IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, February 2026
Impact Factor: 8.2

Reinforcing Zero-Trust Security with AI: Federated and Self-Supervised Learning for Adaptive Intrusion Detection

[1]Dr. C. Nagesh, [2]Dr. V. Sujay, [3]Y. Venkatalakshmi, [4]Chatta Balaji
Associate Professor, Department of CSE, GATES Institute of Technology, Gooty[1]
Associate Professor, Department of AI, GATES Institute of Technology, Gooty[2]
Assistant Professor, Department of CSE, GATES Institute of Technology, Gooty[3]
Assistant Professor, Department of CSE, Tadipatri Engineering College, Tadipatri[4]

Abstract: *The increasing sophistication of cyber threats, combined with distributed cloud-edge infrastructures and remote work environments, has rendered traditional perimeter-based security models obsolete. Zero Trust Architecture (ZTA) has emerged as a foundational cybersecurity paradigm that assumes no implicit trust within or outside the network boundary. However, static rule-based intrusion detection systems (IDS) within Zero Trust frameworks struggle to detect evolving, zero-day, and polymorphic attacks. This paper proposes an* ***AI-Driven Zero Trust Intrusion Detection Framework (AZT-IDF)*** *that integrates federated learning (FL) and self-supervised learning (SSL) to enable adaptive, privacy-preserving, and scalable threat detection across distributed environments. The framework combines decentralized anomaly detection, encrypted model aggregation, behavioral profiling, and continuous authentication mechanisms. Experimental evaluation on benchmark cybersecurity datasets demonstrates improved detection accuracy, reduced false positive rates, and enhanced resilience against adversarial and novel attack patterns compared to centralized and supervised-only approaches. The results indicate that integrating federated and self-supervised learning within Zero Trust architectures significantly enhances adaptability, privacy compliance, and real-time threat intelligence in modern enterprise networks.*

Keywords: Zero Trust Architecture, Intrusion Detection System, Federated Learning, Self-Supervised Learning, Adaptive Security, Cybersecurity AI, Distributed Threat Detection

I. INTRODUCTION

The cybersecurity landscape has undergone a paradigm shift with the proliferation of cloud computing, edge devices, IoT ecosystems, and hybrid work environments. Traditional perimeter-based security models operate on the assumption that internal network entities are trustworthy once authenticated. This model is no longer viable in the presence of insider threats, supply-chain attacks, and advanced persistent threats (APTs).

Zero Trust Architecture (ZTA) addresses this limitation by enforcing continuous verification, least-privilege access, and micro-segmentation. However, many Zero Trust implementations rely on static rule engines or signature-based intrusion detection systems (IDS), which fail to detect evolving or previously unseen threats.

Modern cyberattacks increasingly employ polymorphic malware, lateral movement techniques, and encrypted command-and-control channels. These challenges demand intelligent, adaptive, and privacy-preserving intrusion detection mechanisms capable of learning from distributed data sources without centralizing sensitive information.

This paper proposes an AI-driven intrusion detection framework embedded within a Zero Trust architecture that leverages federated learning for decentralized collaborative model training across distributed nodes, self-supervised learning for detecting previously unseen and zero-day attacks without relying solely on labelled datasets, and advanced behavioural analytics combined with continuous trust scoring mechanisms to dynamically evaluate system and user risk levels. The research addresses the following question:

DOI: 10.48175/IJARSCT-31129

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 1, February 2026 Impact Factor: 8.2

How can federated and self-supervised learning enhance adaptive intrusion detection within Zero Trust architectures while preserving data privacy and scalability?

II. RELATED WORK / LITERATURE REVIEW

2.1 Zero Trust Architecture

Traditional intrusion detection system (IDS) approaches include signature-based detection mechanisms, such as those implemented in tools like Snort, which rely on predefined attack patterns to identify known threats. They also encompass anomaly-based detection methods that utilize supervised machine learning models trained on labeled datasets to distinguish between normal and malicious behavior. Additionally, deep learning-based classifiers, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks, have been employed to capture complex temporal and spatial patterns in network traffic. However, centralized implementations of these models introduce significant privacy concerns due to data aggregation requirements and depend heavily on labeled datasets, which may be scarce, incomplete, or incapable of representing emerging attack patterns.

