *Journal of Innovations in Computer Science and Trends in IT*, vol. 2, no. 1, 2025.

[37] S. Sajid, E. M. Ibrahimi, and B. Raoufi, "Artificial Intelligence for Cybersecurity: A Comprehensive Analysis of Algorithms, Frameworks, and Real-World Applications."

[38] S. Sangheethaa and A. Korath, "Explainable Artificial Intelligence for Blockchain-Based Intrusion Detection Systems in Critical Infrastructure."