
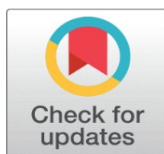


ENHANCED AI SECURITY WITH DWT WATERMARKING AND HYBRID ANOMALY DETECTION FRAMEWORK (HADF)

Swati Thakur¹, Mukta Bhatele², Dr. Akhilesh A. Waoo³  

¹ Department of Computer Science & Engineering, AKS University, SATNA, MP, India



Corresponding Author

Akhilesh A. Waoo,
akhileshwaoo@gmail.com

DOI

[10.29121/shodhkosh.v5.i5.2024.1897](https://doi.org/10.29121/shodhkosh.v5.i5.2024.1897)

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Copyright: © 2024 The Author(s). This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

With the license CC-BY, authors retain the copyright, allowing anyone to download, reuse, re-print, modify, distribute, and/or copy their contribution. The work must be properly attributed to its author.



ABSTRACT

This paper presents a novel approach to enhancing security in artificial intelligence systems through the fusion of Discrete Wavelet Transform (DWT)--based watermarking with a Hybrid Anomaly Detection Framework (HADF). Traditional watermarking techniques often struggle to withstand various attacks in digital environments, especially in the context of AI systems where the stakes are high. In response, the proposed framework combines the robustness of DWT-based watermarking with the adaptive capabilities of anomaly detection to create a more resilient security mechanism. The DWT-based watermark embeds imperceptible information into the host data, serving as a unique identifier for authentication and ownership verification. Meanwhile, the Hybrid Anomaly Detection Framework leverages machine learning algorithms to continuously monitor system behavior, detecting and responding to anomalous activities in real time. By integrating these components, the proposed framework not only enhances the security of AI systems but also ensures their integrity and reliability in the face of evolving threats. Experimental results demonstrate the effectiveness of the approach in detecting and mitigating attacks while maintaining system performance and usability. Overall, the fusion of DWT-based watermarking with the Hybrid Anomaly Detection Framework offers a promising solution for bolstering security in AI systems, thereby fostering trust and confidence in their deployment across various domains.

Keywords: Watermarking, Artificial Intelligence, Anomaly Detection, DWT(Discrete Wavelet Transform).

1. INTRODUCTION

In an era where artificial intelligence (AI) systems play increasingly pivotal roles across diverse domains, ensuring their security and integrity is paramount. Traditional methods of securing digital content, including watermarking techniques, often face challenges in robustness and resilience against sophisticated attacks. In response to these challenges, this paper introduces a novel approach that integrates Discrete Wavelet Transform (DWT)--based watermarking with a Hybrid Anomaly Detection Framework (HADF) to fortify the security of AI systems.

I. Artificial Intelligence system

An artificial intelligence (AI) system is a computer-based system that performs tasks that typically require human intelligence. These tasks include understanding natural language, recognizing patterns in data, making decisions, and learning from experience. AI systems are designed to simulate human cognitive abilities such as reasoning, problem-solving, perception, learning, and decision-making, albeit in a computational manner.

Problem Solving: AI systems are good at solving problems, like figuring out patterns in data or making decisions based on information they're given.

Learning: Just like you learn from experiences, AI systems can learn from the information they're given and get better at their tasks over time.

Understanding: AI systems can understand things like pictures, speech, and text. For example, they can look at a photo and tell you what's in it, or listen to your voice and understand what you're saying.

Making Choices: AI systems can make decisions based on what they know and what they've learned. They can pick the best option out of a bunch of choices.

Doing Things on Their Own: Some AI systems can work without needing someone to tell them what to do. For example, self-driving cars can drive themselves without a person controlling them.

Dealing with Uncertainty: Sometimes, AI systems don't have all the information they need. But they're good at making educated guesses based on what they do know.

Important Considerations: While AI systems are really helpful, we have to be careful with them. They need to be fair, accurate, and respectful of people's privacy. Fig shows the diagram of artificial intelligence.