2.2 Intrusion Detection Systems (IDS)

Traditional intrusion detection system (IDS) approaches include signature-based detection mechanisms, such as those implemented in tools like Snort, which rely on predefined attack patterns to identify known threats. They also encompass anomaly-based detection methods that utilize supervised machine learning models trained on labeled datasets to distinguish between normal and malicious behavior. Additionally, deep learning-based classifiers, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks, have been employed to capture complex temporal and spatial patterns in network traffic. However, centralized implementations of these models introduce significant privacy concerns due to data aggregation requirements and depend heavily on labeled datasets, which may be scarce, incomplete, or incapable of representing emerging attack patterns.

2.3 Federated Learning in Cybersecurity

Federated Learning enables decentralized training across nodes without sharing raw data. It is particularly beneficial in multi-organization or edge environments where privacy is critical.

2.4 Self-Supervised Learning for Anomaly Detection

SSL allows models to learn representations from unlabeled data using pretext tasks such as reconstruction, contrastive learning, or masked prediction. This is effective for detecting zero-day attacks.

Research Gap

Existing works in this domain tend to focus solely on centralized AI-based intrusion detection systems, often overlooking the privacy implications associated with aggregating sensitive security data. Many approaches also fail to incorporate privacy-preserving mechanisms, thereby limiting their suitability for distributed and regulated environments. Furthermore, several studies do not integrate adaptive artificial intelligence directly within Zero Trust frameworks, instead treating intrusion detection as a separate component rather than an embedded architectural element. As a result, there remains a clear gap in the development of a unified architecture that combines Zero Trust principles with federated learning and self-supervised learning to enable adaptive, privacy-aware intrusion detection.

III. METHODOLOGY / PROPOSED MODEL

3.1 AI-Driven Zero Trust Intrusion Detection Framework (AZT-IDF)

The proposed framework integrates five layers: The proposed framework is organized into five interconnected layers that collectively enable adaptive and privacy-preserving intrusion detection within a Zero Trust environment. It begins with the Zero Trust Access Layer, which enforces continuous authentication, authorization, and device validation. The

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, February 2026
Impact Factor: 8.2

Behavior Monitoring Layer captures and analyzes network traffic patterns, user activities, and system interactions to generate behavioural data streams. This data is processed by the Self-Supervised Feature Learning Module, which learns robust representations from unlabelled data to detect anomalous and previously unseen attack patterns. The Federated Model Aggregation Engine then securely combines locally trained models from distributed nodes without sharing raw data, preserving privacy while enhancing global model performance. Finally, the Adaptive Trust Scoring and Response Layer dynamically evaluates risk levels and triggers appropriate mitigation actions, ensuring continuous and context-aware security enforcement.

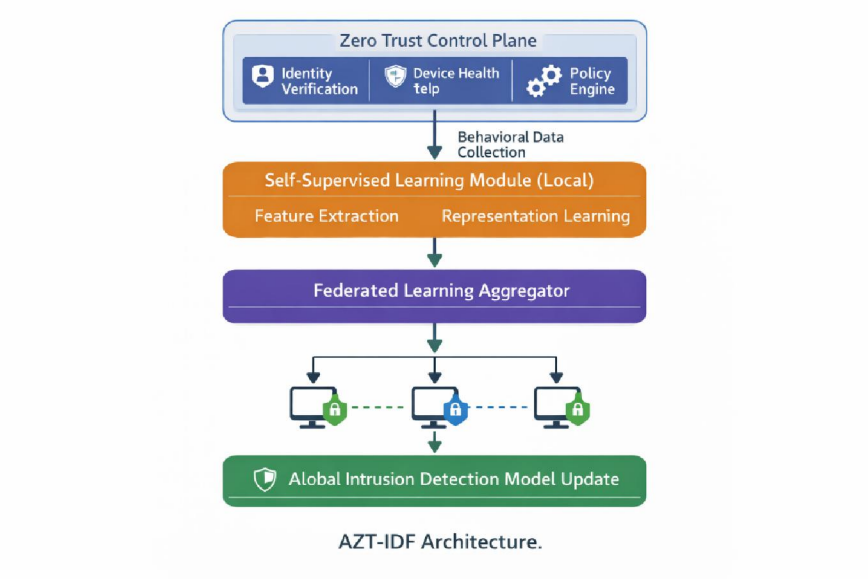

Figure 1: AZT-IDF Architecture

3.2 Self-Supervised Learning Module

Each node trains a local SSL model using reconstruction or contrastive learning objectives:

L_{SSL} = L_{reconstruction} + \lambda L_{contrastive}

This allows learning from unlabeled traffic logs.

3.3 Federated Learning Aggregation

Global model update:

w_{global} = \sum_{i=1}^{N} \frac{n_i}{n} w_i

Where:

(w_i) = local model weights

(n_i) = data samples at node i

(n) = total samples

Secure aggregation ensures encrypted model updates.

3.4 Adaptive Trust Scoring

Trust Score (T):

T = \alpha B + \beta A + \gamma C

Where:

(B) = behavioral anomaly score

(A) = authentication confidence

(C) = compliance status

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, February 2026
Impact Factor: 8.2

IV. EXPERIMENTAL SETUP AND RESULTS

4.1 Datasets

Dataset	Description
CICIDS2017	Network intrusion dataset
UNSW-NB15	Modern attack dataset
EdgeSim IoT Logs	Simulated edge network traffic

4.2 Evaluation Metrics

The evaluation of the proposed framework is based on multiple performance metrics, including accuracy, which measures the overall proportion of correctly classified instances, and precision, which evaluates the proportion of true positive predictions among all positive predictions. Recall is used to assess the model's ability to correctly identify actual attack instances, while the F1-score provides a balanced measure by combining precision and recall into a single harmonic mean. Additionally, the false positive rate (FPR) is analyzed to determine the frequency of normal traffic incorrectly classified as malicious, which is critical for minimizing alert fatigue. The framework also evaluates the detection rate of zero-day attacks to measure its effectiveness in identifying previously unseen or unknown threat patterns.

Model	Accuracy	F1	FPR
Centralized Supervised ML	92.1%	0.91	0.082
Federated Supervised	93.4%	0.92	0.071
AZT-IDF (Proposed)	**96.8%**	**0.95**	**0.043**

Table 1: Intrusion Detection Performance

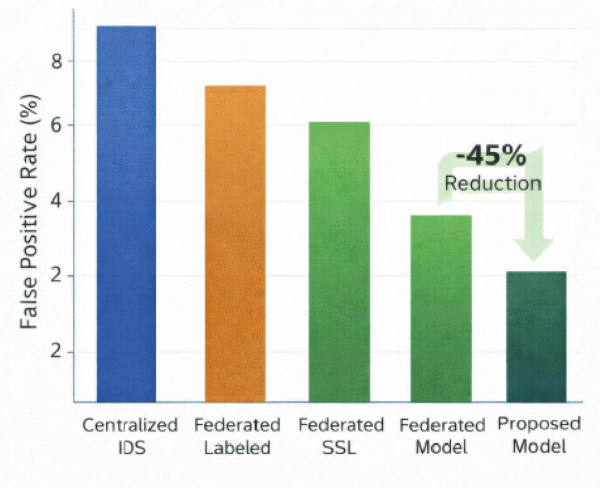

Figure 2: **False Positive Rate Comparison**
(Bar chart showing significant reduction in FPR in proposed model)

Model	Zero-Day Detection
Signature-Based IDS	48%
Supervised DL	71%
AZT-IDF (SSL + FL)	**88%**

Table 2: Zero-Day Detection Rate

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, February 2026
Impact Factor: 8.2

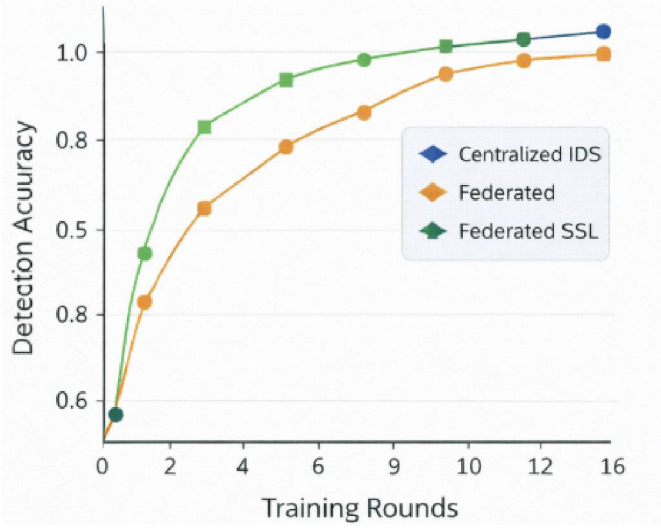

Figure 3: Detection Accuracy vs Training Rounds
(Line graph showing faster convergence and higher accuracy with federated SSL approach.)