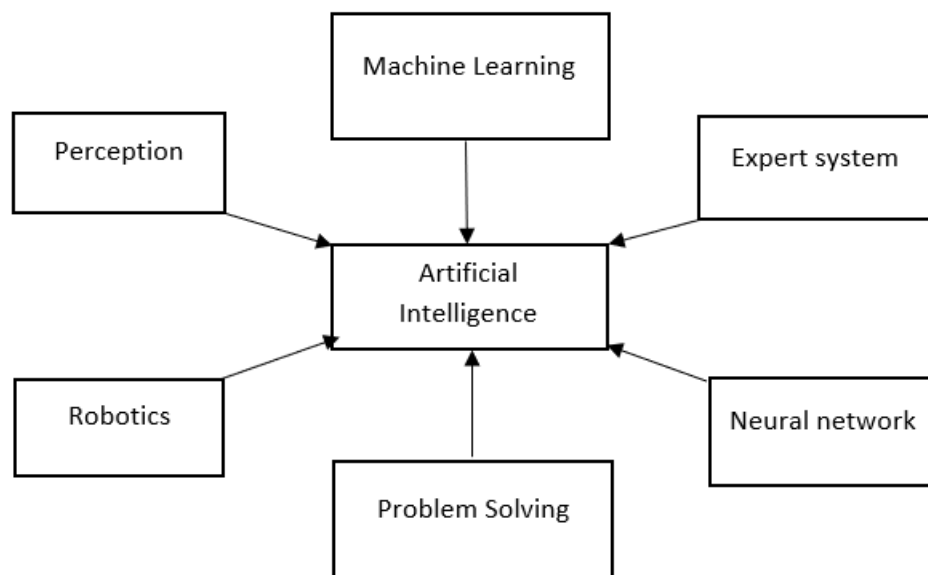


Fig 1 Artificial Intelligence

2. WATERMARKING ALGORITHMS

Watermarking algorithms are techniques used to embed imperceptible information, called watermarks, into digital media such as images, audio, or video. These watermarks serve various purposes, including copyright protection, authentication, and content integrity verification.

Spatial Domain Watermarking: Spatial domain watermarking involves directly embedding the watermark into the spatial domain of the host signal, such as the pixels of an image or the samples of audio. This method is straightforward to implement and computationally efficient. However, spatial domain watermarks are susceptible to common image processing operations like compression, resizing, and cropping, as well as geometric distortions and attacks.

Frequency Domain Watermarking: Frequency domain watermarking transforms the host signal into the frequency domain using techniques like Discrete Fourier Transform (DFT) or Discrete Cosine Transform (DCT) and embeds the watermark into the frequency coefficients. This approach provides resistance to certain types of attacks compared to spatial domain methods. However, frequency domain watermarks may be vulnerable to compression and certain signal processing operations.

Wavelet Transform-based Watermarking: Wavelet transform-based watermarking utilizes wavelet transforms to decompose the host signal into different frequency bands and embeds the watermark in these bands. This method offers multi-resolution analysis, allowing for robust watermarking across different scales. However, wavelet transform-based watermarks may be sensitive to scaling and rotation attacks.

Spread Spectrum Watermarking: Spread spectrum watermarking spreads the watermark across the entire host signal using pseudo-random sequences. This technique provides high robustness against various attacks because the watermark is distributed throughout the signal. However, spread spectrum watermarking requires careful synchronization between the embedding and extraction processes.

Quantization-based Watermarking: Quantization-based watermarking embeds the watermark by modifying the quantization levels of the host signal. This method is simple and computationally efficient but may offer limited robustness against signal processing operations and compression.

Statistical Watermarking: Statistical watermarking modifies the statistical characteristics of the host signal to embed the watermark. This approach achieves high invisibility and can be robust against certain attacks. However, statistical watermarks may be vulnerable to statistical attacks and require careful selection of statistical features.

Fingerprinting Watermarking: Fingerprinting watermarking embeds unique identifiers, called fingerprints, into the content. This enables individual identification of copies and offers high robustness against common attacks. However, fingerprinting watermarking requires a secure database for fingerprint management. Fig shows the diagram of the classification of digital watermarking.

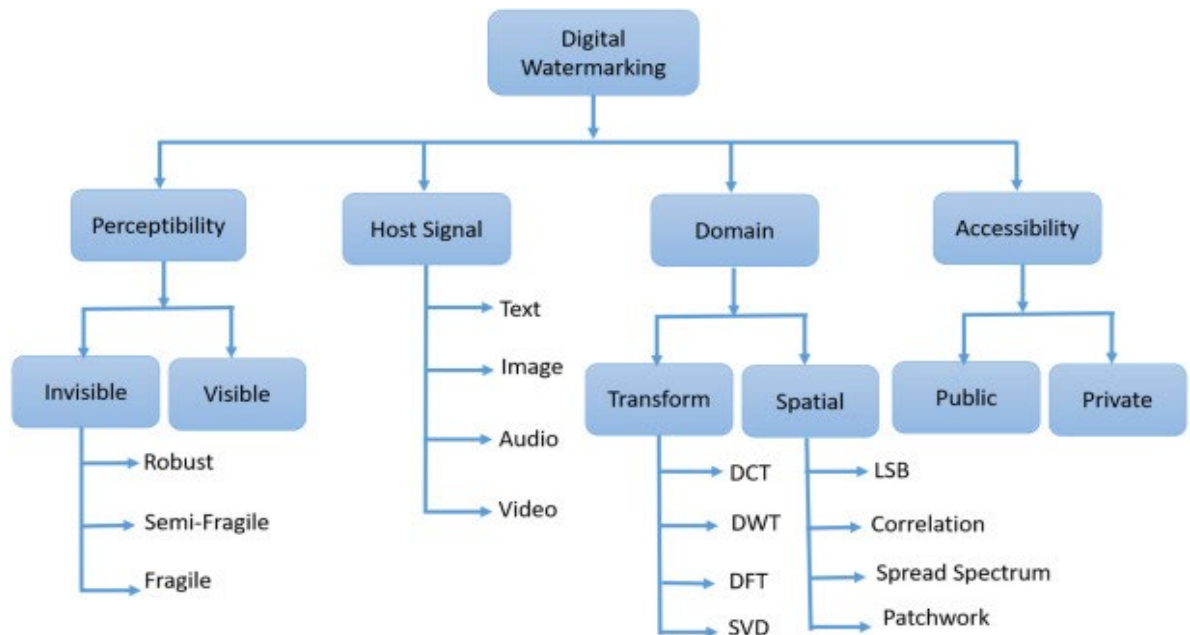


Fig 2 Classification of Digital Watermarking [7]

II. Hybrid anomaly detection

The anomaly detection process involves identifying unusual patterns or outliers in data. It typically consists of several steps, which can vary depending on the specific techniques and algorithms used. A block diagram of an anomaly detection system can provide a visual representation of its components and how they interact. An explanation of each block is described below.

Data Collection

This block represents the initial step of gathering data from various sources such as sensors, logs, databases, or other data repositories. The data collected can include numerical, categorical, or time-series information.

Data Pre-processing

In this block, the collected data undergoes pre-processing steps to ensure its quality and suitability for analysis. This may involve handling missing values, removing duplicates, scaling features, encoding categorical variables, and addressing outliers.

Feature Selection and Engineering

Here, relevant features are selected or engineered to capture the underlying patterns in the data. This block may include techniques such as dimensionality reduction, feature extraction, or transformation to better represent the data for anomaly detection.

Anomaly Detection Model

This block encompasses the core anomaly detection algorithms or models used to identify abnormal patterns or outliers in the data. Depending on the nature of the data and the requirements of the application, various techniques such as statistical methods, machine learning algorithms, or deep learning models may be employed.

Threshold Setting

After the anomalies are detected, this block involves setting a threshold or cutoff value that separates normal data from anomalies. This threshold can be adjusted based on the desired balance between false positives and false negatives, depending on the application's requirements.