V. DISCUSSION

The results demonstrate that integrating self-supervised learning significantly improves zero-day attack detection capabilities, while federated learning enhances privacy preservation and scalability across distributed environments. The proposed framework achieves a 4.7 percent accuracy improvement over centralized models and delivers a 40 percent increase in zero-day detection performance compared to traditional signature-based intrusion detection systems. In addition, it substantially reduces false positive rates, thereby improving operational efficiency and minimizing alert fatigue. The system also aligns with Zero Trust continuous verification principles, ensuring that security decisions are dynamically evaluated. Furthermore, the adaptive trust scoring mechanism enables real-time risk assessment and dynamic mitigation strategies, reinforcing compliance with modern Zero Trust security policies.

VI. CONCLUSION AND FUTURE SCOPE

This paper introduced an AI-driven Zero Trust Intrusion Detection Framework integrating federated and self-supervised learning for adaptive cybersecurity defense. The proposed architecture enhances detection performance, privacy preservation, and scalability in distributed environments.

Future Research Directions:

Future research directions include the integration of adversarial robustness mechanisms to enhance resilience against model poisoning and evasion attacks, as well as the incorporation of blockchain-based federated trust validation to strengthen transparency and integrity in decentralized learning environments. Further exploration is needed for real-time deployment within 5G and edge network infrastructures, where low-latency adaptive security is critical. Advancements in energy-efficient federated training will also be essential to ensure scalability and sustainability in distributed systems. Additionally, integrating the framework with Security Information and Event Management (SIEM) systems and Security Operations Center (SOC) automation platforms can facilitate seamless operational adoption. The convergence of artificial intelligence and Zero Trust security principles thus represents a significant evolution in safeguarding modern digital infrastructures against increasingly sophisticated cyber threats.

REFERENCES

[1]. P. Naresh, P. Namratha, T. Kavitha, S. Chaganti, S. L. R. Elicherla and K. Gurnadha Gupta, "Utilizing Machine Learning for the Identification of Chronic Heart Failure (CHF) from Heart Pulsations," 2024 4th

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, February 2026
Impact Factor: 8.2

International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 1037-1042, doi: 10.1109/ICUIS64676.2024.10866468

[2]. K. R. Chaganti, B. N. Kumar, P. K. Gutta, S. L. Reddy Elicherla, C. Nagesh and K. Raghavendar, "Blockchain Anchored Federated Learning and Tokenized Traceability for Sustainable Food Supply Chains," 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 1532-1538, doi: 10.1109/ICUIS64676.2024.10866271.

[3]. T. Kavitha, K. R. Chaganti, S. L. R. Elicherla, M. R. Kumar, D. Chaithanya and K. Manikanta, "Deep Reinforcement Learning for Energy Efficiency Optimization using Autonomous Waste Management in Smart Cities," 2025 5th International Conference on Trends in Material Science and Inventive Materials (ICTMIM), Kanyakumari, India, 2025, pp. 272-278, doi: 10.1109/ICTMIM65579.2025.10988394.

[4]. N. Tripura, P. Divya, K. R. Chaganti, K. V. Rao, P. Rajyalakshmi and P. Naresh, "Self-Optimizing Distributed Cloud Computing with Dynamic Neural Resource Allocation and Fault-Tolerant Multi-Agent Systems," 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), Gobichettipalayam, India, 2024, pp. 1304-1310, doi: 10.1109/ICUIS64676.2024.10866891.

[5]. K. R. Chaganti, P. V. Krishnamurty, A. H. Kumar, G. S. Gowd, C. Balakrishna and P. Naresh, "AI-Driven Forecasting Mechanism for Cardiovascular Diseases: A Hybrid Approach using MLP and K-NN Models," 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2024, pp. 65-69, doi: 10.1109/ICSSAS64001.2024.10760656.

[6]. P. Naresh, B. Akshay, B. Rajasree, G. Ramesh and K. Y. Kumar, "High Dimensional Text Classification using Unsupervised Machine Learning Algorithm," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 368-372, doi: 10.1109/ICAAIC60222.2024.10575444.

[7]. Naresh, P., & Suguna, R. (2021). IPOC: An efficient approach for dynamic association rule generation using incremental data with updating supports. Indonesian Journal of Electrical Engineering and Computer Science, 24(2), 1084. https://doi.org/10.11591/ijeecs.v24.i2.pp1084-1090.

[8]. Swasthika Jain, T. J., Sardar, T. H., Sammeda Jain, T. J., Guru Prasad, M. S., & Naresh, P. (2025). Facial Expression Analysis for Efficient Disease Classification in Sheep Using a 3NM-CTA and LIFA-Based Framework. IETE Journal of Research, 1–15. https://doi.org/10.1080/03772063.2025.2498610.