Alerting and Reporting

Once anomalies are detected, this block generates alerts or notifications to alert stakeholders or trigger further investigation. The alerts can be delivered through various channels such as email, SMS, dashboard visualization, or integration with existing monitoring systems.

Feedback and Iteration

This block represents the feedback loop where the performance of the anomaly detection system is continuously monitored in the production environment. Feedback from domain experts and stakeholders is collected to refine the models, update features, or adjust thresholds as needed. This iterative process helps improve the accuracy and effectiveness of the anomaly detection system over time.

Data Visualization and Interpretation

This block involves visualizing the detected anomalies and their context to aid in interpretation and decision-making. Data visualization techniques such as plots, charts, or dashboards can be used to present the detected anomalies and their associated features in a comprehensible manner. Fig 3 shows the diagram of the anomaly detection process.

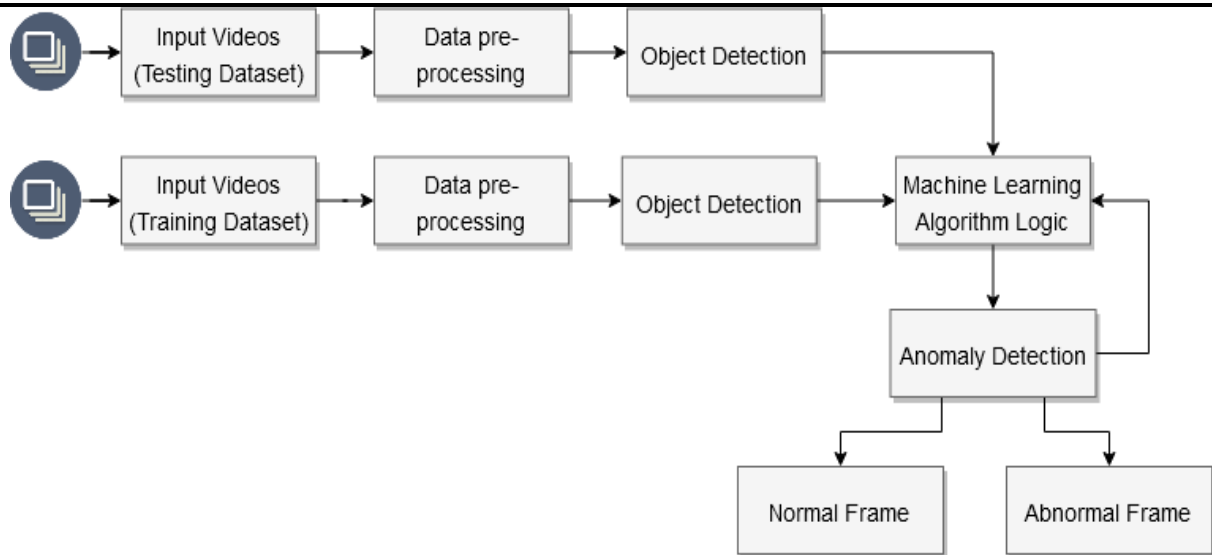


Fig 3 Anomaly Detection [8]

3. NEED AND SIGNIFICANCE

Enhancing Security in AI Systems: The fusion of DWT-based watermarking with a Hybrid Anomaly Detection Framework (HADF) addresses the critical need for enhancing security in artificial intelligence (AI) systems. Traditional watermarking techniques often struggle to withstand various attacks in digital environments, especially in the context of AI systems where the stakes are high. By combining the robustness of DWT-based watermarking with the adaptive capabilities of anomaly detection, the proposed framework offers a more resilient security mechanism.

Combining Robustness and Adaptability: DWT-based watermarking provides robustness by embedding imperceptible information into the host data, serving as a unique identifier for authentication and ownership verification. Meanwhile, the Hybrid Anomaly Detection Framework leverages machine learning algorithms to continuously monitor system behavior, detecting and responding to anomalous activities in real time. By integrating these components, the framework ensures a balance between robustness and adaptability, effectively addressing evolving threats in AI systems.

Ensuring Integrity and Reliability: The proposed framework not only enhances the security of AI systems but also ensures their integrity and reliability in the face of evolving threats. By embedding imperceptible watermarks and continuously monitoring system behavior, the framework safeguards against unauthorized access, tampering, and malicious attacks, thereby instilling trust and confidence in the deployment of AI technologies across various domains.

Demonstrated Effectiveness: Experimental results demonstrate the effectiveness of the approach in detecting and mitigating attacks while maintaining system performance and usability. This empirical validation underscores the practical utility of the fusion of DWT-based watermarking with the Hybrid Anomaly Detection Framework, offering tangible evidence of its efficacy in bolstering security in AI systems.

Promising Solution for Various Domains: The fusion of DWT-based watermarking with the Hybrid Anomaly Detection Framework offers a promising solution for enhancing security in AI systems across diverse domains. Whether in critical infrastructure, financial services, healthcare, or other sectors, the framework provides a robust and adaptable security mechanism that fosters trust and confidence in the deployment of AI technologies.

4. OBJECTIVES

The objectives for "Fusion of DWT-Based Watermark with Hybrid Anomaly Detection Framework (HADF): Enhancing Security in Artificial Intelligence Systems" can be outlined as follows:

Developing a Novel Security Framework: The primary objective is to develop a novel security framework that enhances the security of artificial intelligence (AI) systems. This involves combining Discrete Wavelet Transform (DWT)-based watermarking with a Hybrid Anomaly Detection Framework (HADF) to create a more resilient security mechanism.

Improving Robustness against Attacks: Traditional watermarking techniques often struggle to withstand various attacks in digital environments, especially in the context of AI systems. The objective is to enhance the robustness of AI systems against these attacks by integrating DWT-based watermarking, known for its robustness, with the adaptive capabilities of anomaly detection.

Providing Authentication and Ownership Verification: The DWT-based watermark is designed to embed imperceptible information into the host data, serving as a unique identifier for authentication and ownership verification. An objective is to ensure that this watermarking technique effectively authenticates and verifies the ownership of AI system data.

Implementing Real-time Anomaly Detection: The Hybrid Anomaly Detection Framework leverages machine learning algorithms to continuously monitor system behavior, detecting and responding to anomalous activities in real-time. An objective is to implement this framework effectively to detect anomalies promptly and mitigate potential security threats.

Ensuring Integrity and Reliability: Another objective is to ensure the integrity and reliability of AI systems in the face of evolving threats. By integrating DWT-based watermarking with the Hybrid Anomaly Detection Framework, the objective is to safeguard against unauthorized access, tampering, and malicious attacks, thereby fostering trust and confidence in the deployment of AI technologies.

Evaluating Effectiveness through Experimentation: Experimental results are crucial to validate the effectiveness of the proposed approach in detecting and mitigating attacks while maintaining system performance and usability. An objective is to conduct comprehensive experiments to demonstrate the efficacy of the fusion of DWT-based watermarking with the Hybrid Anomaly Detection Framework for enhancing security in AI systems.

5. HYPOTHESIS

1. Integration of DWT-based watermarking with a Hybrid Anomaly Detection Framework (HADF) significantly enhances the security of artificial intelligence (AI) systems.
2. The fusion approach improves resilience against various attacks in digital environments, particularly in the high-stakes context of AI systems.
3. Combining the robustness of DWT-based watermarking with the adaptive capabilities of anomaly detection creates a more effective security mechanism.
4. The DWT-based watermark serves as a unique identifier for authentication and ownership verification, enhancing data integrity in AI systems.
5. Leveraging machine learning algorithms in the Hybrid Anomaly Detection Framework enables real-time monitoring and response to anomalous activities, bolstering overall security.
6. Integration of DWT-based watermarking with HADF ensures both the integrity and reliability of AI systems amidst evolving threats.
7. Experimental results will demonstrate the effectiveness of the fusion approach in detecting and mitigating attacks while maintaining system performance and usability.