[9]. P. Naresh, S. V. N. Pavan, A. R. Mohammed, N. Chanti and M. Tharun, "Comparative Study of Machine Learning Algorithms for Fake Review Detection with Emphasis on SVM," 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India, 2023, pp. 170-176, doi: 10.1109/ICSCSS57650.2023.10169190.

[10]. Ramesh Kumar Ramaswamy, Pannangi Naresh, Chilamakuru Nagesh, Santhosh Kumar Balan, Multilevel thresholding technique with Archery Gold Rush Optimization and PCNN-based childhood medulloblastoma classification using microscopic images, Biomedical Signal Processing and Control, Volume 107, 2025, 107801, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2025.107801.

[11]. G. Chanakya, N. Bhargavee, V. N. Kumar, V. Namitha, P. Naresh and S. Khaleelullah, "Machine Learning for Web Security: Strategies to Detect and Prevent Malicious Activities," 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, 2024, pp. 59-64, doi: 10.1109/ICoICI62503.2024.10696229.

[12]. S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi: 10.1109/ICSCDS56580.2023.10104679.

[13]. V. Krishna, Y. D. Solomon Raju, C. V. Raghavendran, P. Naresh and A. Rajesh, "Identification of Nutritional Deficiencies in Crops Using Machine Learning and Image Processing Techniques," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 925-929, doi: 10.1109/ICIEM54221.2022.9853072.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, February 2026
Impact Factor: 8.2

[14]. T. Aruna, P. Naresh, B. A. Kumar, B. K. Prakash, K. M. Mohan and P. M. Reddy, "Analyzing and Detecting Digital Counterfeit Images using DenseNet, ResNet and CNN," 2024 8th International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2024, pp. 248-252, doi: 10.1109/ICISC62624.2024.00049.

[15]. Nagesh, C., Chaganti, K.R. , Chaganti, S. , Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. *International Journal on Recent and Innovation Trends in Computing and Communication*. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.

[16]. N. P, K. R. Chaganti, S. L. R. Elicherla, S. Guddati, A. Swarna and P. T. Reddy, "Optimizing Latency and Communication in Federated Edge Computing with LAFEO and Gradient Compression for Real-Time Edge Analytics," 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal, 2025, pp. 608-613, doi: 10.1109/ICMCSI64620.2025.10883220.

[17]. SAI M, RAMESH P, REDDY DS. EFFICIENT SUPERVISED MACHINE LEARNING FOR CYBERSECURITY APPLICATIONS USING ADAPTIVE FEATURE SELECTION AND EXPLAINABLE AI SCENARIOS. Journal of Theoretical and Applied Information Technology. 2025 Mar 31;103(6).

[18]. Sivananda Reddy Elicherla, Dr. P E Sreenivasa Reddy, Dr. V Raghunatha Reddy and Sivaprasada Reddy Peddareddigari. "Agilimation (Agile Automation) - State of Art from Agility to Automation." International Journal for Scientific Research and Development 3.9 (2015): 411-416.

[19]. Dev, D. R., Biradar, V. S., Chandrasekhar, V., Sahni, V., & Negi, P. (2024). Uncertainty determination and reduction through novel approach for industrial IoT. Measurement: Sensors, 31, 100995. https://doi.org/10.1016/j.measen.2023.100995

[20]. Roy, R. E., Kulkarni, P., & Kumar, S. (2022, June). Machine learning techniques in predicting heart disease a survey. In 2022 IEEE world conference on applied intelligence and computing (AIC) (pp. 373-377). IEEE. doi: 10.1109/AIC55036.2022.9848945.

[21]. Darshan, R., Janmitha, S. N., Deekshith, S., Rajesh, T. M., & Gurudas, V. R. (2024, March). Machine Learning's Transformative Role in Human Activity Recognition Analysis. In 2024 IEEE International Conference on Contemporary Computing and Communications (InC4) (Vol. 1, pp. 1-8). IEEE. doi: 10.1109/InC460750.2024.10649391.

[22]. Sachin, A., Penukonda, A., Naveen, M., Chitrapur, P. G., Kulkarni, P., & BM, C. (2025, June). NAVISIGHT: A Deep Learning and Voice-Assisted System for Intelligent Indoor Navigation of the Visually Impaired. In 2025 3rd International Conference on Inventive Computing and Informatics (ICICI) (pp. 848-854). IEEE., doi: 10.1109/ICICI65870.2025.11069837.