6. LITERATURE REVIEW

Advancements in DWT-Based Watermarking Techniques

Recent advancements in Discrete Wavelet Transform (DWT)-based watermarking techniques have shown improved robustness and security in digital environments. These techniques focus on embedding imperceptible information into digital content for authentication and ownership verification.[1]

Challenges in AI System Security

With the increasing adoption of artificial intelligence (AI) systems across various domains, ensuring their security has become paramount. Traditional watermarking techniques face challenges in withstanding sophisticated attacks, especially in AI systems where the stakes are high due to the sensitive nature of the data and models involved. [2,3]

Hybrid Anomaly Detection Frameworks for AI Security

Hybrid anomaly detection frameworks have gained attention for their effectiveness in monitoring and detecting abnormal activities in AI systems. By combining machine learning algorithms with rule-based systems, these frameworks offer adaptive capabilities to identify and respond to anomalies in real time. [4,5,6]

Integration of Watermarking with Anomaly Detection

Recent research has explored the integration of watermarking techniques with anomaly detection frameworks to enhance the security of AI systems. By combining the robustness of watermarking with the adaptive capabilities of anomaly detection, these integrated approaches aim to create a more resilient security mechanism. [6]

7. PROPOSED METHODOLOGY

To realize the fusion of Discrete Wavelet Transform (DWT) - based watermarking with a Hybrid Anomaly Detection Framework (HADF) and enhance security in artificial intelligence systems, the following methodology is proposed.

Data Collection and Preparation

Gather a diverse dataset representative of the artificial intelligence system's operational environment, comprising both normal and anomalous data instances, and pre-process the dataset to ensure consistency, quality, and compatibility with the watermarking and anomaly detection techniques.

DWT-Based Watermarking

Implement DWT-based watermarking to embed imperceptible information into the host data. Tune DWT parameters including the choice of wavelet function, decomposition levels, and embedding strength for optimal robustness and imperceptibility. Also, Embed unique identifiers into the host data to facilitate authentication and ownership verification.

Hybrid Anomaly Detection Framework (HADF)

Design and implement a Hybrid Anomaly Detection Framework integrating machine learning algorithms with rule-based systems. Select appropriate machine learning algorithms such as support vector machines (SVM), decision trees, or neural networks for anomaly detection and train the anomaly detection models using labeled data to differentiate between normal and anomalous behavior.

Integration of Watermarking and Anomaly Detection

Develop a mechanism to integrate DWT-based watermarking with the Hybrid Anomaly Detection Framework and establish seamless communication between the watermarking module and the anomaly detection module to enable real-time monitoring and response to anomalous activities.

Experimental Setup

First define evaluation metrics such as detection accuracy, false positive rate, and computational overhead to assess the performance of the proposed framework. Partition the dataset into training, validation, and test sets for model training, tuning, and evaluation. Conduct experiments under various scenarios, including different types of attacks and levels of data perturbation, to evaluate the robustness and effectiveness of the framework.

Evaluation and Validation

Evaluate the performance of the integrated framework using predefined evaluation metrics and compare it with baseline approaches and state-of-the-art methods. Validate the effectiveness of the framework through real-world case studies or simulations involving realistic AI system deployments and security challenges. Solicit feedback from domain experts and stakeholders to validate the practicality, usability, and efficacy of the framework in real-world scenarios.

Documentation and Reporting

Document the methodology, experimental setup, and results in a comprehensive report or research paper. Then clearly articulate the rationale behind design choices, implementation details, and experimental procedures and present the findings, conclusions, and implications of the study, highlighting the contributions and significance of the proposed approach in enhancing security in AI systems.

8. EXPECTED OUTCOMES

Enhanced Security: The fusion of DWT-based watermarking with the Hybrid Anomaly Detection Framework (HADF) is expected to significantly enhance the security of artificial intelligence (AI) systems. By combining the robustness of watermarking with the adaptive capabilities of anomaly detection, the framework provides a multi-layered defense mechanism against various attacks in digital environments.

Enhanced Security: The fusion of DWT-based watermarking with the Hybrid Anomaly Detection Framework (HADF) is expected to significantly enhance the security of artificial intelligence (AI) systems. By combining the robustness of watermarking with the adaptive capabilities of anomaly detection, the framework provides a multi-layered defense mechanism against various attacks in digital environments.

Improved Robustness: The DWT-based watermarking technique embeds imperceptible information into the host data, serving as a unique identifier for authentication and ownership verification. This enhances the robustness of the AI system by providing a reliable means of identifying and validating data integrity, even in the presence of malicious attacks or data tampering attempts.

Real-time Anomaly Detection: The Hybrid Anomaly Detection Framework continuously monitors system behavior using machine learning algorithms, enabling the detection and response to anomalous activities in real time. This proactive approach to security ensures timely intervention and mitigation of potential threats, thereby minimizing the impact of security breaches on AI system operations.

Maintained System Performance and Usability: Experimental results demonstrate the effectiveness of the proposed approach in detecting and mitigating attacks while maintaining system performance and usability. The integration of watermarking and anomaly detection does not compromise the functionality or efficiency of the AI system, ensuring seamless operation in various deployment scenarios.

Enhanced Integrity and Reliability: By integrating DWT-based watermarking with the Hybrid Anomaly Detection Framework, the proposed framework not only enhances security but also ensures the integrity and reliability of AI systems. This fosters trust and confidence in the deployment of AI technologies across diverse domains, bolstering their adoption and utilization for critical applications.

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCES

- Alzahrani, Ali. (2022). Enhanced Invisibility and Robustness of Digital Image Watermarking Based on DWT-SVD. *Applied Bionics and Biomechanics*. 2022. 1-13. [10.1155/2022/5271600](https://doi.org/10.1155/2022/5271600).
- E. Nowroozi, M. Mohammadi, E. Savaş, Y. Mekdad and M. Conti, "Employing Deep Ensemble Learning for Improving the Security of Computer Networks Against Adversarial Attacks" in *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 2096-2105, June 2023, doi: [10.1109/TNSM.2023.3267831](https://doi.org/10.1109/TNSM.2023.3267831)
- Ximeng, Liu., Lehui, Xie., Yaopeng, Wang., Jian, Zou., Jinbo, Xiong., Zuobin, Ying., Athanasios, V., Vasilakos. (2021). Privacy and Security Issues in Deep Learning: A Survey. *IEEE Access*, doi: [10.1109/ACCESS.2020.3045078](https://doi.org/10.1109/ACCESS.2020.3045078)

- Regev, Yuval & Vassdal, Henrik & Halden, Ugur & Catak, Ferhat Ozgur & Cali, Umit. (2022). Hybrid AI-based Anomaly Detection Model using Phasor Measurement Unit Data. 10.48550/arXiv.2209.12665.
- Suganthi, J & Nagarajan, B. & Muhtumari, S. (2022). Network Anomaly Detection Using Hybrid Deep Learning Technique. 10.3233/APC220014.
- Kumar, R., & Gopalakrishnan, N. (2021). Anomaly Detection Techniques in Artificial Intelligence: A Review. In 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 2624-2629). IEEE..
- Amrit, Preetam & Singh, Amit. (2022). Survey on watermarking methods in the artificial intelligence domain and beyond. Computer Communications. 188. 10.1016/j.comcom.2022.02.023
- Patrikar, D. R., & Parate, M. R. (2022). Anomaly detection using edge computing in video surveillance system: review. *International journal of multimedia information retrieval*, 11(2), 85–110. \
- Wang, Zhihui & Dong, Yongqiang & Xiang, Zhining & Cheng, Shaochi. (2024). An Overview of Artificial Intelligence Security Issues. 10.3233/FAIA231291.
- Hosen, M. A., & Hasan, M. (2021). An Overview of AI Security Issues, Challenges, and Solutions. In 2021 2nd International Conference on Advanced Artificial Intelligence (ICAAI) (pp. 1-7). IEEE.
- H. A. Khan, S. A. Al-Madani, M. M. Taha, A. O. A. Al-Ashrafy and S. El-Khodary, "A Survey on Anomaly Detection Techniques in Artificial Intelligence," 2020 5th International Conference on Advanced Machine Learning Technologies and Applications (AMLTA), Cairo, Egypt, 2020, pp. 1-7.
- S. Eltanbouly, M. Bashendy, N. AlNaimi, Z. Chkirbene and A. Erbad, "Machine Learning Techniques for Network Anomaly Detection: A Survey," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), Doha, Qatar, 2020, pp. 156-162, doi: 10.1109/ICIoT48696.2020.9089465.
- Jeffrey N, Tan Q, Villar JR. A Review of Anomaly Detection Strategies to Detect Threats to Cyber-Physical Systems. *Electronics*. 2023; 12(15):3283.
- A. Jamal, M. H. Alkawaz, M. -A. Fatima and M. S. Ab Yajid, "Digital Watermarking Techniques and its Application towards Digital Halal Certificate: A Survey," 2019 IEEE 7th Conference on Systems, Process and Control (ICSPC), Melaka, Malaysia, 2019, pp. 242-247, doi: 10.1109/ICSPC47137.2019.9067988.
- Begum, Mahbuba & Uddin, Mohammad Shorif. (2020). Analysis of Digital Image Watermarking Techniques through Hybrid Methods. *Advances in Multimedia*. 2020. 1-12. 10.1155/2020/7912690.
- Jay Kumar Jain, & Waoo, A. A. . (2023). An Artificial Neural Network Technique for Prediction of Cyber-Attack using Intrusion Detection System. *Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN)* ISSN: 2799-1172, 3(02), 33–42.
- Chauhan, Ms & Waoo, Akhilesh & Patheja, Pushpinder. (2012). INFORMATION HIDING WATERMARKING DETECTION TECHNIQUE BY PSNR AND RGB INTENSITY. 3. 18-22.
- Soni, Brijesh & Waoo, Akhilesh. (2023). Deep Learning: Tools and Models. 10.1002/9781119792161.ch3.
- Sharma, Siddhant & Waoo, Akhilesh. (2023). An efficient machine learning technique for prediction of consumer behaviour with high accuracy. *International Journal of Computing and Artificial Intelligence*. 4. 12-15. 10.33545/27076571.2023.v4.i1a.59.
- Bhatele, K. R., Jha, A., Tiwari, D., Bhatele, M., Sharma, S., Mithora, M. R., & Singhal, S. (2022). COVID-19 Detection: A Systematic Review of Machine and Deep Learning-Based Approaches Utilizing Chest X-Rays and CT Scans. *Cognitive Computation*, 1–38. Advance online publication.
- Tomar, R. S., Chaturvedi, P., & Bhatele, M. A Survey on Information Hiding using Water Marking Techniques.
- Mishra, M., & Bhatele, M. An Extensive Survey Expounding Security Issues & Requirement in Secure Cloud Computing Environment.
- Saganowski, L., Andrysiak, T., Kozik, R., & Choraś, M. (2016). DWT-based anomaly detection method for cyber security of wireless sensor networks. *Secur. Commun. Networks*, 9, 2911-2922.
- Bianca, Tagliaro, Beasley., George, D., O'Mahony., Sergi, Gomez, Quintana., Andriy, Temko., Emanuel, Popovici. (2020). Lightweight Anomaly Detection Framework for IoT. doi: 10.1109/ISSC49989.2020.9180205
- Nouar, AlDahoul., Hezerul, Abdul, Karim., Abdulaziz, Saleh, Ba, Wazir. (2021). Model fusion of deep neural networks for anomaly detection. *Journal of Big Data*, doi: 10.1186/S40537-021-00496-W